Sonorant voicing specification in phonetic, phonological and articulatory context

Von der Philosophisch-Historischen Fakultät der Universität Stuttgart

zur Erlangung der Würde eines Doktors der
Philosophie (Dr. phil.) genehmigte Abhandlung

Vorgelegt von

Jagoda Bruni
(geb. Sieczkowska)

aus Toruń

Hauptberichter: Prof. Dr. Grzegorz Dogil

Mitberichter: Prof. Dr. Bernd Möbius

Tag der mündlichen Prüfung: 09.05.2011

Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart

2011

Selbständigkeitserklärung

Ich erkläre hermit, daß ich unter Verwendung der im Literaturverzeichnis aufgeführten Quellen und unter fachlicher Betreuung diese Dissertation selbständig verfaßt habe.

Jagoda Bruni

Table of Contents

Acknowledgements

I would like to thank the following people for supporting me during writing this doctoral dissertation:
I would particularly like to thank *Prof. Grzegorz Dogil* for giving me the chance to work in his team, for expanding my interest in linguistics and for his great support and valuable advice.

I would also like to thank *Prof. Bernd Möbius* for sharing his voicing investigation interests with me, as well as for his invaluable professional advice.

Many thanks also to *Antje Schweitzer*, who has contributed a great deal of technical support along with much friendship and interest.

Thanks to the team of the Phonetic Institute of the Cologne University (IFL) for their cooperation: *Prof. Martine Grice, Dr. Doris Mücke, Anne Hermes and Henrik Niemann* as well as subjects JS and NL for their participation in the articulatory study.

Thanks to *Prof. Grazyna Demenko* and her working team from the Institute of Linguistics (IJ) at the Adam Mickiewicz University, who were kind enough to allow me to use their Polish adaptation of the BOSS Speech Corpus.

Similarly many thanks to *SVOX AG*, which enabled me to use their French Speech Corpus.

Moreover, I would like to thank *Marcin Włodarczak* for offering much technical and phonetic advice.

Many thanks to the SFB's director *Prof. Artemis Alexiadou* and SFB coordinator *Sabine Mohr* for their kindness and support.

Thanks also to *Natalie Lewandowski* and *Matthias Jilka* for all their linguistic advice and friendly support.

Many thanks to the *IMS-Phonetik group*, particularly to *Daniel Duran, Kathrin Schneider, Sabine Dieterle, Katie Schweitzer, Michael Walsh, Olga Anufryk, Uta Benner, Jörg Mayer, Britta Lintfert* and *Andre Blessing* for creating very friendly work atmosphere.

This dissertation is part of a research project SFB 732 (A2) 'Specification in Context', founded by the German Research Foundation (DFG). Without this project and financing following work would not be possible.

Special thanks to:

my *husband*, for more than words can describe.

my *parents*, for teaching me the most valuable lessons in life.

all my *family* and my *friends*, especially *Agnes*, for their support.

Madame B.D. for her warmth and care.

Abbreviations

BLF – BOSS Label File

BOSS – Bonn Open Synthesis System

CSM – Context Sequence Model

EMA – Electromagnetic Articulograph

ESPS – Entropic Signal Processing System

HMM – Hidden Markov Model

IPA – International Phonetic Alphabet

IMS Festival – Festival tool, version created at the Institut für Maschinelle Sprachverarbeitung

ISCM – Incremental Specification in Context Model

L1, 2, 3 – first/second/third language

MLM – Multilevel Exemplar Model

OT – Optimality Theory

SAMPA - Speech Assessment Methods Phonetic Alphabet

SPE – Sound Pattern of English

SSFF – Simple Signal File Format

VOT – Voice Onset Time

TIMIT - Texas Instruments, Inc. (TI) and Massachusetts Institute of Technology (MIT)

Summary

Sonorant voicing specification in phonetic, phonological and articulatory context.

This dissertation describes investigation of voicing profiles of sonorants in Polish, American English, French and German. Automatic analysis of voicing described in this paper was first proposed by Möbius (2004). This computational approach enables extraction of complex information with regard to an investigated segment like its position in a word, manner of articulation and voicing probability. Moreover, this dissertation presents also an articulatory study of Polish liquids. Thus, sonorants described in this work are analyzed from phonetic, phonological and articulatory perspective.

Structure of the dissertation

Chapter 1 provides theoretical background to studies concerned with speech production and various perspectives of voicing analysis. In the first part of the chapter following issues are reviewed: articulation of the vocal folds, phonation modes, Quantal Theory (Stevens 1989) and Optimality Theory (Prince & Smolensky 1993), Voice Onset Time measurements (Lisker & Abramson 1964) and description of features [voice] and [spread glottis]. The second part describes phonetic and articulatory properties of liquids of the four investigated languages. Finally, last part of this chapter provides an overview of the basic assumptions of Exemplar Theory (Nosofsky 1988; Lacerda 1995, Pierrhumbert 2001 and others) and a notion of Context Specification (Alexiadou 2006, 2010).

Chapter 2 describes methodology applied in the dissertation, which follows the one presented in Möbius (2004). Automatic analysis of voicing probabilities of sonorants is shown for segments in different contexts and positions. It is based on feature extraction using IMS German Festival tool (2009), which enables tree-structured analysis of speech utterances. Results obtained this way are referred to as *voicing profiles* and illustrate the percentage of voiced exemplars of a segment in a corpus across time. Moreover, structure of speech databases is described for Polish, German, American English and French corpora.

Chapter 3 presents results for the four investigated languages. For Polish and French the analysis of sonorants is narrowed down to liquids /l/ and /r, R/. General tendency shows initial devoicing of German and American English sonorants with left-hand voiceless obstruent context. In case of German, this effect is stronger for sonorants separated from the voiceless segments by a syllable boundary (Möbius 2004). Much stronger devoicing tendencies throughout all time duration of sonorant exemplars are observed for Polish and French. Rhotics /r, R/ devoice entirely in word-final positions with left voiceless obstruent context.

Chapter 4 provides an overview of cross-linguistic influences during second (third, fourth etc.) language acquisition. It is posited that the process of foreign language learning depends on many factors including language experience, categorization of speech events, facilitation and competition processes, markedness degree and phonetic talent. It is however hypothesized, that the strongest role during cross-linguistic exemplar storage is played by the acoustic and linguistic context match, which takes place on a segmental level (Wade et al., 2010). In the chapter results from voicing investigation (chapter 3) are used in the simulation of speech production undergoing cross-linguistic phonetic transfers.

Chapter 5 describes the articulatory study conducted on Polish using Electromagnetic Articulograph (EMA).[1] Acoustic and articulatory recordings obtained in this study resulted in generation of voicing and articulatory profiles of sonorants with left-hand voiceless obstruent context in onset and coda consonant clusters. The purpose of this investigation was to observe whether differences in behavior of voicing are correlated with gestural coordination of the articulators during production of the consonant clusters. Results indicate that Polish C_1C_2V onsets (where C_1 was the voiceless obstruent and C_2 the sonorant) show tendencies of forming the so-called C-Center effect (Browman & Goldstein 1988; Byrd 1995; Hermes et al. 2008; Mücke et al. 2009). It is a global organization of consonant clusters which is formed by consonant positioning at a stable distance with regards to a vowel target. On the other hand, Polish VC_1C_2 coda clusters exhibit less bounding and no C-Center coordination. Thus, the onset consonants show in-phase relation towards each other, whereas the coda clusters anti-phase relation.

Chapter 6 provides a summary of the dissertation concerning voicing and articulatory profiles. Influences of articulatory strengthening are discussed along with new approaches to analysis of feature [voice]. Furthermore, the relevance of context specification is discussed with regard to speech production processes. The chapter is concluded with the thesis of the dissertation that voicing and articulatory changes observed in sonorants in the four investigated languages demonstrate *contextual influences* due to *context specifications* that take place in particular phonetic, phonological and articulatory surrounding.

[1] This study was conducted thanks to the courtesy and in cooperation of Prof. Martine Grice, Dr. Doris Mücke and their colleagues from the Institute of Linguistics at the University of Cologne.

Zusammenfassung

Stimmhaftigkeitsspezifikation von Sonoranten im phonetischen, phonologischen und artikulatorischen Kontext

Diese Dissertation beschäftigt sich mit den Stimmhaftigkeitsprofilen von Sonoranten im Polnischen, Amerikanischen Englisch, Französischen und Deutschen. Die vorliegende Studie benutzt eine neue komputationell-automatische Methode der Stimmhaftigkeitsanalyse, welche zuerst in Möbius (2004) vorgestellt wurde. Dieses Verfahren ermöglicht die Extraktion von komplexen Informationen über die Positionen von Konsonanten innerhalb von Wort und Silbe, sowie deren Artikulationsart und Stimmhaftigkeitseigenschaften, so dass die Stimmhaftigkeitsspezifikationen von Sonoranten in den bereits erwähnten vier Sprachen im Hinblick auf derartige kontextuelle phonetische, phonologische und artikulatorische Aspekte untersucht werden können.

Stimmhaftigkeit in der Sprachproduktion und als kontrastives Merkmal

Der Begriff der Stimmhaftigkeit („voicing“) deckt mehrere verschiedene Aspekte ab. Die vorliegende Untersuchung beginnt dabei mit der Betrachtung der grundlegenden physikalischen Aspekte der Sprachproduktion. Hier wird Stimmhaftigkeit einfach als die akustische Präsenz harmonischer Signale beschrieben, welche in Spektralbildern sichtbar gemacht werden kann, und der artikulatorisch eine reguläre Stimmlippenvibration, bei der die Luft von den Lungen durch den Mund bzw. auch die Nase fließt, entspricht (van den Berg 1958; Saltzman & Byrd 2003; Clark & Yallop 2007). Dieser Vorgang wird als „normale“ Phonation bezeichnet, im Gegensatz

zu „breathy“ und „creaky“ Phonationen, die in Kapitel 1 im Detail beschrieben sind. Ebenfalls in diesem Kapitel diskutiert werden Modelle des menschlichen Vokaltrakts als sogenannte akustische Röhre („Acoustic Tube Model“), insbesondere mit Bezug auf Fants Akustische Theorie der Sprachproduktion (Fant 1970), sowie deren Weiterentwicklung durch Johnson (1997). Auf der Grundlage dieser Form der Modellierung des Vokaltrakts wird im abschließenden Kapitel auch die nicht-lineare Interaktion akustischer und artikulatorischer Parameter im Sinne von Stevens‘ Quantal Theorie (Stevens 1989, siehe auch Johnson 1997) zur Analyse der Stimmhaftigkeit von Sonoranten herangezogen.

Auch die Optimalitätstheorie beschäftigt sich mit dem Phänomen der Stimmhaftigkeit (Prince & Smolensky 1993; Prince & McCarthy 1995; Moosmüller & Ringen 2004). In der OT werden linguistische Prozesse bekanntlich als universelle Beschränkungen, die sich miteinander im Wettbewerb befinden, dargestellt. Das Problem der Stimmhaftigkeit wird als Beispiel für den Konflikt zwischen Markiertheits- („markedness“) und Treuebeschränkungen („faithfulness“) verwendet, z.B. im Englischen (Lombardi 1996), Ungarischen, Russischen (Petrova & Szentgyörgyi 2004) und Türkischen (Kallestinova 2004).

Aus der phonologischen Perspektive erlaubt die Stimmhaftigkeit verschiedene Analysen auf der Grundlage der Opposition der kontrastiven Merkmale [voice] vs. [spread glottis]. Das Merkmal [voice] wurde von vielen Autoren beschrieben, zuerst von Jakobson und Halle (1956), dann von Chomsky und Halle (1968) aufgenommen und als neue phonologische Idee weiterentwickelt. Aufschlussreiche sprachübergreifende Untersuchungen im Hinblick auf den Kontrast mit [spread glottis] folgten durch Keating (1998), Kingston und Diehl (1994, 2010), Halle und Stevens (1971), Lombardi (1991) und andere.

In der wissenschaftlichen Literatur wird „voicing“ sehr oft durch den Parameter der „Voice Onset Time“ (VOT) beschrieben. Diese Form der Analyse von Verschlusslauten wurde zuerst von Lisker und Abramson (1964) demonstriert, und danach beispielsweise von Ladefoged (1971), Keating (1980), Poon & Mateer (1985) anhand vieler weiterer Sprachen fortgeführt.

Der darauf folgende Abschnitt des ersten Kapitels konzentriert sich auf die akustische und artikulatorische Analyse von Sonoranten im Polnischen, Deutschen, Amerikanischen Englisch und Französischen, mit besonderem Fokus auf Liquide, die in Stimmhaftigkeitsprofilen die größten Tendenzen Richtung Stimmlosigkeit nachgewiesen haben. Außerdem enthält diese Sektion eine Beschreibung der phonotaktischen Regeln dieser vier Sprachen, welche in Kapitel 3 als Basis für die Sonorant Extraktion aus den Korpora genutzt werden.

Als letzte der theoretischen Grundlagen in Kapitel 1 wird die Exemplartheorie (Nosofsky 1988; Lacerda 1995; Pierrhumbert 2001, Bybee 2006) vorgestellt. Besonderes Gewicht wird auf die phonetischen/phonologischen und artikulatorischen Kontextfaktoren (z.B.Frequenzphänomene) gelegt (Schweitzer & Möbius 2004; Walsh et al. 2007), welche durch das Context Sequence Model (Wade et al. 2010), das Multi-Level Exemplar Model (Walsh et al. 2010) und das Incremental Specification in Context Model (Alexiadou 2006, 2010; Dogil & Möbius 2001; Möbius & Dogil 2002; Schneider et al. 2006; Dogil 2010) beschrieben werden und als Grundlagen für die Schlussdiskussion dienen.

Ergebnisse der vorliegenden Studie

Stimmhaftigkeitsprofile

In den durchgeführten Experimenten wurden die Stimmhaftigkeitsprofile des Polnischen, Amerikanischen Englisch, Französischen und Deutschen untersucht. Die Stimmhaftigkeitsprofile

von polnischen Sonoranten wurden darüber hinaus auch mit artikulatorischen Profilen verglichen und analysiert.

Die Stimmhaftigkeitsprofile wurden durch automatische Analyse erlangt, indem vier mit professionellen Sprechern aufgenommene Korpora (Demenko et al. 2008; SVOX AG; Ostendorf et al. 1995; SmartKom 2003, Schweitzer et al., 2003) mittels IMS Festival Tool (2009) untersucht wurden. Die Festival-Software enthält Funktionen für die Text-to-Speech Analyse und generiert linguistische Repräsentationen einer Äußerung aus dem eingegebenen Text und wiederum akustische Eigenschaften aus den linguistischen Repräsentationen. Die akustischen Details hängen somit von den phonologischen und anderen linguistischen Eigenschaften (wie sie in Festival implementiert sind) dieser Repräsentationen ab. Das Programm baut zwei Relationssorten: „flache“ Relationen, die den Ebenen in der linguistischen Struktur entsprechen (zum Beispiel ‚segment‘, ‚syllable‘, ‚word‘) und „hierarchische“ Baumrelationen, die die Ebenen verbinden (zum Beispiel ‚phrase‘, ‚intonation‘, ‚syllable structure‘). Mittels dieser Beschreibungsebenen wurden die vier oben genannten Korpora analysiert und folgende Parameter extrahiert: linker und rechter Kontext der Sonoranten, Stimmhaftigkeit/Stimmlosigkeit in neun Abstufungen, Position in Wort und Silbe, Artikulationsart und Artikulationsort der Sonoranten und der benachbarten Segmente. Das Stimmhaftigkeitsprofil eines Sonoranten wurde durch die Messung von neun (von 10% bis 90% der Lautdauer), äquidistanten Stimmhaftigkeit/Stimmlosigkeitswerten erstellt, die durch das ESPS Tool get_F0 ermittelt und in binärer Form (1 oder 0) ausgegeben wurden, basierend auf je 10 ms des akustischen Signals. Diese so genannte „Frame-by-frame“-Analyse ermöglicht die Extraktion von beliebigen Kontexten und Positionen der zu untersuchenden Sonoranten

und erlaubt dadurch den Aufbau eines graduell differenzierbaren Stimmhaftigkeitsbildes. Die statistische Analyse bestimmt jeweils die Mittelwerte jedes Sonoranten in einem gegebenen Kontext.

Die Ergebnisse der Untersuchung der polnischen und französischen Sonoranten zeigen ähnliche Tendenzen. Es ist zu beobachten, dass Sonoranten, die Vokalen oder anderen Sonoranten folgen, egal ob sie am Wortbeginn oder Wortende stehen, sehr geringe Neigung zur Entstimmlichung zeigen. Sonoranten, welche stimmlosen Obstruenten folgen, weisen dagegen eine Entstimmlichung von im Durchschnitt 25% (Polnisch) bis zu 60% (Französisch) auf. Eine noch stärkere Tendenz in Richtung Stimmlosigkeit (100%) kann für das polnische /r/ und das französische /R/ im wortfinalen Kontext festgestellt werden.

Im Gegensatz hierzu scheinen die Stimmhaftigkeitsprofile des Amerikanischen Englischen und des Deutschen mehr vom linken Kontext und Silbengrenzen zwischen Obstruenten und Sonoranten abhängig zu sein als von der Position im Wort. Im Deutschen behalten Sonoranten mit Vokalen oder anderen Sonoranten im linksseitigen Kontext fast immer 100% ihrer Stimmhaftigkeit, mit vorausgehenden stimmlosen Obstruenten jedoch zeigen sie signifikante Tendenzen zur Entstimmlichung, die zwischen 50% und 100% schwanken (die Stimmhaftigkeit steigt mit der Lautdauer). Außerdem steigt die Wahrscheinlichkeit einer stimmlosen Aussprache, wenn es keine Silbengrenze zwischen den stimmlosen Obstruenten und den Sonoranten gibt. Im Amerikanischen Englischen wurden zwei unterschiedliche Beobachtungen gemacht[2]: die Ergebnisse des Korpusteils „lab news“ zeigen wortinitial und final eine Entstimmlichung von bis zu 15% im Kontext vorausgehender Vokale/Sonoranten. Stimmlosen Obstruenten folgend weist

[2] Es wurden zwei verschiedene Teile des Korpus (identischer Sprecher) analysiert: „Laboratory News“ – manuell und automatisch annotiert und „Radio News“ – nur automatisch annotiert.

das Amerikanische Englisch Entstimmlichungstendenzen von 5% bis 15% nur am Anfang der Sonoranten auf. Im Korpusteil „radio news“ werden wortinitiale Sonoranten mit vorausgehenden Vokalen/Sonoranten dagegen am Ende des Segments bis zu 30% entstimmlicht. Im Falle vorausgehender stimmloser Obstruenten sind hauptsächlich /m/ und /w/ am Anfang des Wortes stimmlos (bis zu 17%).

Es ergibt sich also der Eindruck, dass der segmentale Kontext besonders einflussreich ist. In Anlehnung an Gussmann (1992, 2007) kann die wortfinale Stimmlosigkeit von polnischen Sonoranten im Kontext linksseitiger stimmloser Obstruenten als ein Resultat von „Desyllabifizierung“ interpretiert werden, mit anderen Worten, der Sonorant ist stimmlos, weil er von der Silbe ausgeschlossen ist. Gussmann postuliert, dass die Lizensierung von [voice] im Polnischen nur stattfinden kann, wenn das Segment Teil einer Silbe ist. Deswegen sind Sonoranten wie das finale [r] in [vjat̪r̥] *Wind* stimmlos und extrasyllabisch. Auf der Grundlage dieser Erkenntnisse kann ebenfalls erklärt werden, warum französische Sonoranten in Clustern mit stimmlosen Obstruenten (wie zum Beispiel in qua[tR̥] *vier*) Messergebnisse von bis zu 100% Stimmlosigkeit liefern.

Anderseits zeigt ein Modell der Wahrscheinlichkeit der Entstimmlichung im Deutschen und Englischen einen sehr starken Einfluss des linksseitigen Kontexts (der Präsenz von stimmlosen Obstruenten vor die Sonoranten). Tendenzen zur Stimmlosigkeit am Wortanfang rühren vom Anteil des [spread glottis] Merkmals her, das sich vom stimmlosen Obstruenten zum Sonoranten ausbreitet (Kingston & Diehl 1994).

Nach Recasens (1989) und Stevens (1972, 2010) wird postuliert, dass die Laute, die in kleineren Teilregionen des Vokaltrakts produziert werden (wie z.B. Palatale), weniger Koartikulationseffekte erlauben als Laute, die mit größeren Konstriktionen des Vokaltrakt

produziert werden. Aus diesem Grund ist es nachvollziehbar, dass Laute mit schmaleren Quantalregionen mehr koartikulatorische Resistenz aufweisen als Laute mit breiten Quantalregionen, da sie weniger „Platz“ während der Sprachproduktion verbrauchen und dadurch auch weniger Varianz ermöglichen. Deswegen ist es durchaus begründet zu argumentieren, dass die Stimmlosigkeitstendenzen von /w/ und /m/ am Wortanfang nach stimmlosen Obstruenten, wie sie im Amerikanischen Englischen gefunden wurden, durch eine geringere koartikulatorische Resistenz bedingt sind.

Artikulatorische Profile

Zusätzlich wird auch eine artikulatorische Analyse polnischer Sonoranten vorgenommen. Untersucht werden die Positionen am Wortanfang und –ende mit jeweils vorausgehenden stimmlosen Obstruenten. Ziel dieser Studie war es, eine artikulatorische Erklärung für das Stimmlosigkeitsverhalten in der Coda zu finden. Die Analyse wurde in Kooperation mit dem Institut für Linguistik an der Universität zu Köln durchgeführt.

Die Hypothese dieser Studie folgt älteren Untersuchungen von Browman und Goldstein (1988), Byrd (1995), Hermes et al. (2008), sowie Mücke et al. (2009), die sich auf die Analyse des C-Center Effekts konzentriert haben. Dieser Effekt besagt, dass in Onset-Konsonantenclustern ein temporales Gravitationszentrum (das Mittel der konsonantischen Einzelelemente) existiert, welches stets die gleiche Distanz zum Vokal hat (Browman & Goldstein 1988). Beispielsweise im Italienischen (Hermes et al. 2008) bauen Konsonanten am Anfang des Wortes den C-Center Effekt auf, wenn neue Konsonanten (C_2 und C_3) zu einem schon bestehenden Cluster C_1V addiert werden, wie in ‘Lina’ (Eigenname) - C_1V → ‘plina’ (Logatom) → C_1C_2V und ‘splina’ (Logatom) → $C_1C_2C_3V$. Entscheidend für den Effekt ist, dass sich C_2 (ursprünglich C_1) nach

rechts in Richtung des Vokals verschiebt, um Platz für ein neues C_1 zu machen. Das Phänomen der Konsonantenkoppelung (consonant coupling) am Anfang des Wortes ist sehr weitgehend beschrieben (Browman & Goldstein 1988; Hermes et al. 2008; Mücke et al. 2009 et al.), aber es wurde auch entdeckt (Honoref & Browman 1995; Nam 2007), dass Cluster in Coda Positionen mehr Variabilität und Irregularität in Kupplungsrelationen zeigen. Für die hier beschriebene artikulatorische Studie wurden drei Muttersprachler des Standardpolnischen mit einem 2D Elektromagnetischen Artikulographen (EMA), Carstens AG100, aufgenommen (10 Kanäle). Die artikulatorische Analyse wurde mit dem EMU Speech Database System (Cassidy & Harrington 2001) gelabelt. Die akustische Analyse, also die Extraktion der Stimmhaftigkeitsinformationen, wurde mittels Praat durchgeführt. Die zu analysierenden Sonoranten wurden von Hand aus dem akustischen Signal geschnitten, während die temporale Bestimmung der individuellen Bereiche der Stimmhaftigkeitsabstufungen durch Interpolation erlangt wurde und die statistische Analyse mit R (2009) durchgeführt wurde.

Die Ergebnisse lassen zwei Tendenzen erkennen. Erstens, wie in den bereits genannten Studien gezeigt, gibt es klare Evidenz für Konsonantenkopplung, indem sich in /kr/, /pr/ und /pl/ Onset-Clustern der erste Konsonant C_1 nach links verschiebt und der zweite, C_2, nach rechts – ein deutliches Beispiel für Koordination mittels des C-Center Effekts. Zweitens konnte beobachtet werden, dass sich die C_2-Elemente in VC_1C_2 Clustern (/kr/, /pr/ und /pl/) am Ende des Wortes nach rechts verschieben. Die C_1 Bewegungen sind dagegen nicht regelmäßig und zeigen bei allen drei Sprechern verschiedene Tendenzen.

Die Stimmlosigkeits- bzw. Stimmhaftigkeitsprofile weisen ähnliche Ergebnisse auf wie die Profile, die in der ersten in dieser Dissertation beschriebenen Stimmhaftigkeitsuntersuchung

ermittelt wurden, wo in der Coda positionierte Liquide, welche stimmlosen Obstruenten nachfolgen, die stärksten Neigungen zur Entstimmlichung demonstrieren.

Der Kontext beeinflusst auch die artikulatorischen Profile polnischer Sonoranten mit linksseitigen stimmlosen Obstruenten, besonders am Ende des Wortes. Die Tatsache, dass der C-Center Effekt in der Coda Position nicht gefunden wurde weist auf die besondere phonologische Lizensierung von Liquiden und die Auswirkung ihrer Position außerhalb der Silbe hin.

Aufbau der Arbeit

Kapitel 1 führt neben Studien der theoretischen Grundlagen der Sprachproduktion spezifische Forschungsarbeiten ein, die sich mit dem Thema Stimmhaftigkeit beschäftigen. Stimmhaftigkeit wird aus den Perspektiven der Stimmlippenfunktion und des Mechanismus des Lautaufbaus beschrieben, sowie im Hinblick auf akustische quantale Spracheigenschaften, die Optimalitäts-Theorie und die Unterscheidung der distinktiven Merkmale [voice] vs. [spread glottis]. Im zweiten Teil dieses Kapitels werden die akustischen und artikulatorischen Eigenschaften der Sonoranten des Polnischen, Französischen, Deutschen und Amerikanischen Englischen illustriert. Abschließend wird die Exemplartheorie im Zusammenhang mit der Idee von Prozessen der Kontextspezifikation vorgestellt und erläutert.

In *Kapitel 2* wird die Methode zur Bestimmung der Stimmhaftigkeitsprofile dargestellt. Zuerst werden vier mit professionellen Sprechern aufgenommene Korpora mit einer genauen Beschreibung der Phonemdatenbank und der Formatierung dieser Sprachressourcen vorgesellt. Nachfolgend wird die komputationelle Bearbeitung der Korpora illustriert, besonders im Hinblick auf die Nutzung des Festival TTS Tools und die statistische Bearbeitung der Messergebnisse.

In *Kapitel 3* werden die Untersuchungsergebnisse präsentiert und zu den Stimmhaftigkeitsprofilen der Sonoranten im Polnischen, Deutschen, Amerikanischen Englischen und Französischen in Abhängigkeit von vorangehenden Vokalen, Sonoranten und stimmlosen Obstruenten verarbeitet.

Kapitel 4 beschreibt sprachübergreifende Einflüsse während des Zweit- (Dritt-, Viert-, etc.) Spracherwerbs. Es wird postuliert, dass der Vorgang des Erlernens einer Fremdsprache von vielen Faktoren abhängt, etwa Spracherfahrung, Kategorisierung sprachlicher Ereignisse, Förderungs- und Wettbewerbsprozessen, Grad der Markiertheit sowie phonetischem Talent. Es wird aber auch die Hypothese vertreten, dass die wichtigste Rolle bei der Speicherung sprachübergreifender Exemplare von der Abgleichung des akustischen und linguistischen Kontextes, die ja auf der segmentalen Ebene stattfindet (Wade et al. 2010), übernommen wird. Das Kapitel nutzt die Ergebnisse der Stimmhaftigkeitsuntersuchung (Kapitel 3) bei der Simulation von Sprachproduktionsvorgängen, bei denen sprachübergreifende phonetische Transfers ablaufen.

In *Kapitel 5* werden die artikulatorischen EMA-Profile (welche zusammen mit dem entsprechenden akustischen Signal aufgenommen wurden) von polnischen Sonoranten mit vorangehenden stimmlosen Obstruenten in Onset- und Coda-Positionen vorgestellt. Die theoretischen Grundlagen des C-Center Effekts und der Konsonantenkopplung werden erklärt und von einer Darstellung der Ergebnisse gefolgt.

Das abschließende *Kapitel 6* beschäftigt sich mit der Analyse der präsentierten Ergebnisse. Im ersten Teil werden die Untersuchungen der Stimmhaftigkeitsprofile aus der Perspektive der artikulatorischen Verstärkung und der quantalen Relationen diskutiert. Das Kapitel endet mit einer Diskussion der neuesten Ansichten zu [voice] als distinktivem phonologischem Merkmal

und der Illusionen, mit denen die Wahrnehmung von Merkmalen verbunden sein kann. Diese Illusionsdiskussion führt direkt zur Grundhypothese dieser Dissertation – Kontextspezifizierung und ihre Mechanismen.

CHAPTER 1

Introduction

This dissertation is concerned with the phonological, acoustical and articulatory study of voicing contrasts in Polish, French, German and American English sonorants. In Chapter 1 I will introduce general issues concerning voicing and its properties. I will go on to describe the methodology (computational procedure of obtaining 'voicing profiles') in Chapter 2. In Chapter 3 I will present results for four languages and provide a description of the cross-linguistic second and third language acquisition processes in Chapter 4. The articulatory study of Polish obstruent-sonorant clusters will be presented in Chapter 5. The dissertation will conclude with a discussion, to be found in Chapter 6.

1.1 Motivation

Pinker (1995:18) said that language is "a distinct piece of the biological makeup of our brains. (…) It is a complex, specialized skill, which develops in the child spontaneously, without conscious effort or formal instruction, is deployed without awareness of its underlying logic, is qualitatively the same in every individual, and is distinct from more general abilities to process information or behave intelligently." Speech production involves many complex processes, starting from the physiological production of a wave signal, using motor control and neural networks, going through the phonological organization of speech into units and contrasting them by differentiating production manners, and ending with speech perception, speaker-listener interaction and adjustment (Saltzman & Byrd 2003: 1072-1076). Voice is a natural human

property of speech, as much as language is a natural ability of a human being. The term *voicing*[3] can be characterized in many dimensions with regard to various aspects. From the perspective of speech production it is defined as a periodic vocal fold vibration produced as a result of laryngeal actions occurring along most or all of the length of the glottis. As investigated in numerous studies (Chomsky & Halle 1968: 1-470; Browman & Golsdtein 1992: 155-180; Hawkins 2010: 60-89), the feature [voice] corresponds to the normal mode of phonation, i.e. periodic vibration of the vocal folds. The feature [spread glottis] serves as a characteristic to describe the wide state of the glottis, in which a large airwave flows through the vocal folds inhibiting the voicing process.

Another way to define voicing, introduced by Lisker and Abramson (1964: 384–422) and investigated by many researchers cross-linguistically (e.g. Keating 1984: 286-319; Ladefoged & Maddieson 1996: 1-425; Shimizu 1990: 1-13; Poon & Mateer 1985: 39-47), is through the analysis of the voice onset time (VOT) - the time interval between the release of a stop occlusion and the onset of vocal fold vibration. A variety of categories which define voicing with regard to voice onset (fully voiced, voiceless unaspirated and voiceless aspirated categories defined by Lisker & Abramson 1964; and two additional categories: partly voiced, voiceless slightly aspirated defined by Ladefoged 1971: 1-122) are well-established forms of analyzing stop consonants.

Jessen (2000: 11-64) proposed a set of eight correlates to classify consonants (aspiration duration, closure voicing, fundamental frequency onset, first formant onset, closure duration,

[3] In this work I will refer to the features [voice] and [spread glottis] as factors determining or influencing the voicing profiles of sonorants in four languages (Polish, French, German and American English). However, the usage of those terms will apply to a different extent depending on the language under investigation.

preceding vowel duration, following vowel duration, difference between the amplitude value of the first and second harmonics). These are relevant for the new definition of the features [voice] and [tense] in a new model of the range of acoustic/auditory correlates of these features.

While all the methods listed above have been investigated and implemented successfully in the cross-linguistic studies of voicing, they have not proven to be a satisfactory way to analyze classes of consonants like sonorants or fricatives. In order to investigate voicing dependencies, the occurrence and change of sonorants in Polish, French, German and American English, I have applied an automatic analysis method proposed by Möbius (2004: 5-26). As phonotactic studies require large datasets, this frame-by-frame analysis seems to be a reasonable solution to provide rich voicing information based on the time duration of the investigated segment, as well as the frequency of its occurrence. This computational, data driven method of looking at voicing might be an alternative to previous voicing investigations (Jessen 2000; Keating 1984; Lisker & Abramson 1964). Already applied in German, Mandarin Chinese, Hindi, Mexican Spanish and Italian (Möbius 2004), it has now been expanded and slightly modified for the purpose of the study of Polish, French and American English (Sieczkowska et. al 2010: 1549-1552).

1.2 Background

The following sections will provide the traditional background information to frame the investigation concerning voicing and its phonetic/phonological properties. Section 1.2.1 describes the processes of articulation and phonation, as well as voicing properties from the perspective of speech production. It will also outline the basic concept of the Acoustic Theory of Speech Production (Fant 1970). The following section, section 1.2.2, will demonstrate quantal properties of speech and their impact on voicing descriptions. Section 1.2.3 presents selected insights on

voicing with regard to Optimality Theory. Section 1.2.4 will address the distinction between the features [voice] and [spread glottis] in the view of voicing classification. Section 1.2.5 provides an overview of the analysis of voicing by means of the feature Voice Onset Time (VOT) and is followed by a phonotactic, articulatory and acoustic analysis of the sonorants occurring in the languages to be examined in this thesis (the four subsections of section 1.3 corresponding to the four languages: Polish, French, German and American English under investigation). Finally, section (1.4) describes basic assumptions of Exemplar Theory and the models of Specification in Context developed in SFB 732[4]. A brief summary follows in section 1.5.

1.2.1 Speech production, phonation and tube models

The production of normal speech involves the lungs, trachea, larynx and vocal cavities. It is held by the vocal folds which are activated by a stream of air delivered trough the lungs and trachea (van den Berg 1958: 227-243). Constrictions formed in the supralaryngeal vocal tract by the lips and tongue parts (tip, body and root) enable the creation of resonance tubes with varying resonance frequencies along the vocal tract (Saltzman & Byrd 2003).

Phonation is a result of the vibratory cycle of the vocal folds. It is driven by the opening and closing phases, where the Bernoulli effect takes place. Periodic sound generated at the larynx and through the vocal tract is later shaped by the cavities of the tract and transformed by egressive pulmonic airflow. Three auditory dimensions define phonation – timbre (sometimes also referred to as quality of the sound), pitch and loudness. The first is said to be determined by a mode of vocal fold vibration during phonation. It can be measured with an opening quotient (dividing the glottal opening during one cycle by the duration of the entire cycle). Pitch, on the

[4] Sonderforschungsbereich 732 – research grup at the Stuttgart University founded by the German Science Foundation.

other hand, is the perceptual correlate of the frequency of vibration of the vocal folds, while perceived loudness depends on the level of pressure maintained below the glottis during speech production, shortly defined as subglottal pressure (Clark & Yallop 2007: 184-189).

Voice quality variation differs in relation to variations in glottal opening. Figure 1 (Klatt & Klatt 1990: 822) shows positions of the arytenoid cartilages during laryngealized, modal and breathy phonation. In the laryngealized phonation (Fig.1: first row, first picture) the arytenoid cartilages are allocated in a way as to close off the glottis. After the application of lung pressure, the vocal folds vibrate and produce a waveform at glottal volume velocity (Fig.1: second row, first picture), where the duration of the open portion of the fundamental period is relatively short. Fundamental frequency decreases over time during laryngealization, with a reduction in the fundamental component of the source spectrum. In the modal voice, the vocal folds are closely approximated (Fig.1: second row, first picture), which results in an opening quotient of around 50-60% of the period (Fig.1: second row, second picture); a normal voicing source has an average decrease. In the glottal/breathy mode the arythenoid cartilages are separated at the back (Fig.1: first row, third picture) while the vocal folds vibrate. This causes a large amount of air leakage (Fig.1: second row, third picture). An increased airflow results in the generation of turbulent aspiration noise, which occurs along with the periodic voicing (Klatt & Klatt 1990).

Normal vocal fold vibration, which results in producing voiced sounds, was analyzed and described by van den Berg (1958) in his Aeorodynamic Myoelastic Theory of Phonation. In this theory, the author presents a complex mechanism of muscle and tissue cooperation along with accompanying aerodynamic forces which, taken together, form a system of vocal fold vibration. According to van den Berg (1958), the process starts with abduction of the vocal folds (closure of the glottis), behind which airflow builds up, forcing the folds to spread apart in order to allow

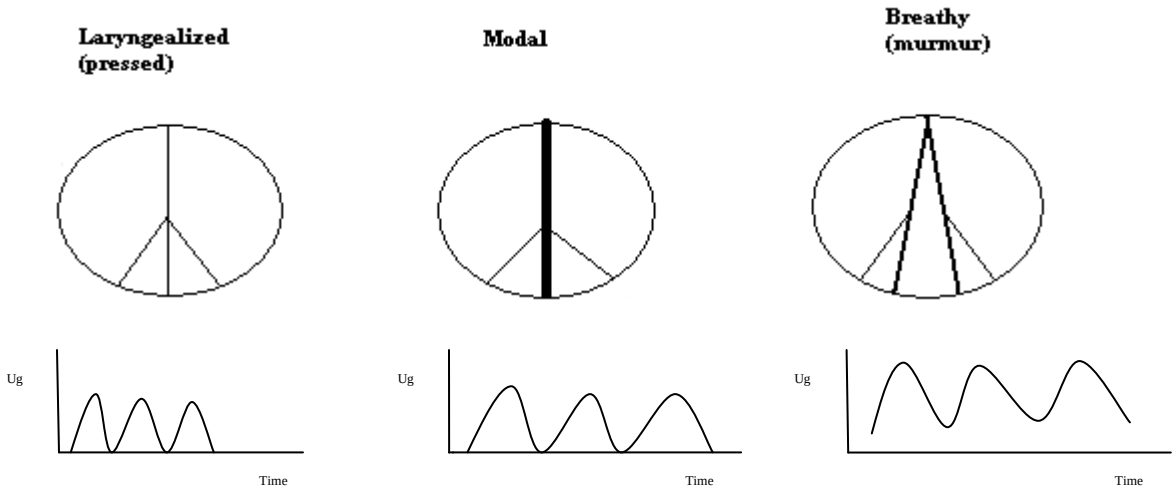


Fig.1: Glottal configurations in first row, opening at the arytenoids and resulting volume velocity waveforms (second row) (aKlatt & Klatt 1990: 822)

airflow through the glottis. In the next step, air passing through the narrow opening accelerates and its pressure drops to ultimately stop completely (Bernoulli Effect), which in turn causes the focal folds to close again due to pressure suction. These movements are possible thanks to the elasticity of the vocal folds.

Following the assumptions of the Myoelastic-Aerodynamic Theory of Phonation (van den Berg 1958) in their voicing investigation, Keating & Westbury (1986: 145-166) have proposed an aerodynamic model of voicing for stops in order to investigate when and to what extent voicing is likely to occur. Their expectations depend on two assumptions: (1) that voicing will occur whenever the states of the glottis and vocal folds are suitable for voicing and there is a sufficient pressure drop between the trachea and the pharynx; and (2) that the acoustic and physiological realization of an utterance depends upon its articulation. Taking the example of modern Polish utterance-final stops /b,d,g/, which are said to be devoiced, and final voiced stops in the speech of children acquiring American English, the authors report that oscillographic analyses show final

/b,d,g/ in both Polish and the developing speech of young children to have more closure voicing (ca. 30-40 ms) than their underlying voiceless counterparts (c.10-20ms). Overall the model provides evidence for the distribution of acoustically voiced and voiceless stop consonants in the pre-contrast stages of children's speech. No relevant data, however, is provided for consonants in clusters. In languages with no stop consonant voicing contrast, these segments tend to be voiceless in all positions. In languages with contrast, however, relatively few examples of variation have been found (Keating & Westbury 1986).

In the Acoustic Theory of Speech Production, Fant (1970: 1-328) describes relations between speech production and the acoustical data (speech wave). His theory bases its assumption on the source-filter properties resulting from the characteristics of the vocal tract and its resonating tubes. The author investigates articulatory patterns by conducting X-ray studies of Russian articulations in order to reconstruct spectral images of speech production. Furthermore, Fant (1970) proposes a source-filter model where the vocal fold vibration serves as a staple of sound energy, while the vocal tract is thought to serve as an acoustic filter that modifies a sound. Voiced sounds have their source in a periodic glottal excitation and a filter, depending on lip protrusion or tongue position. Fricatives, by contrast, are a result of turbulent noise produced at a constriction in the oral cavity (voiced fricatives have two sources: at the glottis and the supra-glottal constriction). Thanks to this model, it has become possible to calculate formant frequencies of a sound when one has information about the length of the vocal tract, which is described as a set of 'tubes' where the acoustic effects take place.

"There is some degree of correspondence between the phonetic term *phonation* and the technical term *source* and similarity between *articulation* and *filter*. This analogy implies, of course, that phonation is held apart from articulation in the sense of the generation of sound versus the specific shaping of its phonetic quality. The vocal tract system is dependent on the position of the

articulators and a direct translation is possible, at least when dealing with idealized vocal tract models (…).” (Fant 1970:17)

Fant (1970) proposed modeling the acoustic effects of speech production by illustrating the vocal tract as a set of tubes, and it is possible to calculate the resonant frequencies of this tube from its length. Following this reasoning, Johnson (1997: 104) demonstrates vocal tract modeling and a calculation of its length. According to the author vocal tract configuration producing [ɑ] requires a two-tube model. As shown in Figure 2, the back tube has a cross-sectional area $A_{b,}$ while the smaller front tube cross-sectional area A_f.

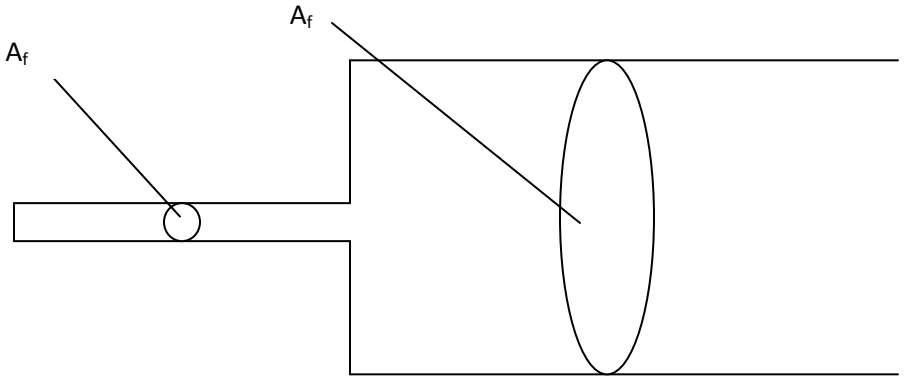


Fig. 2: A two-tube model of the vocal tract for [ɑ] (Johnson 1997: 104).

Following Johnson (1997), the back tube is closed at the glottis and open to the front tube, while the front tube is closed at the junction with the back tube and open at the other end, corresponding to the lips. Since both tubes are closed at one end and open at the other, Johnson (1997) proposes a vocal tract resonance formula (Eg. (1)) to calculate the resonances of the front and back tubes, where *n* is the formant (number of the resonance), *c* is the speed of the sound in the warm, humid air (35,000cm/sec), and *L* is the length of the tube in cm,

(1) $F_n = (2n-1)c / 4L$

Figure 3 shows the resonant frequencies produced by this model, including when the front and back cavities have different lengths. This shows very high resonant frequencies in the back, short cavity and the lowest frequencies in the front cavity. As stated by Johnson (1997:104) "when the back cavity is a little over 4cm long, its lowest resonance is lower than the second resonance of the front cavity. So when the cavity is between 4 and 8cm, the lowest resonance of the tube model (F1) is a resonance of the front cavity, while the second resonance (F2) is a resonance of the back cavity."

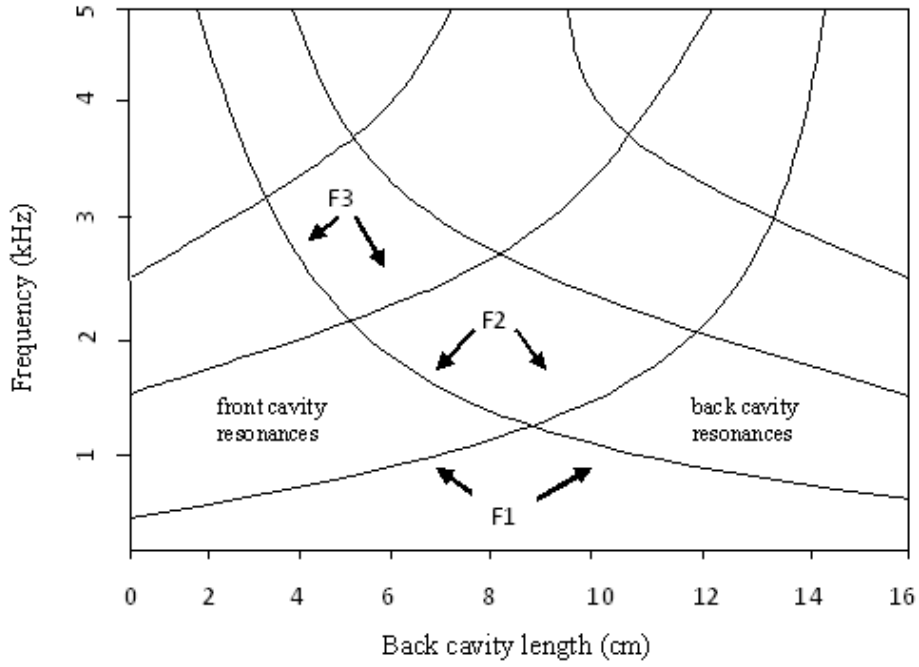


Fig. 3: Resonant frequencies of the back tube and front tube for the tube model shown in the Fig.2. (Johnson 1997: 104).

Just as he does with vowels, Johnson (1997) proposes tube models for the nasals and the laterals. Nasals are formed as a uniform tube with closing at the glottis and the vocal tract and opening at the nostrils, as in the Figure 4, which illustrates the vocal tract configuration and its tube model representation. By contrast, the tube model representation for laterals introduces the presence of a side branch which corresponds to the anti-formant in the output spectrum (Fig.5). Johnson (1997)

describes this side channel as "formed by a pocket of air over the tongue, while the outlet channel is formed around one or both sides of the tongue" (Johnson 1997:161).

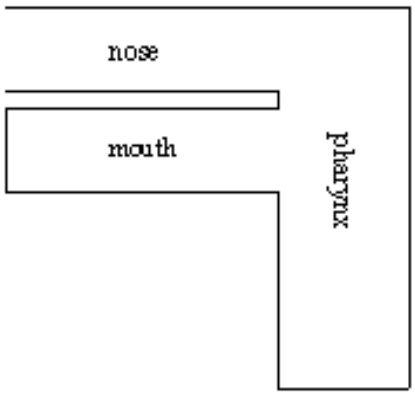


Fig. 4: A tube model of the vocal tract configuration for [m] (Johnson 1997: 154).

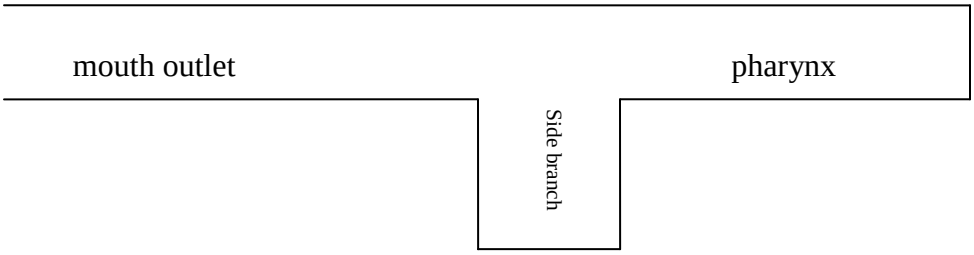


Fig.5: A tube model of the vocal tract configuration for [l] (Johnson 1997: 161).

1.2.2 Articulation of the vocal folds and its quantal effects in acoustics

The Quantal Theory of speech production, as proposed by Stevens (1989: 3-45) as a further development of Fant's (1970) Acoustic Theory of Speech Production, assumes that articulatory-acoustic relations are quantal because while there is a large variation in the acoustic pattern, the articulatory parameter is actually manipulated to a rather small degree. Thus, a small change in the articulatory parameter, produces a large effect in the acoustic dimension. As noticed by Johnson (1997: 82), "the action of the vocal folds provides one of the clearest examples of the

quantal theory of speech". The phenomenon applies also to the voicing qualities (glottis states), showing a nonlinear mapping between the glottal width and its acoustic output. The opening of the glottis during speech varies, from a wide opening (like during deep breath), to a tight closure (such as during a glottal stop). Following Johnsons' analysis (1997), at the beginning of the process when the vocal folds are open, a voiceless sound is produced, until at some point the folds start to vibrate and produce a voiced sound. A closure produces a glottal stop. In this way, the acoustic parameter undergoes significant changes when the articulatory parameter (the closing movement of the vocal folds) passes through the regions that are critical. Those horizontal regions (I and III in Fig. 6) are considered regions of stability in the acoustic-articulatory mapping (Stevens 1989)[5]. Stevens (1989; 2010: 10-19) claims that a complete inventory of the nonlinear mappings between the acoustic and articulatory dimensions expresses the list of distinctive phonetic features in the language. In his hypothesized acoustic-articulatory relation (Fig.6), region I is associated with the minus value for feature F ([-F]); region III, its counterpart [+F]. Because phonation shows the quantal properties described above, natural languages make use of distinctive features within the stable areas of the acoustic-articulatory mapping (voiceless/voice). In this way the Acoustic Theory of Speech Production explains the universality of voicing distinctiveness in natural languages. The two areas of acoustic stability (I and III) and the critical articulatory quantal area demonstrate the distinctive acoustic distributions.

[5] Johnson (1997:83) has suggested that: "A certain amount of articulatory slop can be tolerated, because a whole range of different glottal widths produce practically the same output. In this way, the natural nonlinearity in the mapping from articulation to acoustic output leads to natural classes of speech sounds."

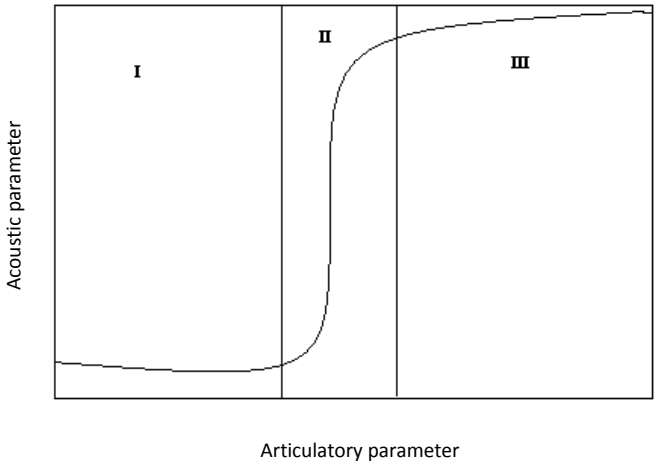


Fig.6: Schematization of the nonlinear mapping between acoustic and articulatory dimensions (Stevens 1989: 357)

1.2.3. Voicing and Optimality Theory

Voicing phenomena have also been analyzed from a universalist, but strictly descriptive phonological perspective, with an application of Optimality Theory as a methodological basis. The authors of OT (Prince & Smolensky 1993: 1-304) assume that forms of language are a result of conflicting universal constraints. In the grammatical conflict between markedness and faithfulness, voicing is said to be prohibited in the obstruents and facilitated by the markedness in sonorants. Faithfulness, on the other hand, disallows voicing assimilation, as it does not fulfill the requirement of identical input-output specification (Prince & Smolensky 1993; Moosmüller & Ringen 2004: 43-61).

In Lombardi's (1995: 89-115) seminal study a set of constraints within Optimality Theory was proposed, accounting for the patterns of obstruent devoicing and voicing assimilation. The author claims that voicing assimilation is always regressive unless additional constraints are active. Proposing supplementary mechanisms to account for a wide range of languages, Lombardi listed a set of 23 constraints which predicts a generalization according to which the

alternation of voicing assimilation in obstruents is restricted in an either morphological or phonological way. The following is a list of the most crucial ones:

"1.*voicing assimilation in obstruent clusters*:

a. With word-final neutralization (for ex. Polish, Dutch, Catalan, Sanskrit)
b. With word-final faithfulness (for ex. Yiddish, Romanian, Serbo-Croatian).

2. *IDentOnset(Laryngeal) (IDOnsLar):*
   Onsets should be faithful to underlying laryngeal specification
3. *IDent(Laryngeal) (IDLar):*
   Consonants should be faithful to underlying laryngeal specification)
4. **Lar: Don't have Laryngeal features*
5. *Agree: Obstruent clusters should agree in voicing*"

Lombardi (1995:2).

In conclusion, the author claims that all obstruents are subject to regressive voicing assimilation. Progressive voicing assimilation alternation, however, is restricted by phonological and morphological limits.

Petrova and Szentgyörgyi (2004: 87-116) investigated voice assimilation of /v/ in Hungarian and Russian with regard to OT. Their analysis of phonetically ambivalent /v/ behavior accounts for voice assimilation in Russian and Hungarian, as well as the ambivalent sonorancy behavior of /v/ and sonorant transparency in Russian within the Optimality Theory constraint hierarchy. The authors employ Sonorant Default (Rubach 1997), which requires that all and only syllabified sonorants are specified for voice, serving as an explanation for the sonorant transparency. According to their description, Russian word-initial sonorants followed by an obstruent permit assimilation if preceded by a cliticized prefix ending in an obstruent, like in i[s#mt^{s}]enska 'out of Mtensk' and in cf. i[z#o]kna 'out of the window', where the final clitic voiced obstruent (in the former example) becomes voiceless under the influence of the voiceless obstruent in the onset of the word-initial syllable, (as in the latter example), despite of the intervening sonorant nasal (Petrova & Szentgyörgyi 2004). On the other hand, word-final

sonorants in Russian are said to be non-transparent and thus not affected by the devoicing from the preceding obstruent, like in ze[zl] ## , *ze[sl] 'staff' nom.sg. or ze[zl#t]o, *ze[sl#t]o 'staff' emph.sg. (Petrova & Szentgyörgyi 2004). Due to the differences in syllabification patterning in Russian sonorants[6], the authors adopt Rubach's (1997) Sonorant Default rule, which states that "all and only syllabified sonorants are specified for voicing" (Rubach 1997: 302). Finally, Petrova & Szentgyörgyi (2004) state that the sonoracy patterns of /v/ result from a sonorant's faithfulness in sonoracy and a restriction that /v/ is a sonorant before a syllabified sonorant.

Further studies by Kallestinova (2004: 117-143) describe the analysis of voice assimilation processes in Turkish stops within the OT framework. By analyzing a set of universal markedness constraints on voicing, the author claims that the three-way contrast in Turkish stops (voiced vs. voiceless aspirated vs. voiceless unaspirated) explains the low ranking of the features [spread glottis] and [voice] (Fig. 7).

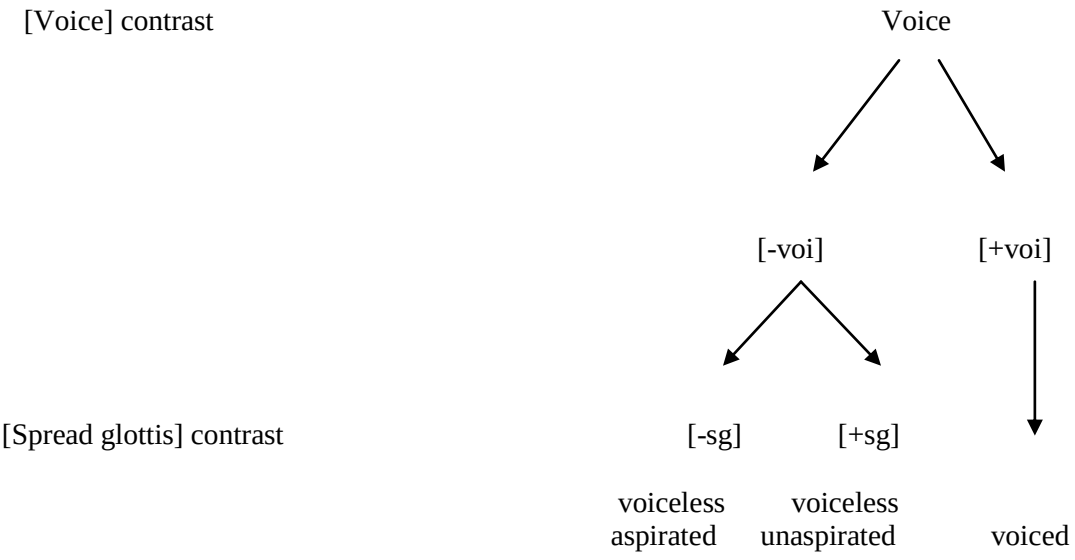


Fig.7: Three-way voicing contrast in Turkish stops (Kallestinova 2004:132)

[6] Transparent sonorants are said to be unsyllabified; non-transparent ones, syllabified (Petrova & Szentgyörgyi 2004).

Kallestinova also reviews phonological licensing of stops in word-initial and final positions, and in stop clusters. Showing the relevance of the syntagmatic and paradygmatic contexts based on syllable structure, Kallestinova (2004) provides a final ranking of the constraints presented in figure 8.

In conclusion, the constraints proposed by Kallestinova (2004) assume that (1) voiced obstruents and [spread glottis] segments are prohibited; (2) input-output segments have specifications for all features; (3) voiced spread glottis stops are prohibited; (4) obstruents in clusters must share voice specifications; and (5) obstruents are voiced between vowels. Such a set of descriptive contrasts as postulated in OT is claimed to be universal, but their ranking is language-specific.

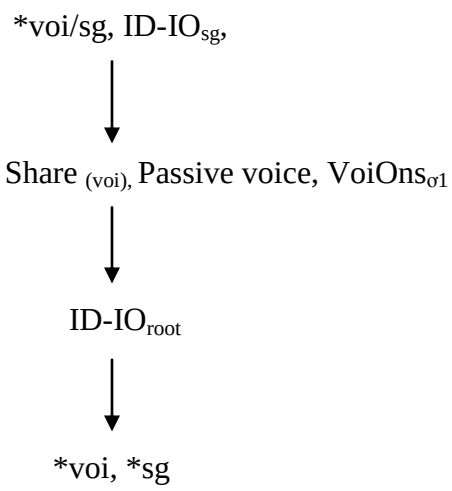


Fig. 8: Ranking of the constraints applying to voicing patterns in Turkish stops (Kallestinova 2004:192).

1.2.4. Features [voice] and [spread glottis]

Features are a necessary ingredient of phonological analyses and should be defined and motivated in phonetic (articulatory, acoustic and auditory) terms. Classical feature systems have been

introduced by Jakobson and Halle (1956: 1-88) and Chomsky and Halle (1968). The latter study relates to the "physical" properties of the features specific for languages.

> "The total set of features is identical with the set of phonetic properties that can in principle be controlled in speech; they represent phonetic capabilities of man and, we would assume, are therefore the same for all languages." Chomsky and Halle (1968: 298)

Laryngeal features of consonants are used to define phonetic dimensions such as voicing or aspiration. In Sound Pattern of English voicing was defined with the features [voice], [tense], [heightened subglottal pressure], [aspirated] and [glottal constriction]. However, as noted by Keating (1988: 275-292), the SPE proposal have never been widely accepted. Despite introducing an innovative glottal configuration, it lacked information about the results of this configuration. Halle and Stevens (1971: 1-46), on the other hand, explain [+voice] in obstruents as a result of slack vocal folds, contrasting it with the stiff vocal folds of voiceless obstruents (thus the renaming of the feature [voice] as [slack/stiff vocal cords]). These features describe the position and state of the vocal folds at the moment of their release in the segment and characterize aspects of laryngeal distinction, for example airstream mechanisms, phonation types, aspiration, voicing and fundamental frequency (Keating 1988). Halle and Stevens (1971) have also related the feature [voice] to tone. Keating (1988: 139) noted also that "stiff vocal cords raise f0 on a sonorant while slack vocal folds lower it; thus low tone is represented by the combination [-stiff, +slack], mid tone by [-stiff,-slack] and high tone by [+stiff, -slack]". According to Keating (1988), what is problematic in this system is that sonorants affect f0 and tones just as much as the obstruents do, especially in voiced/voiceless pairs.

Acoustic characteristics of voicing presented by Jakobson and Halle (1956: 1-108) define voicing by the presence of a low-frequency component – the so-called voice bar, and the periodicity in the spectrum as a result of vocal fold vibration, which can be measurable not only

in acoustic terms but also using phonetic and articulatory methods. The authors use the feature [tense] to distinguish aspirated from unaspirated stops in Germanic languages.

An auditory investigation of the feature [voice] has been conducted, among others, by Kingston and Diehl (1994: 419-454). In their work on automatic (phonetic implementation as a form of overlearned, automatic process) and controlled (characteristic of fluent mature speaking and listening as a product of controlled and well-practiced behaviors) phonetics, the authors present their views on the feature [voice] with regard to English, Swedish, German, Icelandic and Dutch. While the first three languages contrast voiceless aspirated stops with initially unaspirated or prevoiced stops and voiceless unaspirated stops with intervocalically prevoiced ones, Icelandic contrasts voiceless aspirated stops with an unaspirated set that is never voiced during the closure. Finally, Dutch contrasts voiceless unaspirated stops with regularly prevoiced stops (Kingston & Diehl 1994). The question that the authors attempt to answer is whether all the languages under investigation contrast in the feature [voice] or whether it only applies to Dutch stops, since in Icelandic the distinctive feature is [spread glottis] and in English, German and Swedish there might be a third laryngeal feature contrasting them. Their research provides evidence to back up the conclusion that all these languages contrast in the same distinctive feature [voice]. In addition, in all five of them voicing begins earlier relative to the stop release and F0 is consistently depressed in vowels next to [+voiced] stops, regardless of prevoicing or the short lag employed by the language as the realization of the phonation type. It is claimed that f0 values vary only along the [voice] contrast and not just with regard to the presence or absence of phonetic voicing.

In her 1991 study on laryngeal features, Lombardi introduces the features [voice] for the voice vs. voiceless opposition of the obstruents; [glottalization] for the feature [constricted

glottis] for implosives, ejectives and laryngeal sounds; [aspiration] to correspond to the voiceless aspirated consonants; and the feature [spread glottis]. In her description, laryngeal features are concentrated by the Laryngeal node. This is justified by evidence from languages in which more than one laryngeal feature is distinctive.[7]

The features [voice] and [tense] have been widely described by Jessen (1998, 2000). In his investigation on German obstruents, Jessen argues that German does not employ the feature [voice] as a distinctive feature in the stop consonant system, but it does employ the feature [tense]. The results of his experiments (Jessen 1998: 1-347) show that the duration of aspiration is the correlate of the feature [tense]. This occurs in most of the tense/lax opposition contexts in German. In his later studies, Jessen (2000: 11-64) described the feature [voice] and proposed a model of the range of the acoustic/audiotory correlates of [tense] and [voice]. In the model (Fig.10) a distinction between basic and non-basic correlates is made. Basic correlates are those with particularly high contextual stability (meaning that the relevant distinction in a language is described by the correlate which is considered basic) and perceptual salience (meaning that this correlate after manipulation in speech perception experiments leads to categorical perception of the feature). In contrast to basic correlates, the non-basic ones are those that do not have perceptual salience and have less contextual stability. Their function is to support/replace the basic correlates in cases when they are weak or unavailable.

[7] For more discussion on Lombardi's [voice] licensing and sonorants' devoicing, see chapter 5.

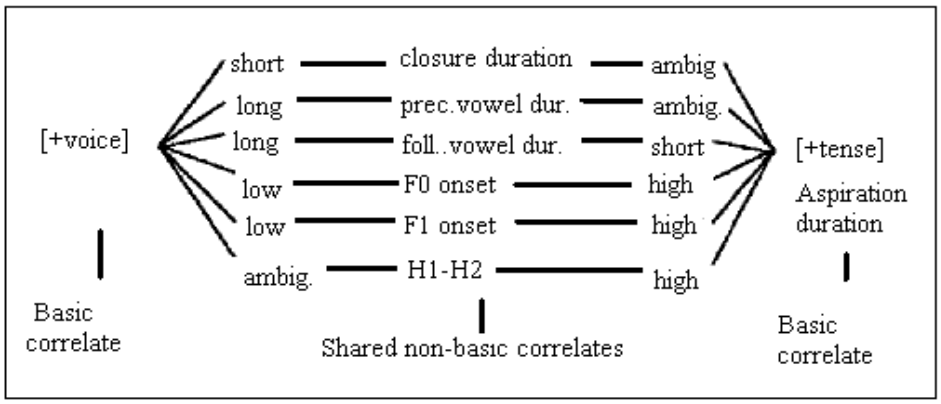


Fig.10: Model of the range of acoustic/auditory correlates of [tense] and [voice] (Jessen 2000:19).

Jessen (2000) specifies the feature [tense] in terms of duration of aspiration as its basic correlate, and the basic correlate of the feature [voice] in terms of closure voicing. The author suggests that for specifying the contextual stability criterion, [tense] should be employed in the Germanic languages for the representation of /b,d,g/ vs. /p,t,k/. He argues that the synthetic manipulation of the voice onset time values into positive ones for the duration of aspiration leads to categorical perception. Basically he claims that [tense] is more important perceptually for aspiration duration than other durational correlates. Jessen (2000) also states, that there is a contextual stability of closure voicing and that full categorical perception can be achieved due to the manipulation of voicing duration in [voice]. He defines the low-frequency property suggested by Kingston and Diehl (1994) as the denominator of the feature [voice], while duration is the denominator of the feature [tense]. Furthermore, within this study Jessen (2000) created a new way of classifying for consonant inventories by proposing acoustic parameters like closure voicing, fundamental frequency onset, preceding vowel duration, following vowel duration, first formant onset, aspiration duration and the difference between amplitude values of the first and second harmonics.

The feature [spread glottis] was first proposed by Halle and Stevens (1971) as a binary feature used to distinguish aspirated from unaspirated stops. It is now considered a part of the set

of universal features describing large, glottal opening. It has been assumed that this feature is not distinctive for obstruents in English. Therefore it does not appear in the phonological representation specifying any given English obstruent. In later studies (Iverson & Salmons 1995: 369-396), the feature [spread glottis] was used for the analysis of fricatives. The authors claimed that it allows a unified treatment of stop deaspiration after fricatives and sonorant devoicing after stops and fricatives. The authors also state that there is a close correlation between glottal opening duration, aspiration and sonorant devoicing.

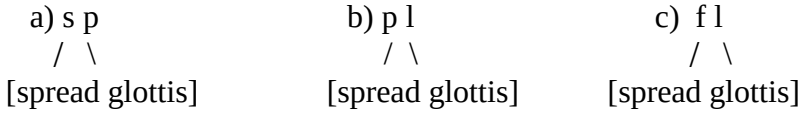


Fig. 11: Relation between sonorant devoicing, glottal opening, duration and aspiration (Iverson & Salomns 1995).

Using examples of words like shrimp [ʃ̥ɹɪmp] or fleet [f̥lit]), it has been demonstrated that the feature [spread glottis] is shared in the syllable onset between the obstruent and the sonorant (11b and 11c) as well as /s/ and a stop (11a), and that the aspiration equates with voicelessness in the sense that both phenomena are the realizations of an open glottis. However, Iverson and Salmons (1995) claim that for languages where voiceless stops are uniformly unaspirated and where [spread glottis] plays neither a phonemic nor a phonetic role (like Polish and French), sonorants remain voiced after initial voiceless obstruents. The issue will be further discussed in detail in the following chapters.

Beckman, Ringen and Jessen (2009: 231-268) investigated the features [voice] and [spread glottis] as contrast features in German fricatives. In their experiment 32 native speakers of Standard German were recorded reading a list of 75 sentences containing the contexts under investigation. Measurements consisted of factors such as the beginning and end of a fricative, end

of voicing, voicing duration and voicing percentage defined with regard to the total duration of the fricative. The tokens were classified into those produced with syllabic and those produced with non-syllabic sonorants. Furthermore, voicing percentage was divided into three classes: 100% (fully voiced), higher than or equal to 25%, and lower than 25%[8]. Additionally, for the analysis of fricative-sonorant German clusters Beckman, Ringen and Jessen (2009) used the following constraints[9]:

1. VoiCODA – "voiced obstruents are prohibited in codas" (Beckman, Ringen & Jessen 2009:2)
2. ID-PRESONORANT VOICE – "an obstruent in presonorant position must be faithful to the input specification for [voice]" (Beckman, Ringen & Jessen 2009:2).

The positional faithfulness of voiced fricatives in word-final position is further said by the authors to be motivated by two additional constraints:

3. ID-PRESON-f – "an input fricative and its output correspondent must have the same specifications for [voice] in pre-sonorant position" (Beckman, Ringen & Jessen 2009:5).[10]
4. FRIC-SG – "fricatives are [spread]" (Beckman, Ringen & Jessen 2009:5). [11]

The claim was that the interaction of these constraints results in the wide range of behavior types of German fricatives. The authors observe two tendencies in their results, one being variation of the syllabicity of a liquid in fricative-sonorant clusters, and the other being variation in the duration and percentage of voicing in voiced fricatives. The first variation, (consider 'grus.lig' and 'gru.s̩l.lig') is decribed as a result of the Optimality Theory (1) **PEAK/LIQUID* constraint – no liquid in syllable peak (Prince & Smolensky 1993), and (2) **ZL/ZR* – no [zl] or [zr] clusters in coda position. The second variation is said to be due to difficulties in producing voiced

[8] The authors claim that the 25% boundary was chosen because not many predicted voiceless fricatives have more than 25% voicing.
[9] The authors credit the first constraint to the work of Ito & Mester (1998), and the second one to Steriade (1997), Padgett (1995), Lombardi (1995), Beckman (1998) and Petrova et al. (2000, 2006).
[10] Same references as for the first and second constraint.
[11] The authors refer to the work of Vaux (1998).

fricatives. It is concluded that German voiced fricatives retain their voicing when followed by a sonorant segment, regardless of their syllabification, as a result of the positional faithfulness. This, however, tends to be problematic for the coda devoicing constraint, due to the above-mentioned sonorant syllabicity and fricative voicing variation (Beckman, Ringen & Jessen 2009).

In her study of phonological features, Keating (1984: 286-319) proposes improvements to the SPE model (Chomsky & Halle 1968), where the binary phonological features are implemented as categories chosen from a fixed and universally specified set, consisting of three categories: fully voiced, voiceless unaspirated and voiceless aspirated stop consonants. Despite the correspondence to the standard VOT divisions, the new terms are viewed as abstract categories which include a number of acoustic correlates and articulatory mechanisms:

"The occurrence of a phonological rule in languages, should not depend on, or be correlated with, the phonetic details of the language (…). A distinction between phonological and phonetic category levels of representation offers an important advantage in describing phonological rules. In a system like SPE, which equates phonological with phonetic representation, rules that occur across languages will look different in each language, depending on the phonetics. In this system, which distinguishes the two levels, these results will look the same regardless of the phonetics. Thus, if rules affect voicing recur consistently across languages, but differ in their phonetic categories, there is an evidence in favor of distinguishing phonological from phonetic representation. The generalization that certain rules occur across languages will be missed of phonological rules apply to phonetic features which are different across languages, but it will be expressed if such rules apply to phonological features that are similar across languages" (Keating, 1984: 292).

Following this assumption, in the chapters to follow I will apply the feature [spread glottis] to American English and German voicing profiles and the feature [voice] to French and Polish voicing.

1.2.5. Voice Onset Time

While numerous studies have employed the Voice Onset Time factor cross-linguistically (e.g. Keating 1984; Ladefoged & Maddieson 1996; Shimizu 1990; Poon & Mateer 1985), the notion itself was first addressed by Lisker and Abramson (1964: 384-422). Their research conducted on word-initial stops showed that the best physiological dimension in phonated speech of these segments is the measurement of the voice pulses relative to the stop release. In recognition of this, the authors named this dimension voice onset time and showed the applicability of the following categories to cross-linguistic examples from several languages (among them: English, Dutch, Hungarian, Spanish, Tamil, Cantonese): 'fully voiced' (typically negative VOT), 'voiceless unaspirated' (with VOT around zero) and 'voiceless aspirated' (with long positive VOT). In later studies, Abramson and Lisker (1967) carried out psychoacoustic experiments with speech and speech-like stimuli and found that the ability of listeners to detect differences between stimuli depends on the phonetic categories. Application of voice onset time has been crucial for this distinction, as VOT values are used in many languages to distinguish either two or three stop categories (Abramson & Lisker 1967). Despite this new dimension proposed by Lisker and Abramson (1964), Chomsky and Halle (1968) maintained binary features rather than a scalar division at the phonological level, proposing a set of features which incorporate phonetic descriptions. According to Ladefoged (1971: 20) there are five values of voice onset for stops and fricatives (Fig.12): (1) fully voiced, (2) partly voiced, (3) voiceless slightly aspirated, (4) voiceless unaspirated and (5) voiceless aspirated. The divisions presented in the figure 12 are said to be arbitrary, as the VOT feature is a continuum. Ladefoged (1971) claims that no language contrasts in more than three points on this scale. The author also describes basic differences between voiced-voiceless distinctions, as well as aspirated-unaspirated, pointing out

that there are languages whose sounds cannot be characterized simply in terms of two states of the vocal cords – voiced and voiceless. For example, in Gujarati there is a distinction in ordinary informal speech between two sets of vowels, and in both sets the vocal cords are vibrating. Ladefoged (1971) proposes that it is due to two kinds of phonation – voice and murmur.

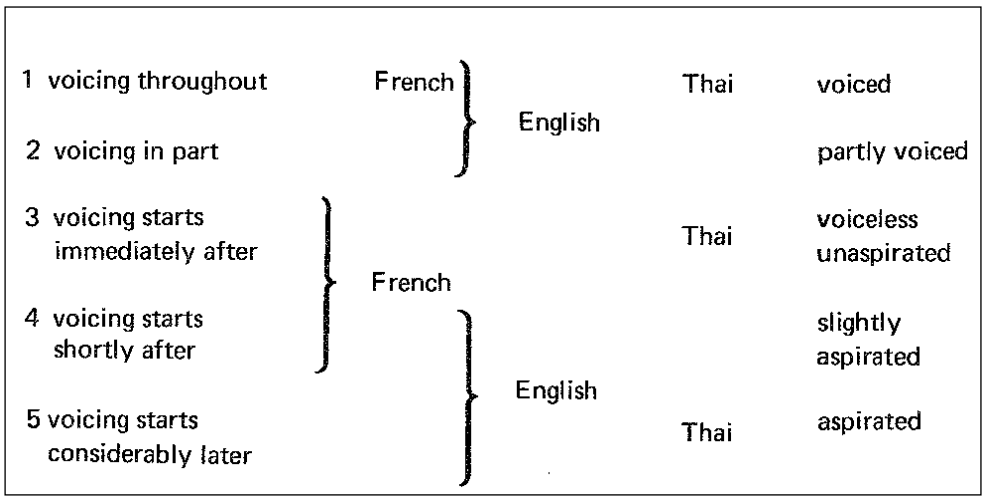


Fig.12: The feature voice onset on the example of English, French and Thai stops and fricatives (Ladefoged 1971: 20).

A study by Keating (1980: 1-242) conducted on Polish provides another justification for this dimension's usage for [voice] contrasts in Polish, compared with, for example, burst parameters. In her investigation, VOT values of apical stops /t/ and /d/ in a non-cluster environment were measured. Perception and production tests were conducted on 24 native speakers of Polish. The author recorded minimal pairs containing various stops in question, as well as words embedded in phrases. Voice onset time was measured from the beginning of the release burst to the beginning of voicing. Results for the post-pausal /t/ and /d/ have shown that, despite the lack of stability in VOT values across speakers, the combined values for 24 speakers yield a normal distribution of each phonemic category. Results reported for the /t/ and /d/ occurrence in running speech have revealed that speech context does affect voice onset.

Moreover, it has been pointed out that Polish has very little category overlap compared to English, where voiced and voiceless stops merge in casual speech.

The voice onset dimension has been investigated also with Nepali stops (Poon & Mateer 1985: 39-47). VOT distributions were investigated on samples of 720 CVC words from ten adult male Nepali speakers. The results showed that only voice lead, short-lag and long-lag stops could be analyzed in terms of VOT values.

Similarly, VOT measurements were successful in research involving Asian languages. Studies conducted by Shimizu (1989: 1-13) on Japanese, Burmese, Thai, Korean and Hindi show that Japanese reveals large a variability for VOT values in each of the three categories. Burmese and Thai voiceless unaspirated stops show a restricted amount of voice onset variation. Moreover, Burmese aspirated stops, however, exhibit a slight delay in voicing when compared to Korean and Hindi. In conclusion, Shimizu advocated the necessity of postulating an additional category in order to differentiate between the voicing categories of Korean and Hindi.

Voice onset measurements cover various laryngeal and supralaryngeal events associated with the timing relation between the release of a stop consonant occlusion and the onset of vocal-fold vibration. In acoustic practice, this is the time between the release burst and the first periodicity in the acoustic signal. Values for voice onset time refer to the moment of the stop release, which is a reference point in time (0 msec) and its values are measured in relation to that point: VOT is negative when the onset of the voice occurs before the stop release, but it is positive when the voice onset occurs after the stop release. 0 msec VOT is when onset of the voice occurs at the same time as the stop release. However, as has been suggested by Keating (1984), this measurement causes difficulties for stops in positions that are not word-initial. For example, if a stop follows a sonorant, then the voicing of the sonorant and the stop closure will be

continuous and the VOT measurement will not be possible, since the voicing will be already in progress. Similarly, for a voiced stop in word-final position or before another stop which might not be released, the amount of voicing during the closure should be measured.

While VOT measurements have been successfully investigated in stop consonants, a more robust method of temporal voicing analysis has been proposed by Möbius (2004: 5-26) to investigate other consonant classes like sonorants and fricatives. The method will be described in detail in Chapter 3.

1.3 Contrasting features in Polish, French, German and English sonorants

Sonorant consonants are produced when the air passage between the glottis and the output of the vocal tract has a constriction greater than during vowel production but not to such an extent as to allow turbulence at the point of constriction (Stevens 1989; Clark & Yallop 2007). As noted by Stevens (1989) the vibration of the vocal folds under these conditions does not change with the presence of an adjacent vowel, and shows very little diversity in the first and the second harmonic as the vocal tract undergoes the shift from the sonorant to the vowel. This in turn suggests that there is continuity in the low-frequency energy amplitude between a sonorant consonant and a vowel. In their study on the sounds of world's languages, Ladefoged and Maddieson (1996) described rhotics as segments which behave similarly in the phonological sense, occupying privileged places in the syllable cross-linguistically. They observed that rhotics occur very often in onset syllable clusters as the second member and in coda syllables as the last member, and are usually placed close to the nucleus of the syllable (a fact which serves as an explanation for the historical division of English r-sounds within dialects into 'rhotics' and 'non-rhotics', depending

on whether the postvocalic /r/ occurred in the prepausal or preconsonantal positions in the respective pronunciation varieties).

As mentioned by Ladefoged and Maddieson (1996), rhotics form a group with a variety of manners of articulation, including trills, taps, flaps, fricatives, approximants and 'r-coloured' vowels. In their acoustic description of rhotics, the authors suggest a lowered third formant as the common factor for different kinds of segments, mainly the American English /ɹ/. They point out that differences in the location of the formants of the r-sounds serve as important cues as to the location of the segments' constriction. For example, a high third formant is an indicatior of an uvular as well as a dental rhotic location. It is also worth noting that Ladefoged and Maddieson (1996) show the similarities between the spectral properties of trills, flaps and taps. According to their proposal, trills (particularly those in intervocalic positions), tend to be reduced to a single period, the so-called 'one-tap trill', which may sometimes be manifested with a co-occurring frication or trilling in fricative rhotics (like in French). Moreover, the authors notice a tendency of transformation from trills into approximants, which is caused when one or more closures are produced, such that the following opening phase is prolonged. This results in the approximant rather than the further production of shorter openings with a closure. With regard to the lateral approximants, the authors present articulatory properties following their palatographic and x-ray studies conducted on several languages. Those investigations have shown that a dental/alveolar occlusion is "very often limited to a few millimeters on the alveolar ridge in the area behind the incisors and perhaps extending to the premolars. It does not extend back to the molar regions but instead the body of the tongue is relatively low in the mouth behind the closure, permitting lateral air escape as far forward as the front of the palatal region" (Ladefoged & Maddieson 1996: 198). However, the authors also note that the area of constriction may extend further back in the

mouth or that the constriction may be incomplete. Based on the cross-linguistic examples from Ladefoged and Maddieson (1996), production of the voiced lateral approximant should be possible in nine places of articulation, eight of which participate in pairs which are distinguished between the features apical and laminal, being independent of other features.

The acoustical properties described in the same study (Ladefoged & Maddieson 1996) characterize these segments by the presence of "well-defined, formant-like resonances". The first formant is usually placed low in frequency, while the second formant may occur anywhere within a wide range in the center frequency. The third formant is then relatively strong in amplitude and has a high frequency. If a vowel is adjacent to the lateral segment, an abrupt change in formant placement may occur, which can be observed in both situations - when the medial closure for the lateral is formed and released (especially in apical articulation; in laminal and dorsal articulation, the transitions may be slower due to the adjacent vowel). Although the most common laterals are voiced approximants, it has been noted that they may differ with regard to phonation type as well. Thus, apart from voiced exemplars, there are cases of voiceless, breathy voiced and laryngealized occurrences as well. Voiceless laterals occur as contrastive segments in languages like Burmese, Tibetan and some Irish dialects. In other languages, voiceless laterals may also contrast with their voiced counterparts at different places of articulation.

Based on experimental results (Sieczkowska et al. 2009; Sieczkowska, Möbius & Dogil 2010) this dissertation will focus on single- and obstruent-cluster-occurring sonorants in Polish, French, German and American English, particularly the liquids /r/ and /l/, which show the biggest tendency of devoicing. Their distribution and feature contrasts differ depending on the language under investigation. These contrasts will be described in the following sections.

1.3.1. Polish sonorants

The sonorants occurring in Polish are /r, l, w, m, n, ɲ, j/. While most of them are considered to be voiced, both as single segments and in clusters (Gussmann 2007: 29-56), this dissertation will focus on the liquids /l/ and /r/, which have a devoicing tendency in certain phonological and phonotactic contexts (obstruent-sonorant clusters).

Polish [r] is articulated by producing a trill that is formed by an egressive pulmonic airstream, which causes vibration of articulators resulting in multiple, brief and fast touching of the tongue tip against the alveolum and its release (Dukiewicz 1995: 40). This movement can occur as much as 25, 30 or even 40 times per second. Both liquids voiced [r] and voiceless [r̥] are articulated at the same place in the oral cavity, although it has been observed that in fast speech there exists an allophone of [r] which is no longer a trill but resembles more a tap due to a single touching of the articulators (Dukiewicz 1995). Ladefoged and Maddieson (1997: 217) describe the production of a trill as "the vibration of one speech organ against another driven by the aerodynamic conditions, where one of the moveable parts of the vocal tract is placed close enough to another surface, so that when a current of air of the right strength passes through the aperture created by this configuration, a repeating pattern of closing and opening of the flow channel occurs". The authors compare the trilling action to the vibration of the vocal folds, where there is a certain degree of variation from trilled to non-trilled production. Sonorants [l] and [l̥] are both articulated as laterals, forming a constriction between the tongue tip and alveolum, which disallows the airflow through the central part of the oral cavity (Dukiewicz 1995).

Due to Gussmann's description, Polish sonorants are usually considered voiced "since the cavity configuration in their production encourages or facilitates spontaneous voicing" (Gussman 2007:295). Spontaneous voicing refers to the narrowing of the air passage to the point where the

rate of flow is reduced below the critical value needed for the Bernoulli effect to take place (Chomsky and Halle 1968), which can be formally expressed as the description of the laryngealy unmarked sonorants with the glottal tension dimension (Iverson 2007). Polish sonorants are voiced except for a position between voiceless consonants or after a voiceless obstruent before a pause. Consider the list below (Gussmann 1997: 295).

krwi [krfji] 'blood, gen. sg.'
kadr [katr̥] 'frame'
narośl [narɔɔ̯l] 'growth'
baśń [baɕn̥] 'fairy tale'
rytm [rɨtm] 'rhythm'
fanatyzm [fanatɨs(m̥)] 'fanaticism'
piosnka [pjɔsn̥ka] 'song, dim.'
jabłko [jap(w̥)ko] 'apple'
wiatr [vjatr̥] 'wind')
bóbr [bupr̥] 'beaver'
módl [mut(l̥)] 'prey, imp.'
bojaźń [bɔjaʑn̥] 'fear'
kosmka [kɔsm̥ka] 'villus, gen. sg.'
wydm [vɨdm̥] 'dune, gen.pl'
mielizn [mjɛlisn̥] 'shoal, gen. pl.'
jadł [jat(w̥)] 'he ate'

Fig.13: Examples of Polish sonorants' devoicing (Gussmann 2007:295)

Devoicing of sonorants can also depend on the dialect or socio-phonetic features. For example, if the obstruent which precedes a sonorant preserves its voicing, the word final sonorant is voiced as well. In some cases a devoiced cluster might sound unnatural (like in wydm [vɨdm̥] *dune, gen. pl*). Despite that, however, there are variants of sonorants that can only be devoiced, e.g. the ones in derivatives *–izm, -yzm* (marks-izm [markɕism̥] *Marxism*) (Gussmann 2007). Gussmann notes that devoicing might vary from speaker to speaker, depending on both the carefulness and the variety of his or her speech. This applies to the sonorant as well as to the preceding obstruent, which in emphasized pronunciation might preserve its voicing, like in [kadr], [bubr] or [kadm]. As has been discussed by Gussmann (1997) and Rubach (1996: 69-100), voicing assignment in Polish sonorants depends on their phonological licensing, including

extrametricality, segment-government and empty nucleus. The phonetic manifestation of these obstruent phonological analyses, however, has bever been investigated systematically.

The acoustical study of voicing in Polish sonorants presented in this dissertation is based on the investigation of the feature [voice], which has neen analyzed using the results of automatic voicing detection and the calculation of fundamental frequency values along with the occurrence of periodicity using a component of the ESPS get_F0 tool, as well as the German Festival (2009) application. The articulatory investigation was completed using 2D electromagnetic midsaggital articulography (EMA).

1.3.2. French sonorants

While the French lateral sonorants have large speaker-variation in the manner of articulation (Ladefoged & Maddieson 1996), the most common way they are articulated has been described as a voiced approximant, with a strong tendency towards being an apical alveolar sound. Delattre (1971: 129-155) investigated the French and German rhotic sound [R] by conducting x-ray studies on several speakers of these two languages. The author reported that uvular trills are produced with an initial backward tongue root movement, which is followed by an upward movement towards the uvula. The uvula is in turn moved forward, enabling the occurrence of trilling (Fig. 14).

According to Tranel (1987: 1-252), there are numerous variations of the liquid /r/ in French. He distinguishes two groups with regard to their manner of articulation: (1) the occlusive type, produced by a closure in the oral cavity caused by articulators touching (for example tongue tip against the alveolar ridge) and (2) the constrictive type, a result of production involving some kind of constriction in the oral-pharyngeal cavity, which may but need not occur with frication

noise. Similarly, Meunier, in her 2007 (11-29) study, differentiates between two possible realizations of the liquid /r/, one being the trill /R/ (articulated relatively rarely according to the author), and the second one being its devoiced counterpart, the fricative /ʁ/.

Tranel (1987) presents a comparison between the syllabicity of English and French liquids. While English /l/ and /r/ are syllabic at the end of the word in consonant cluster, like in the words 'table' [tɛjbl̩] and 'sugar' [ʃʊr̩], in French, word-final consonants are never syllabic, for example 'table' [tabl] and 'sucre' [sykr̥]. Thus, as will also be mentioned in the following section, English liquids play the role of a vowel and constitute the syllable nucleus. Alternatively they can behave as a consonant and be part of the margin of the syllable. French liquids, by contrast, play only a consonantal role. This is shown in Figure 15.

In his acoustical studies concerning French sonorants, Chafcouloff (1980: 7-56) describes the characteristic of spectral properties of /r/. The author demonstrates that it is the distance between the second and third formant (particularly variation in F3 placement) which distinguishes the liquid from other sonorants. According to his studies, the distances between F1 and F2 are smaller than between F2 and F3. As an example Chafcouloff (1980) shows a /r/ formant measurement in the intervocalic contexts /i, y, a, u/, as produced by four speakers. The average value taken from all four recordings demonstrates differences of 1075Hz between F1 and F2 in the /i/ context, up to 335Hz in the /u/ context. Similar measurements conducted for the F2 and F3 show differences of 1130Hz for /i/, up to 1300Hz for /u/ context. Moreover, Chafcouloff (1980) and Meunier (1989) notice final devoicing in the French rhotic consonant when it is preceded by a voiceless consonant (like in 'quatre' [katr̥]). This process is reflected in the average

duration[12] of the /r/ segment, which appears to be much longer than comparable word final sonorants: /j/ - 64.6ms; /l/ 89.2ms; /r/ 160,6ms (Chafcouloff, 1980).

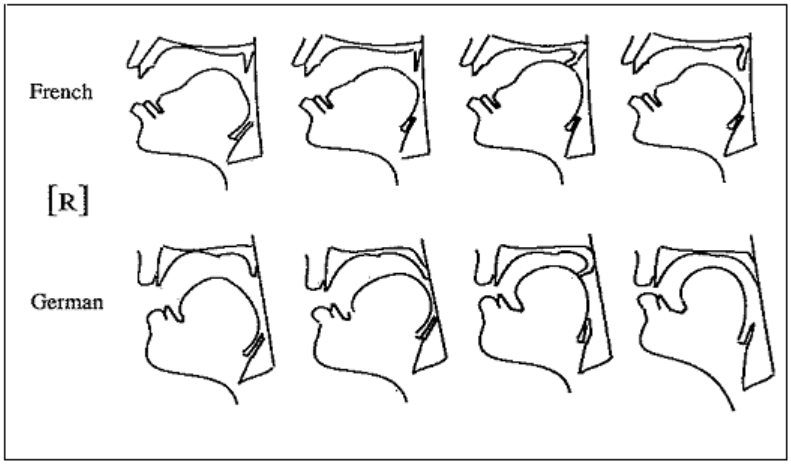


Fig.14: X-ray films of intervocalic French and German uvular trills. Second frame of rows shows tongue retraction, third frames show backward movement of the tongue and raising of its body in order to front the uvula (Delattre 1971; Ladefoged & Maddieson 1996: 229)

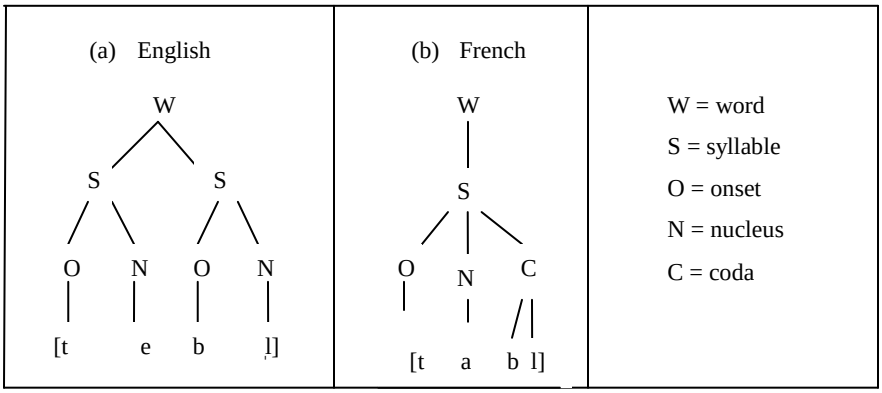


Fig.15: The syllabicity of liquids in English and French (Tranel 1987:135).

1.3.3 German sonorants

According to Wiese (1996: 170), German /R/ is most commonly articulated as a uvular sound. In non-standard varieties, however, like in the Bavarian dialect, other places of articulation can occur. The amount of constriction differs both within and across dialects ranging from fricative to

[12] Average values were determined by the recordings of four speakers.

vocalic, but the r-sound occurring in Modern Standard German can best be described as an approximant. The author claims that /R/ in postvocalic position becomes vocalized to a large extent, especially in northern dialects and in the standard variety, where it tends to be totally vocalized (/ɐ/). As has been pointed out by Wiese (1996), in the case of the postvocalic position and complete vocalization of the sonorant, it often occurs that the /R/ merges with the preceding vowel in a word-final sequence, for example: Honorar [hono:ʁaɐ̯] vs. hurra [huʁa:]. The fricative variant /ʁ/ is said to be found in initial position, with more than just fricative realizations in certain dialects of the Lower Rhine.

	l	R	n	m	s	v
p	+	+	(+)	-	+	-
t	-	+	-	-	-	(+)
k	+	+	+	(+)	(+)	+
b	+	+	-	-	-	-
d	-	+	-	-	-	-
g	+	+	+	(+)	-	-
f	+	+	-	-	-	-
v	(+)	+	-	-	-	
ts	-	-	-	-	-	-
pf	+	+	-	-	-	-
ʃ	+	+	+	+	-	+

	R	l	m	n	f	s	ʃ	ç	p	t	k
R	-	+	+	+	+	+	+	+	+	+	+
l	-	-	+	+	+	+	+	+	+	+	+
m	-	-	-	-	+	+	+	-	+	+	-
n	-	-	-	-	+	+	+	+	?	+	?
ŋ	-	-	-	-	-	+	+	-	-	+	+
s	-	-	-	-	+	-	-	-	(+)	+	+
f	-	-	-	-	-	+	-	-	-	+	-
X	-	-	-	-	-	+	-	-	-	+	-
ʃ	-	-	-	-	-	+	-	-	-	+	-
t	-	-	-	-	-	+	+	-	-	-	-
k	-	-	-	-	-	+	-	-	-	+	-
p	-	-	-	-	+	+	+	-	-	+	-

Fig.16: German onset (left) and coda clusters (right) (Wiese 1996).

German sonorants, like Polish ones, are said to have various phonotactic possibilities, particularly in the onset, as shown in Fig. 16. (Wiese 1996: 33-49). In terms of features, Wiese (1996) describes the German r-sound as [+continuant] and [+low] in varieties in which it merges

with the vowel /a/ and receives its place of articulation. The feature [continuant] allows a differentiation between the German sonorants /r/ and /l/, as the latter is produced with a mid-sagittal complete closure, an articulatory feature accounting for [-continuant]. In his study on the phonology of German /R/, Hall (1993: 83-105) proposed a specification of the privative feature [voice], which can only be positive if it is to play a phonological role. Thus the author claims that German sonorants are not marked for [voice]. These results form the implication that there is no phonemic contrast between voiced and voiceless sonorants.

For the sonorant /l/, Wiese (1996) proposes incorporating the feature [lateral], which is an alveolar sonorant articulated by touching the tongue tip against the alveolum and placed lower in the mouth below the front palate than, for example, when articulating the stop /t/ (Ladefoged & Maddieson 1996). This comparison shows also that the pharynx is much more open during the articulation of the lateral than it is during the stop. The jaw opening is also more extended for the articulation of /l/ than for /t/ (this factor facilitates the lateral escape of air) (Fig. 17).

In his study on voicing, Möbius (2004) observed that German sonorants show similar voicing patterns to the voiced fricatives, inheriting their voicing properties from the left-hand segmental context. The author claims that in a sonorant or vocalic context, all sonorants are almost fully voiced, with the exception of voiceless exemplars of obstruent- /R/ clusters. While /l/ voicing rises from voiceless to high degrees of voicing near its end, /R/ on the other hand tends to be fully devoiced if preceded by a voiceless obstruent. The realization of /R/ therefore ranges from an uvular trill, through to a velar and a uvular voiced fricative, to its devoiced variant (which is 'virtually undistinguishable from /x/ when played in isolation' [Möbius 2004:18]). Moreover, the probability of voicing in obstruent-sonorant clusters is said to increase when a

syllable boundary separates the segments, as in Stecknadel [ʃtɛk.na:dəl] ‘pin’ vs. knapp [knap] ‘tight’ (Möbius 2004).

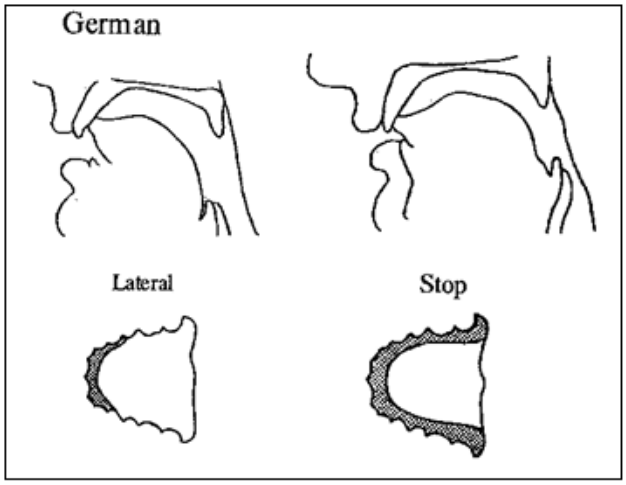


Fig.17: X-ray tracings and palatograms for /l/ and /t/ in German (Ladefoged and Maddieson 1996:184)

1.3.4 English sonorants

The English liquid /l/ is a lateral approximant articulated by forming a constriction between the tongue tip and the alveolum. It may be described as the ‘dark’ [ɬ] (produced with slight velarization) or ‘clear’ [l] (produced with simultaneous palatalization), depending on the phonetic context (Clark and Yallop 2007). ‘Clear [l] occurs in the onset of the syllable, whereas ‘dark’ [ɬ] occurs elsewhere, i.e. in the nucleus or coda position’ (Tranel 1987: 53). Consider the examples in Figure 18.

Clear *l*	Dark *l*
look	cool
leaf	feel
left	felt
lark	Carl
plane	apple

Fig. 18: Examples of clear and dark l in English (Tranel 1987:53).

The articulation of ‘clear’ [l] is in the front part of the tongue dorsum, which is raised towards the hard palate, as during the production of a relatively closed vowel. On the other hand, ‘dark’ [ɬ] production involves raising the back part of the tongue towards the velum, causing the velarization effect.

Sonorant /l/ is voiced in most of the occurrences, with a tendency to devoice (Fig.19) after voiceless obstruents (both stops and fricatives), with the exception of /s/ (Tsuchida et al. 2000: 167-181). It has been observed (Tsuchida et al. 2000) that devoicing of sonorants in English may only be partial and that its extent depends on the manner of articulation and the preceding consonant (there is less devoicing following fricatives than there is following stops).

plea [p̥l] flee [f̥l] spleen [spl]

Fig.19: Example of English sonorant devoicing (Tsuchida et al. 2000: 167)

Espy-Wilson et al. (2000: 344) listed three major ways of articulating the American English r-sound: (1) tip-up retroflex /r/, (2) tip-up bunched /r/ and (3) tip-down bunched /r/. The first type, placed at the alveolar ridge, is formed only by the tongue tip. The second takes place in the palatovelar region and is made solely by the tongue dorsum and the lowered tongue tip. The third type happens in both alveolar and palatovelar regions, produced by raising the tongue tip and the dorsum simultaneously. Its positional placement is said to occupy the syllable nucleus and consonantal positions when classified as a sonorant liquid. Ladefoged and Maddieson (1996: 235) have presented x-ray films of the syllabic variant of the American English /ɹ/ (Fig.20) produced by six speakers, in which it can be seen that the constriction for the so-called ‘bunched r’ is made in the low pharynx and at the mid-palatal region with the absence of tongue tip raising.

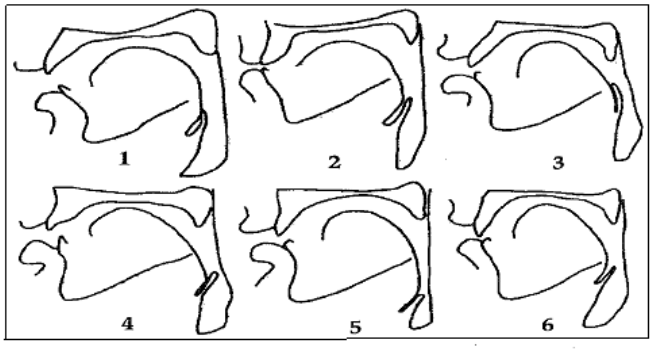


Fig. 20: X-ray films of American English 'bunched r' (Ladefoged & Maddieson 1996:235).

The analysis of voicing properties of English sonorants has been widely presented in the studies concerned with the feature [spread glottis] (Chomsky & Halle 1968; Browman & Goldstein 1986; Ladefoged & Maddieson 1996), which all assume that sonorant devoicing is a result of a widely open glottis that allows a large airflow. The [spread glottis] occurrence is said to be a result of the process of spreading this feature from the adjacent obstruent and sharing it with a following sonorant (Iverson & Salmons 1995: 369-296).

The positional placement of English sonorants has also been described by Kreidler (2004: 90-116) as occurring word initially, medially and finally in free-standing segments as well as in obstruent clusters. Liquids are said to be placed after both voiceless and voiced stops and voiceless fricatives in word initial and medial position. When positioned in word-final clusters, liquids precede obstruents or other sonorants, as in 'harp' [harp] or 'barn' [barn] (Kreidler 2004).

1.4 Exemplar Theory and Specification in Context

Exemplar Theory is a usage-based account of language and its changes, in which experience with a language plays a crucial role in grammar formation (Bybee 2006: 711-733). In the exemplar-theoretic view (Nosofsky 1988: 700-708; Lacerda 1995: 140-174; Pierrehumbert 2001: 137-157),

speech events, which represent various levels of categories like phonemes, syllables and words, are stored in the memory as exemplars (i.e. actually experienced instances of those categories) in a perceptual space. A closely-linked production-perception loop thus operates on the exemplar level by comparing stored events between collections (Wade et al. 2008: 151-152). Percepts contain ample phonetic and phonological information and are grouped together in exemplar clouds. The ones which are similar are placed in closer vicinity, while the dissimilar ones are located farther apart in the perceptual space. Exemplar clouds are said to represent the categories of a given language (Pierrehumbert 2001). In the process of language experience, new categories may emerge after receiving a sufficient number of stimuli (Pierrehumbert 2003: 177-228). If, however, the exemplars in a given category are not reactivated often enough, the categories may change over time or disappear due to memory decay (Goldinger 1997: 251-279). Thus, in the exemplar-theoretic approach, frequency of occurrence and frequency of experience play crucial roles (Pallier et al. 2004: 78-91, Wade et al. 2010: 227-239, Walsh et al. 2007: 481-484, Schweitzer et al. 2009: 728-736).

1.4.1 Exemplar-based models, frequency and context effects

The use of models within the exemplar-theoretic account enables us to provide an excellent explanation for discrete and gradient phenomena like phonetic neutralization and frequency of occurrence matters (Wade et al. 2008; Bybee 2002, 2006; Johnson 1997; Pierrehumbert 2001, 2006).

Johnson (1997: 145-165) proposed a model of speech perception where the dimensions along which exemplars of speech are assessed comprise features pertaining to the speaker's voice and various levels of context rooted in the properties of the auditory system. The model preserves

speaker-specific details in the set of exemplars which are instances of an experienced category. Categorization process involves comparison between new items and already stored ones, which are further constructed as sums of similarity in relation to each category. According to Johnson's (1997) exemplar model, the process of speech perception is based on "association between a set of auditory properties and a set of category labels. The auditory properties are output from the peripheral auditory system, and the set of category labels includes any classification that might be important to the perceiver, and which was available at the time that the exemplar was stored" (Johnson 1997: 147). According to the author, this kind of model results in finding that the already stored exemplars, which are the most similar to the new ("to-be-stored") items, are the ones which were produced either by the same or a similar speaker. Johnson further claims that the model enables categorization of the new exemplars by analyzing speaker-specific information within the existing categories (containing prior exemplars). Thus, the model takes all the speaker-specific information (like gender, age, etc.) into account during the perception process by analyzing acoustical differences in the speech signal.

Wade and his colleagues (2010) have, by contrast, proposed a computational Context Sequence Model (CSM), in which target acoustic patterns are based on previously heard or produced sounds from memory storage. It is assumed that speech events always appear in continuous stretches, in which individual sounds occur in a larger context. This framework claims that the production of exemplars stems from weighing the similarity of the original context in which they occurred, compared with the target production context. The authors defined the left context as recently-produced acoustic information, and the right context as "estimation of what is likely to be produced in the future" (Wade et al. 2010: 229). Thus the matching process involves a comparison of the preceding acoustic exemplar with the current acoustic context and the

following linguistic context. In the course of experiments conducted on a large single-speaker German corpus (Schweitzer et al. 2003: 1321-1324; Schweitzer & Möbius 2004: 459-462), Wade and colleagues (2010) demonstrated that "up to about 1s of surrounding context (0.5s preceding and 0.5s following) was useful in determining the acoustic shapes of phoneme categories" (Wade et al. 2010: 236). Thus, it has been illustrated that the exemplars (i.e. words, segments or features) are stored in the memory as a continuum with the adjacent exemplars with which they originally occurred, including their neighboring and overlapping segments. Moreover, the authors view the production processes as a selection of stored exemplars and probabilistic degradation. Crucially for the production of the entirely stored utterances, their acoustic information is claimed to be stored 'bit-by-bit', i.e. where the new exemplars are specified in the produced context. In so assuming, Wade et al. (2010) have been able to experimentally demonstrate the central role of context during speech production, a process of the token selection on the segment-level placed in their acoustic surroundings. Additionally, it was found that "segments produced as part of more frequent syllables were selected more efficiently and gradually took on context-specific patterns, becoming more variable and more affected by lenition processes than the same segments produced in less frequent contexts" (Wade et al. 2010: 237). Finally, it has also been demonstrated that context selection takes priority over unit selection during speech production.

Frequency effects have also been investigated by Walsh and colleagues (2007: 481-484) who conducted experiments involving syllable duration measurements. In their Syllable Frequency Effects Model, a hypothesis formed by Schweitzer and Möbius (2004) is assumed, which states that high frequency syllables have a significant number of exemplars which act as production targets, whereas low frequency ones have a low number of exemplars and "have to be computed online from exemplars of their constituent segments or segment clusters" (Walsh et al.

2007: 481). The authors' study explains their hypothesis as a computational process, by modeling competition between syllables accessed from their exemplar clouds and syllables accessed as sequences of phoneme-sized units. As a result, Walsh et al. (2007) were able to show that while frequent syllables are accessed as units during speech production, the infrequent ones are produced on-line from exemplars corresponding to their constituent segments.

As a follow-up to this study, Walsh et al. (2010: 537-582) proposed the Multilevel Exemplar Model (MLM), which demonstrates the relationship between exemplars on the constituent level and the unit level. In the phonetic aspect the MLM models syllable frequency effects, positing the argument that syllable duration variability is a function of segment duration variability for infrequent syllables. The authors' view is that frequent syllables are accessed as units, while infrequent syllables are more likely to be produced on-line, i.e. using the available phone-sized exemplars. On the syntactic level, MLM was designed to correctly predict grammaticalisation of *going to* as future tense and no grammaticalisation for other verbs of movement which have maintained their original sense.

Context Sequence Model (Wade et al. 2010) and Multilevel Exemplar Model (Walsh et al. 2010) apply context at various levels (phonological, morphological, syntactic etc.) as a crucial factor for modeling behavior of the exemplars in their cognitive representation. Contextual effects in phonological representations have also been investigated on the prosodic level (Dogil & Möbius 2001: 2737; Möbius & Dogil 2002: 523-526; Schneider et al. 2006: 335-361, Dogil 2010: 343-380). The Incremental Specification in Context (ISC) Model resulting from these studies sees phonetic and phonological speech representations as regions stored in the speaker's perceptual space, where "category-specific exemplars emerge from the internal analysis-by-synthesis process and a successful match to patterns derived from the input speech signal" (Dogil

2010: 356). Conclusions drawn up in this study will also be further analyzed (see chapter 6) in relation to the specification of voicing by various contexts.

1.4.2 Context specification

The results presented in this dissertation find their justification in the voicing specification in context[13]. As will be further discussed in the following chapters, voicing probabilities of Polish, French, German and American English sonorants undergo changes due to contextual variations on the articulatory and consequentially phonetic and phonological levels. Voicing dependencies in Polish and French seem to lie in the changes of the context of phonological licensing, whereas German and American English voicing probabilities depend more on phonetic contextual variation. Moreover, Polish coda devoicing in obstruent-sonorant clusters seems to be influenced by the non-coupling articulatory patterns in this position, which differ when one looks at the word initial C1 and C2 relation, where the consonants in the onset cluster undergo C-center effects by forming leftward and rightward consonant shifts maintaining a stable distance with regard to the vowel target (see Chapter 6). Voicing is universal but it is dependent on many factorial changes (phonological, phonetic and articulatory phenomenon) that it demands an analysis which includes all contextual specifications.

1.5 Summary

In this chapter I have made an attempt to describe the most important and crucial studies concerning voicing as a result of speech production which have directly and indirectly influenced the emergence of the studies described in this dissertation. As has been presented above, the topic

[13] This study is part of the SFB 732 research grant "Incremental Specification in Context" (Alexiadou 2006).

varies cross-linguistically as well as methodologically, and has been widely investigated over the past few decades. This makes new studies all the more challenging. The investigations conducted within this dissertation employ some of the previous assumptions and adopt an automatic framework concerning voicing extraction as described by Möbius (2004). On the basis of his studies, it has been observed that this new method to analyze voicing is very important, particularly for the voicing of sonorant consonants, as the previously conducted investigations lack temporal information and thus a more detailed voicing-tracking.

CHAPTER 2

Methodology

In order to characterize the voicing profile of Polish, French, German and American English sonorants in all possible phonotactic contexts, I ran a computational investigation of large speech corpora based on a design from Möbius (2004: 5-26). The study uses professional speech corpora, which are analyzed by extracting positional and segmental properties of sonorant consonants in order to glean temporal data about their voicing probabilities. Section 2.1 describes the databases used for the four languages and their format. The following section (2.2) characterizes speech processing (feature extraction, definition and the preparation for the computational analysis), and defines methods used for the automatic analysis and the statistical computing. Finally, in section 2.3 a set of predictions will be made. The results will be presented in Chapter 3.

2.1 Databases

2.1.1 German

The analysis of consonant voicing conducted by Möbius (2004), incorporated as a model study in this dissertation, was carried out on a large speech database recorded by one male, professional German speaker. The MS corpus[14] was designed for the purpose of the unit selection speech synthesis project, SmartKom (SmartKom 2003; Schweitzer et al. 2003). It was designed to provide coverage not only for domain-specific, but also for open domain output, where the entire

[14] So called because of the initials of the speaker.

language was the target. As a consequence, sentences for the corpus were selected from a large database in order to achieve a maximum number of combinations of speech sounds and the contexts in which they occur. The database thus contains a full set of German diphones, as well as rich combinations of phones and contexts in which they occur, including segmental context. Moreover it contains prosodic information about syllabic stress and syllable structure, pitch-accents and boundary tones. The database lasts 160 minutes, includes 17489 words embedded in 2301 sentences (for the frequency of sonorants in the MS corpus, see Table 1 below). The MS corpus was segmented automatically on the phone, syllable and word levels using HMM-based forced alignment (Rapp 1995). Prosody labels (GToBI) were performed manually.

m	n	ŋ	l	R	j
2708	8285	669	3329	2472	472

Tab. 1: The inventory and frequency of occurrence of German sonorants in the MS corpus.

2.1.2 Polish

The Polish Speech Corpus (Demenko et al. 2008: 1650-1653; Demenko et al. 2008: 85-101) was designed for the Polish version of the Bonn Open Synthesis System (BOSS) (Klabbers et al. 2001), originally aimed at German and Dutch speech. The database comprises 115min (3249 utterances) of speech recorded by a professional speaker during several recording sessions supervised by an expert phonetician (for the sonorants' frequency of occurrence in the BOSS corpus, see Table 2 below). It consists of several databases:

- Base A: Phrases with the most frequent consonant structures. Polish has a number of complex consonant clusters. 258 consonant clusters of various types were used;
- Base B: All Polish diphones, realized in 92 grammatically correct but semantically nonsense phrases;

- Base C: Phrases with CVC triphones (in non-sonorant voiced context and with various intonation patterns). 664 phrases were recorded for triphone coverage;
- Base D: Phrases with CVC triphones (in sonorant context and with various intonation patterns). The length of the 985 phrases varied from 6 to 14 syllables to provide full coverage of suprasegmental structures;
- Base E: Utterances with the 6000 most frequent Polish vocabulary items. 1109 sentences were recorded (Demenko et al. 2008)

m	n	ɲ	l	r	j	w
4216	7065	2540	2600	3394	4605	3532

Tab. 2: The inventory and frequency of occurrence of Polish sonorants in the BOSS corpus.

The computer coding conventions were drawn up in SAMPA for Polish and in the IPA alphabet (SAMPA 2009; IPA 2009). [15]

2.1.3 French

The extraction of voicing information for French was based on the data produced and processed by the SVOX AG. Information included in the files contains data concerning the duration of the phones, their syllabic, word and phrase position, as well as the binary voicing decision corresponding to each time step of the duration of the phone. These data were based on the professional female speaker speech corpus designed for embedded automotive speech applications, including speech recognition and text-to-speech solutions (SVOX 2009). The corpus comprises a rich inventory of phones, including sonorants (Table 3).

[15] For all transcriptions, see Appendix.

m	n	ɲ	l	R	j	w
2616	2597	2597	1893	6982	1893	684

Tab. 3: The inventory and frequency of occurrence of French sonorants in the SVOX corpus.

2.1.4 American English

Studies conducted on American English were based on the Boston University Radio News Corpus (Ostendorf et al. 1995). This database contains recordings of FM radio news announcers associated with WBUR, a public radio station. The corpus was originally designed for text-to-speech synthesis, particularly prosody patterning. It consists of two parts: radio news and lab news. The first part of the corpus (radio news database) embodies stories recorded during broadcasts by seven professional speakers. As presented in Table 4, three female and four male speakers were divided into two groups: A speakers being those whose job is to read news live, and B speakers, who normally pre-record and edit their stories.

Speaker	F1A	F2B	F3A	M1B	M2B	M3B	M4B
Minutes	52	49	107	48	58	32	91
Stories	43	34	340	36	35	21	62
Clean Paragraphs	276	124	341	161	214	126	236
Noisy Paragraphs	1	40	51	108	102	32	41
Words (times 1000)	11.9	12.2	28.6	15.7	18.4	10.5	25.6

Tab. 4: List of speakers and the amount of the recorded material.

The second part of the corpus (lab news database) combines laboratory recordings of the same speakers reading stories previously read only by the B-type speakers. The phonetic labels were generated automatically and based on the TIMIT phonetic labeling system (Lamel at al. 1986: 100-109). Since not all the recordings have been labeled and corrected manually, I decided to use

the data from the F1A speaker, whose lab news materials were fully annotated. The F1A radio news recordings were only automatically corrected and their analysis serves as a comparison, though they do contain a larger coverage of phones (Table 5).

	m	n	ɲ	l	r	j	w
BOSTON radio news	1658	4145	33	2435	3733	663	660
BOSTON lab news	179	447	59	302	1559	71	103

Tab. 5: The inventory and frequency of occurrence of English sonorants in the BOSTON RADIO corpus.

The original format of the corpus contained files with word and phone division. Thus, for the Festival tree structure building, it was necessary to create also syllable files. This task was performed using Daniel Kahn's syllabification theory (Kahn 1968: 1-218), which was later implemented by William Fisher (Fisher 1997) in his syllabification tool. It has provided reliable syllabification output, choosing always one out of three syllabification possibilities, which, according to Fisher's algorithm, is the most reliable one. Inspection of random subset has not revealed any major errors.

2.2 Feature extraction and analysis

The analysis method is based on the previous study of German consonant voicing conducted by Möbius (2004). The author proposes automatic analysis of voicing performed using F0 information of each phone along with its durational and positional specification. This input data later served for the generation of a voicing profile using speech analysis software, ESPS/xwaves (Entropic Inc.).

2.2.1. Speech processing and voicing profiles

The data processing involved automatic analysis, but it did have to be computed and adjusted for each language separately. Figure 21 illustrates all steps of the study (using the example of the BOSS BLF-text files), starting from the collection of the data from the corpora, aligning text information with the speech signal, through the generation of frame-by-frame voicing information and producing Festival utterances in order to form the output table of the result and its statistical analysis.

For each of the corpora (excluding French, where voicing information had already been extracted), temporal information about phone, syllable and word boundaries were extracted and stored in separate files. After defining the phone sets of the languages under investigation, symbols corresponding to their features were adjusted to SAMPA or TIMIT system annotations[16]. This information, along with the speech signals, served as an input for further processing.

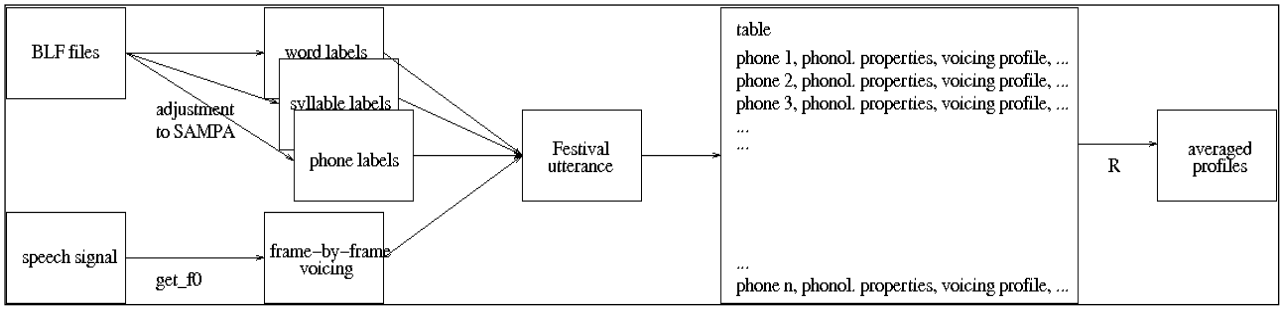


Fig. 21: *Illustration of the analytical steps of the study.*

2.2.1.1 *Computing issues*

Preparation of the text files containing the linguistic information needed for further processing

[16] Slight changes were made if the Festival tool complained of graphical errors.

required adjustment to the ESPS format [17](Fig. 22).

Thus, a number of computing issues emerged, most of which were solved by using simple PERL and PYTHON scripts.

```
0.0975625 121 p
0.17825 121 t
0.21825 121 e
0.26825 121 l
0.31825 121 e
0.3835 121 m
0.43825 121 a
0.49825 121 r
0.587375 121 k
0.6673125 121 e
0.72825 121 t
0.7890625 121 e
0.851375 121 r
0.91825 121 k
0.957 121 a
```

Fig.22: Example unit of a word 'telemarketerka' in ESPS format, preceded by a pause.

The BOSS label files were converted to phone, syllable and word label files in ESPS format by reprogramming them from the BLF format (Polish annotation interface). Furthermore, the Polish SAMPA phone set provided with the Polish BOSS corpus (Demenko et al. 2008) was modified for further processing using the Festival tool. This transition was justified by the emergence of errors due to unknown symbols which occurred while running Festival (see next section) and resulted in the deletion and replacement of some of the symbols.

Preparation of The Boston University Corpus required translation from the TIMIT annotation system into the SAMPA system and was based on the feature description provided by the corpus documentation in comparison with the TIMIT phone classification (Keating et al. 1994: 91-120).

[17] ESPS data format contains three columns of information: (1) end time of the segment, (2) color code number and (3) the name of the segment.

The data obtained from the SVOX French corpus were provided in its format, which differed slightly from the output files obtained for the other investigated languages. The data for the French phones contained voicing binary information, which consisted of more than nine time intervals. In these cases the number of voicing steps was interpolated in order to obtain nine time points of voicing decision, analogous to other profiles. Furthermore, the French phone set was defined along with the consonantal features, which were subsequently extracted for the final result file. The format of this file was similar to the Festival output, which in turn enabled parallel statistical analysis (for further information, see scripts in the Appendix).

2.2.1.2. *Festival utterances*

Festival is a text-to-speech software developed by Edinburgh University and Carnegie Mellon University. Implementation is based on linguistic and phonological knowledge. This software generates a linguistic representation of an utterance from a given text and acoustic properties from the linguistic representation and phonological features (Fig. 23).

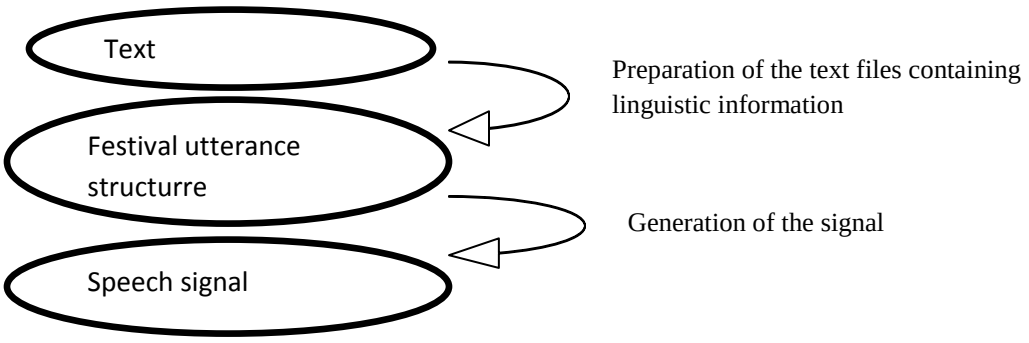


Fig. 23: Festival utterance formation.

Festival can analyze all linguistic and phonetic properties of utterances after having been given the linguistic details about the investigated signal, such as structure of the utterance, phone set of a language with feature description, and information about voicing.

The input data provided in the ESPS-format label files are later integrated with the speech signal. The structure of the Festival utterances entails linguistic items, which contain feature values and are joined by relations, sharing the same properties within one level. Segment, syllable and word items belong to the so-called 'flat' relations, while syllable structure information and intonation belong to the 'tree-organized' relations (Fig. 24).

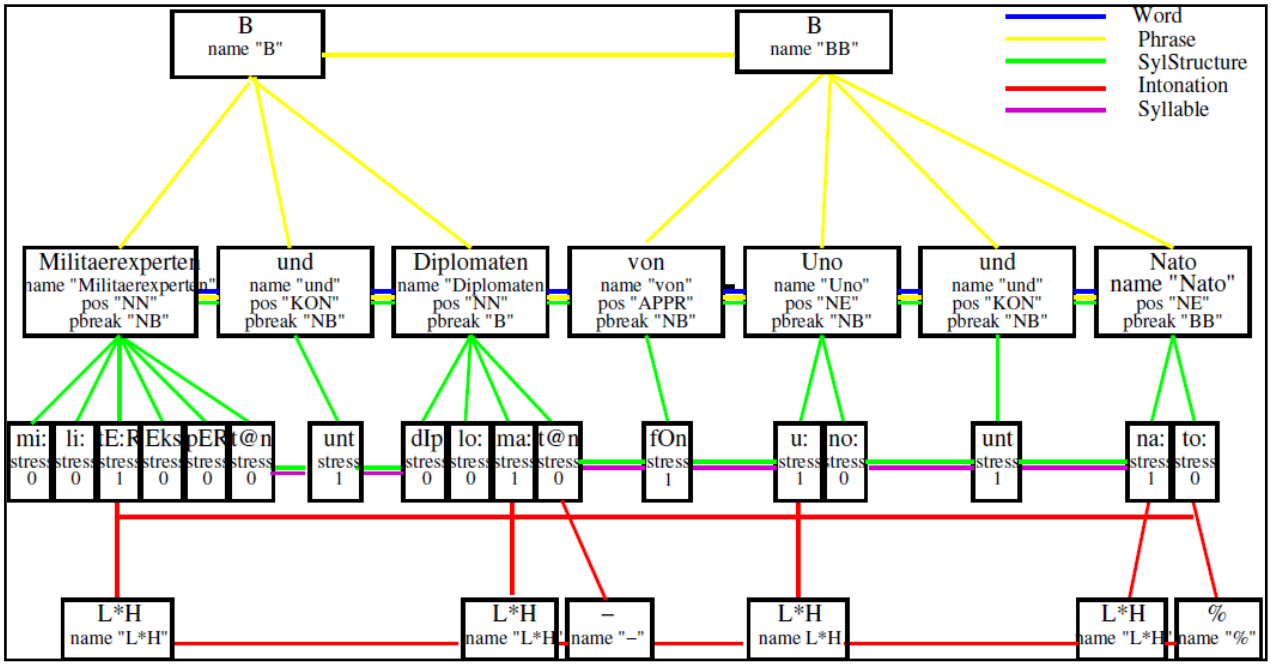


Fig. 24: Hierarchical representation of Festival utterance structure (Schweitzer 2008).

From the linguistic information containing a given text, Festival provides the data about various features and relations. For the purpose of this study, the following information (Table 6) was obtained: the name of the phoneme (name); its preceding and following segment (p.name and n.name); consonant type (ph_ctype); voicing temporal binary information (lisp_voice10-90), file name (lisp_fileid); position of the investigated phoneme in the syllable (pos_in_syl); and its structure regarding the word in which it occurred (R:SylStructure.parent.position_type). This type of output file allows thorough, multi-factor analysis of voicing properties, showing binary

voicing values changing over time and over contexts. Moreover, Festival enables the extraction of many other features, not investigated in this study, such as, for instance, word accent, position in the phrase and pitch value.

name	r	k
p.name	a	r
n.name	k	e
ph_ctype	r	s
lisp_voice10	1	1
lisp_voice20	1	1
lisp_voice30	1	1
lisp_voice40	1	0
lisp_voice50	1	0
lisp_voice60	0	0
lisp_voice70	0	0
lisp_voice80	0	0
lisp_voice90	1	0
lisp_fileid	A0001	A0001
pos_in_syl	2	0
R:SylStructure.parent.position_type	mid	mid

Tab. 6: An Example of Festival output information.

2.2.1.3 *Voicing Profiles*

After having specified linguistic properties and features, Festival aligned the textual information with the speech signal and generated F0 values for each phone using an integrated ESPS get_F0 tool, which reports a binary voicing decision for each 10msec analysis frame, with values '1' signifying voiced and '0' indicating unvoiced. This frame-by-frame analysis enabled the generation of nine equidistant time intervals for each investigated sonorant from 10% to 90% of its duration. Voicing status was extracted at those points. Thus, voicing probability at a given temporal position is a computed mean percentage value of all the phone exemplars in the analyzed speech corpus that are voiced at this position. If the time point of the phone in question happened to lie between two different voicing status values, its value was obtained by

interpolation rather than by a categorical decision. Probability of vocing of a sonorant derived in this way is referred to as the *voicing profile* of a segment.

2.2.2 Extraction of the liquids

Due to the phonotactical constrains of Polish, German, American English and French, segments in the following environments were extracted.

For Polish:

- intervocalic [r, l, m, n, w, j, ɲ]
- [r, l, m, n, w, j, ɲ] with left hand voiceless obstruent context in word-initial positions
- [r] with left hand voiceless obstruent context in word initial and final positions

For French:

- intervocalic [R, l, m, n, w, ɲ]
- [R, l, m, n, w, ɲ] with left hand voiceless obstruent context in word-initial positions
- [R] with left hand voiceless obstruent context in initial and final positions

For German (Möbius 2004):

- intervocalic [R, l, m, n, j]
- [R, l, m, n, j] with left hand voiceless obstruent context in all word positions with and without an intervening syllable boundary between the obstruent and the sonorant

For American English:

- intervocalic [r, l, m, n, w, j, ŋ]
- [r, l, m, n, w, j, ŋ] with left hand voiceless obstruent context in word-initial and medial positions with and without an intervening syllable boundary between the obstruent and the sonorant

2.2.3 Statistical analysis

Results of the frame-by-frame voicing profiles were statistically analyzed with R. The graphs produced show normalized time (percentage of the duration of the segment) plotted on the x axis, while the voicing probability (referred to as voicing profiles) computed as the percentage of the phones undergoing de-/voicing (the number of exemplars in the corpus) is plotted on the y axis. Thus, the results show sonorant exemplars in the corpora and their voicing probabilities over time.

The extraction of the phonological environment in the statistical analysis is possible due to previously defined consonantal Festival features, including information about the manner and the place of articulation of the sonorants.

2.3 Predictions

Now that the background literature has been reviewed (Chapter 1), as well as the frame-by-frame analysis of the voicing profiles has been performed, some hypotheses concerning the nature of voicing in German, Polish, French and American English liquids can be formulated.

According to the studies concerning German sonorants, Hall (1993) claims that [voice] can only be a privative feature, playing a phonological role only when its value is positive. The author argues that there is no phonemic contrast between voiced and voiceless sonorants in German, which means they are not marked for the feature [voice]. Contrary to these assumptions, a model study on the voicing profile of German sonorants (Möbius 2004), showed that phonological context and intervening syllable boundary influence voicing probability of the sonorants, which varies depending on these conditions. As a result, it has been further assumed

that positional and contextual factors are crucial for the voicing profile of sonorants investigation in the other three languages.

As was pointed out by Gussman (1997, 2007) Polish sonorants do not devoice word-finally after a vowel (e.g. sy[n] 'son'), before a voiceless obstruent (e.g. la[mp] 'lamp, gen.pl.') and in sonorant clusters (e.g. se[jm], 'parliament'). However, they do devoice between voiceless consonants or after a voiceless obstruent before a pause. Moreover, the author claims that their devoicing behavior, which in Polish occurs in word-final position following an obstruent, is governed by the syllable structure. Going against this theoretical hypothesis, Rubach (1996) states that voice assimilation is not connected with the licensing principles for onsets and codas, but is rather a result of the linear adjacency of Laryngeal nodes. From these assumptions we can predict that devoicing of sonorants in Polish will either be a result of their word and syllable position, their contextual placement, or both of these factors.

Similar questions arise for French liquids, as their distribution is comparable to the Polish ones. According to Dell (1995: 5-26) word-final consonant clusters in French represent sequences of a coda followed by an onset, like /dr/ cluster in *mordre*. As was pointed out by the author, a word-final cluster consisting of an "unpaired, branching onset" whose second member is a sonorant may only belong to an obstruent-sonorant sequence, forming a "degenerate syllable" whose rhyme is composed of a nucleus that is not associated to any distinctive features. My prediction is that the voice licensing will demonstrate a pattern similar to the one occurring in Polish, where word-final position governs devoicing of liquids, being influenced by the left segmental context of the obstruent.

Voicelessness of English sonorants (equated with aspiration), as described by Iverson and Salmons (1995), is a phenomenon caused by the state of an open glottis. As a result of sharing the

feature [spread glottis], sonorants following voiceless obstruents become devoiced as well. The authors point out that such a pattern holds only for those languages in which stops can be characterized with an aspiration, i.e. where feature [spread glottis] plays either the phonemic or a phonetic role. This is not the case in Polish obstruent-sonorant clusters, which, according to Iverson and Salmons (1995), serves as an explanation for the sonorants being voiced in the word-initial position. In the study conducted by Tsuchida and colleagues (2000) it was observed that sonorant devoicing in English may only be partial and depends on contextual factors (there seems to be less devoicing after voiceless obstruents than after stops). The authors share the same view as Iverson and Salmons (1995) that there is one glottal opening per onset which influences the obstruent and the following sonorant. The hypothesis presented by the authors is based on a distinction between stops and fricatives in which, during the production of the former, the glottis is open, while in the latter the glottal opening is only formed at the midpoint of the fricative. Studies conducted by Tsuchida and colleagues (2000) involved experiments with a flexible fiberoscope, which was inserted in the subject's nasal cavity in order to capture the size of the glottal opening. These studies completely reversed Iverson and Salmons' (1995) assumption. As a result, it is now assumed that fricative glottal gestures are much longer than those occurring during the production of a stop. My voicing profile investigation aims at verifying both of those assumptions by analyzing obstruent and stop contexts followed by a sonorant, particularly a liquid, on a large corpus of data.

CHAPTER 3

Results

This chapter presents results from the statistical analysis of voicing profiles in Polish, German, French and American English sonorants, each of which are described in separate subsections for each language. The chapter will be concluded with a general discussion. The next chapter will contain a more detailed analysis of the results concerning Polish liquids and [voice] licensing.

Grounding the studies in the previously proposed voicing method analysis from the investigation made by Möbius (2004), as well as in the theoretical descriptions of [voice] licensing in Gussmann (1992, 2007), Lombardi (1991, 1995), Rubach (1996) and others, the results presented in this chapter encompass influences of two factors governing voicing probabilities: segmental context and syllable structure information. In the figures presented in the following chapter, normalized time (percentage of the time duration of the segment) is plotted on the x axis, while the percentage of the phones undergoing de-/voicing is plotted on the y axis[18]. Thus, the results show all sonorant exemplars in the corpora and their voicing probabilities over time. The analysis focuses mainly on the liquids, since their voicing probabilities seem to vary to the largest extent. However, for each of the languages under investigation, the voicing profile of all sonorants is given for the sake of comparison.

3.1. German

The following graphs (Fig. 25 and Fig. 26) show the results of a voicing study conducted by

[18] The graphs vary sometimes in the 'percentage voiced' scale, since in some cases devoicing amount was more legible when shown in smaller scale.

Möbius (2004), which served as a methodology example. In Figure 25 voicing probabilities of German sonorants preceded by a voiceless obstruents are presented, where there is a distinction between sonorants occurring in the same syllable as voiceless obstruents (dashed lines) and those separated by a syllable boundary (solid lines). As pointed out by Möbius (2004), voicing probability increases for those sonorants which are separated from the voiceless obstruents by a syllable boundary, like in Stecknadel [ʃtɛk.na:dəl] 'pin'. The strongest effect was observed for [j] and [R]; the weakest effect, for nasals. Furthermore, the left segmental context was investigated (Möbius 2004) and it has been shown to be a crucial factor influencing the voicing probabilities of German sonorant segments. Figure 26 shows results reported from vocalic/sonorant-sonorant occurrences, as well as voiceless obstruent-sonorant occurrences. The author describes clear contextual dependencies, showing initial devoicing in sonorants only when preceded by voiceless obstruents (with exception of [R] minor devoicing after left vocalic/sonorant segments).

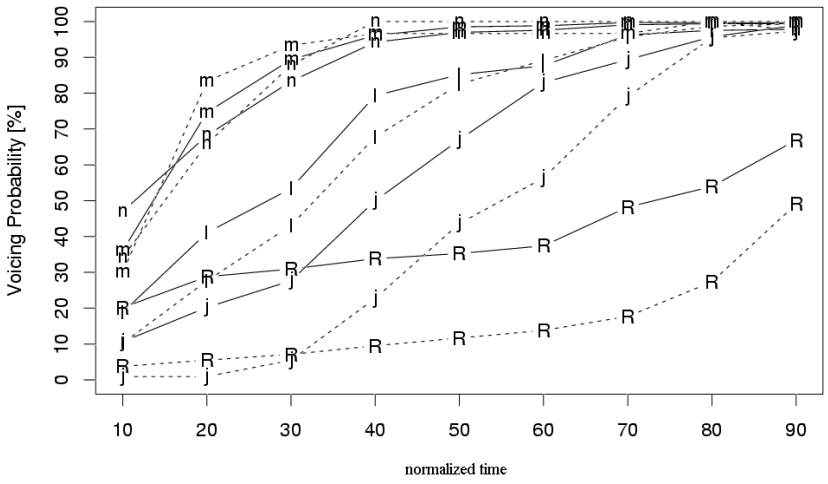


Fig.25: German sonorants with left voiceless obstruent context within one syllable (dashed lines) and those separated by a syllable boundary (solid lines) (Möbius 2004: 19).

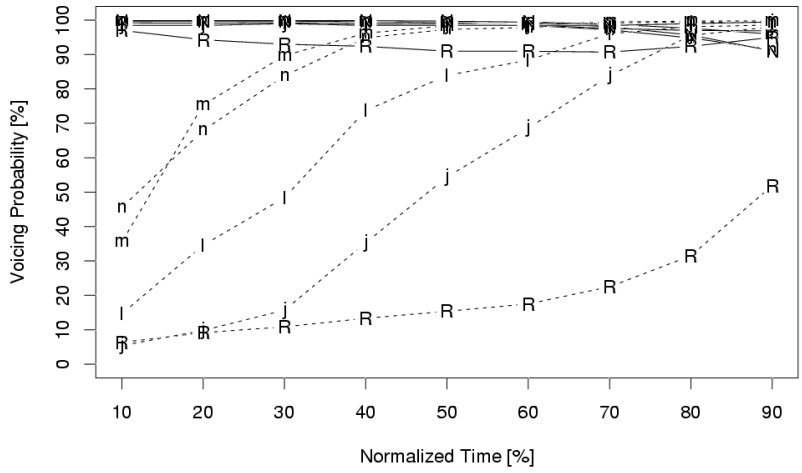


Fig. 26: German sonorants with left voiceless obstruent context (dashed lines) and those with vocalic/sonorant context (solid lines) (Möbius 2004: 17).

3.2 Polish

Similar conclusions concerning contextual effects have been drawn for Polish sonorants, where the left sonorant/vocalic environment did not trigger any major devoicing processes. Figure 27 shows voicing probabilities of Polish sonorants in all word positions with the left vocalic/sonorant context. A general tendency can be observed here: the majority of sonorants is voiced in nearly all exemplars occurring in the corpus. Minor deviation concerns particularly the liquid [r], which will thus be the focus of the analyses of other segmental contexts. It has been observed that the probabilities of voicing in word-medial position reach almost full voicing, both in obstruent/sonorant clusters and in vocalic/sonorants contexts. Thus, the results presented below show only word-initial and final sonorant positions with voiceless obstruent context. Figure 28 depicts voicing probabilities of the word-initial sonorants preceded by a voiceless obstruent, like in trawa [trava] 'grass'. The probability of voicing for almost all sonorants starts from a high level of around 93% and rises to 100% of exemplars.

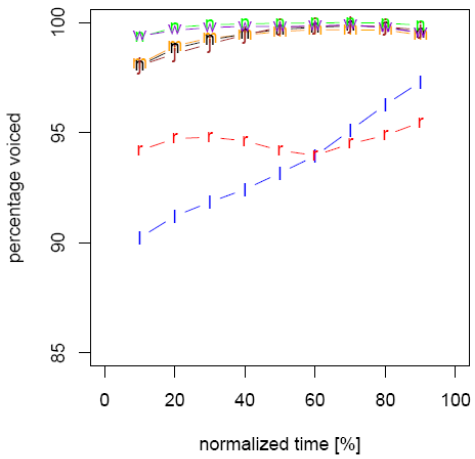


Fig. 27: Voicing profile of Polish sonorants with left vocalic/sonorant context.

As with the previous examples, an exception is the profile of [r], which demonstrates a devoicing tendency through its duration, starting at 75% and rising to 92%. Observing similar voicing tendencies for sonorants in word-final positions which do not exhibit major devoicing processes, I have focused on the liquid [r][19]. The figures presented below (Fig. 29 & 30) show the sonorant in word-initial and final positions in the left voiceless obstruent context. The initial position of the sonorant, such as the word 'pranie' [praɲe] *laundry*, seems to display a devoicing probability that starts at 75% and rises to 90% towards the end of its duration. However, the word-final sonorant, as in the word 'wiatr' [vjatr̥] *wind*, undergoes devoicing in nearly 100% of the exemplars in the corpus.

[19] Voicing probabilities of the liquid [l] in the left voiceless obstruent context have shown only slight devoicing tendency, which does not fall below 90% of exemplars.

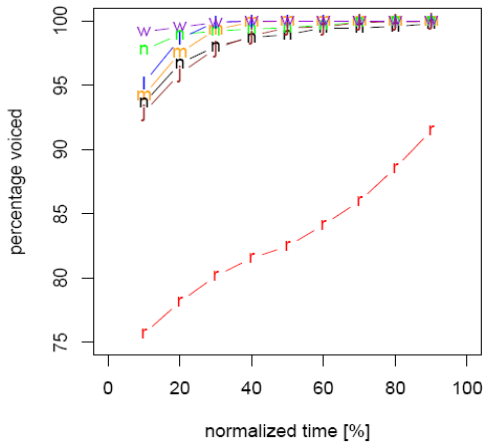


Fig. 28: Voicing profile of Polish word-initial sonorants with left voiceless obstruent context.

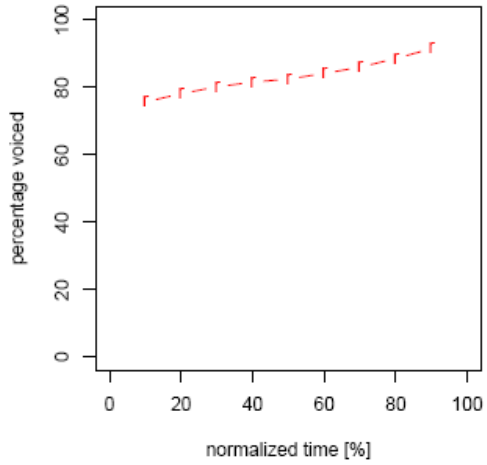


Fig. 29: Voicing profile of Polish word-initial [r] with left voiceless obstruent context.

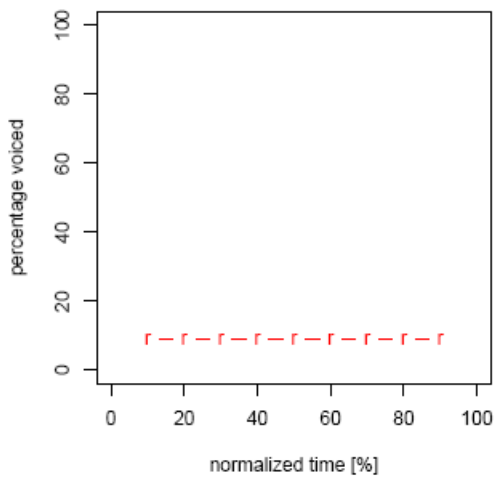


Fig. 30: Voicing profile of Polish word-final [r] with left voiceless obstruent context.

3.3 American English

The results of the voicing investigations conducted on American English may be divided into two parts, demonstrating the voicing probabilities of the sonorants in the laboratory news part of the corpus and those in the radio part. The lab news database of the Boston University Corpus was annotated and corrected both automatically and manually, while the radio news database was processed automatically. The results from two databases serve as a comparison of the voicing probabilities in American English liquids.

3.3.1 Laboratory news database

The analysis of the voicing probabilities of English sonorants in the laboratory news part of the corpus showed no significant tendencies in word-initial positions. On the other hand, medial positions showed large similarities in voicing patterning. Additionally, it appeared that a left voiceless obstruent context did not suppress devoicing (in the positions where it was

phonotactically allowed), which is why sonorants were investigated including all possible left segmental contexts. As is demonstrated in Figure 31, American English sonorants with left vocalic/sonorant context tend to undergo devoicing in a different fashion than the German and Polish ones. Their probabilities vary with regard to the sonorant. The majority (more than 95%) of the occurrences of [l, n, r, j, ŋ[20]] display a stability between 0-20% of their duration and devoice after that time in a large number of exemplars (85%). The biggest variability can be observed in the profile of [w], which devoices at the beginning of its duration to a larger extent than the rest of the sonorants, starting at 84% of the exemplars voiced, reaching 100% towards the end of its duration.

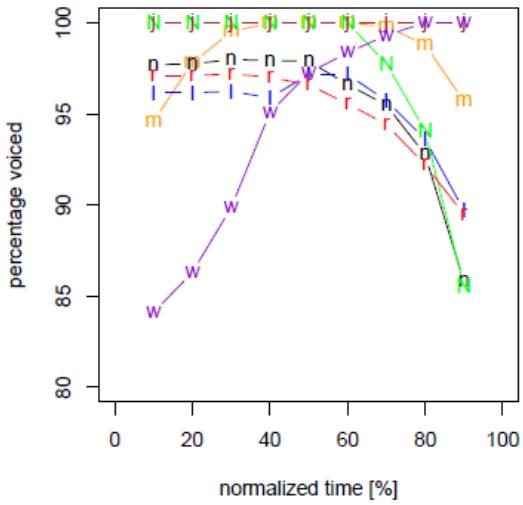


Fig. 31: Voicing profile of American English sonorants (lab news corpus) with left vocalic/sonorant context.

Voicing probabilities of American English sonorants in word-initial and medial position with left voiceless obstruent context show effects of initial devoicing, to approximately 40% of their time

[20] Sonorant [ŋ] is represented by a letter [N] in order to differentiate it from sonorant [n].

duration (Fig.32). For the purpose of comparison, Figure 33 presents the voicing profile of [ŋ, w, r] with left vocalic/sonorant context (like in the word 'miracle' [mɪrɪkl], 'bank' [bæŋk] and 'bowling' [boʊlɪŋ]). Probability of voicing varies remarkably between [w] and [ŋ, r]. The former devoiced in 84% at the beginning of its duration, but tends to be voiced in 100% towards its end. The latter show the opposite tendency: voiced in almost 100% at the start, but begin to devoice between 40% and 60% of their duration. The final figure 34 presents the results from the lab news part of the corpus and shows voicing probabilities of sonorants [r] and [w] in initial and medial word positions with left voiceless obstruent context. Unlike the results observed with Polish and German, it can be seen that the liquid [r] maintains almost 100% voiced exemplars (after slight initial devoicing of not less than 97%) through its duration, whereas [w] preserves its minor devoicing tendencies in no less than 85%.

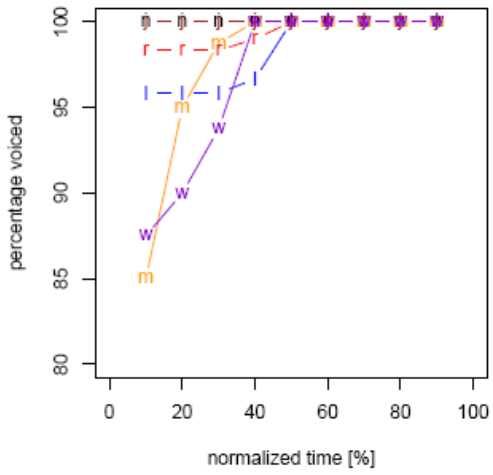


Fig. 32: Voicing profile of American English sonorants (lab news corpus) with left voiceless obstruent context.

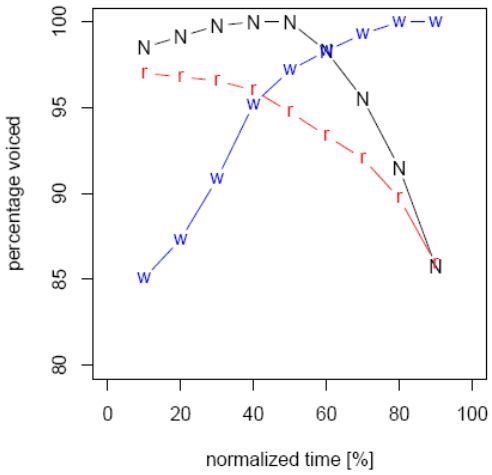


Fig. 33: Voicing profile of American English [ɹ, w, r] (lab news corpus) with left vocalic/sonorant context.

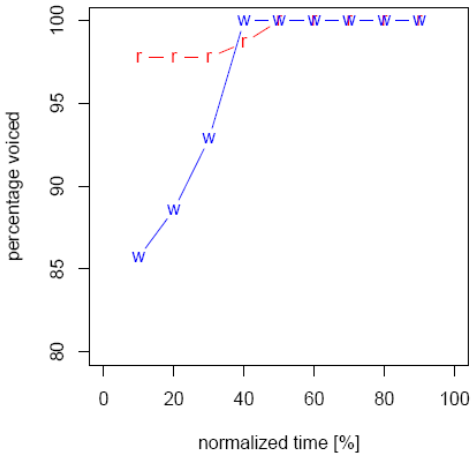


Fig. 34: Voicing profile of American English [w, r] (lab news corpus) with left voiceless obstruent context.

3.3.2 Radio news corpus

When compared to the results from the lab news corpus, the voicing probabilities of the sonorants from the radio news corpus with left vocalic/sonorant context show less variability at the beginning

of their time duration, exhibiting significant devoicing towards their end, particularly with regard to the sonorant [ŋ] like in the word ‘ongoing’ [ɒŋgəʊ.ɪŋ] (Figure 35).

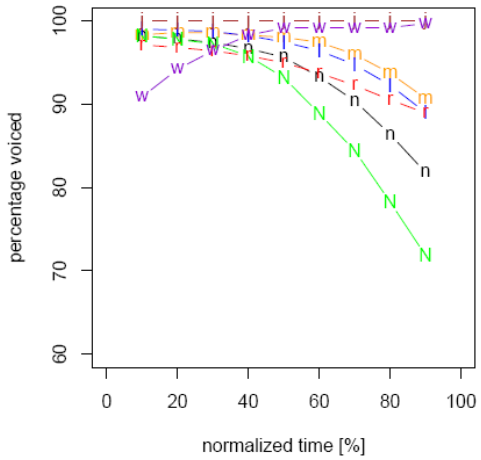


Fig. 35: Voicing profile of American English sonorants (radio news corpus) with left vocalic/sonorant context.

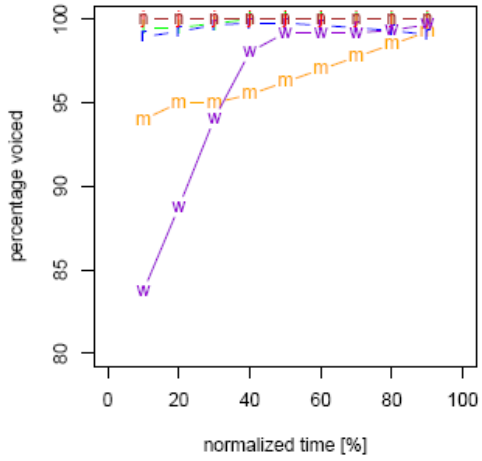


Fig. 36: Voicing profile of American English sonorants (radio news corpus) with left voiceless obstruent context.

Figure 36 demonstrates voicing probabilities of sonorants in word-initial and medial position with left voiceless obstruent context. It is demonstrated that only [m] (little below 95% of exemplars) and [w] (slightly below 85% of exemplars) undergo minor devoicing.

3.4 French

The results of the voicing investigation conducted on French show similar patterns to Polish voicing probabilities, where it is the liquid /R/ which displays the biggest devoicing tendencies, particularly with the left voiceless obstruent context. Figure 37 shows the voicing probabilities of French sonorants in all word positions with left vocalic/sonorant context. The tendency is for most of the sonorant to raise voicing probability to 100% of exemplars towards the end of their duration. Most of them start to be fully voiced from around 80% of the occurrences, with exception of [w], where the initial devoicing reaches a little below 40% and rises to 100% in a similar way as the rest of segments; and [R], which exhibits devoicing tendencies oscillating between 70% and diminishing to 50% of the corpus exemplars.

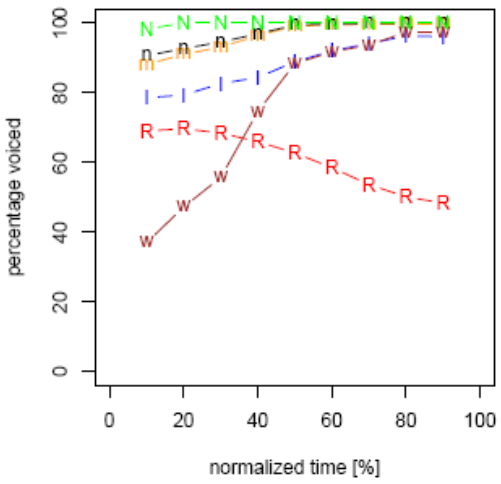


Fig.37: Voicing profile of French sonorants
with left vocalic/sonorant context in all word positions.

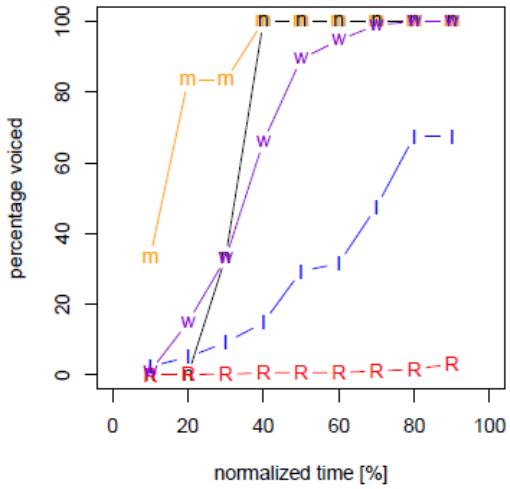


Fig.38: Voicing profile of French sonorants
with left voiceless obstruent context in word-final position.

Figure 38 shows the voicing profiles of French word-final sonorants[21] which changed significantly under the segmental left-hand context variation. While nasals and [w] seem to devoice in the initial period of their duration reaching full voicing towards their end, [R] and [l] devoice to a much larger extent, reaching relatively low voicing probabilities (slightly more than 60% for [l]; [R] has almost no voiced exemplars). In order to verify the influence of left-hand context with regards to word position, Figure 39 displays the voicing profile of word-initial [R] with left voiceless obstruent context, as in the word [tR]aducteur *translator*. The voicing probability of the liquid starts from a relatively high level (around 70% of exemplars) and diminishes slightly below 60% towards its end (it resembles the pattern from figure 37 with left vocalic/sonorant context). On the other hand, [R] with the same left voiceless obstruent context but in word-final position, like in the word qua[tR] *four*, devoices in almost 100% of corpus occurrences (Fig.40).

[21] Word-initial sonorants with left voiceless obstruent context showed similar voicing probabilities to the ones in vocalic/sonorant context.

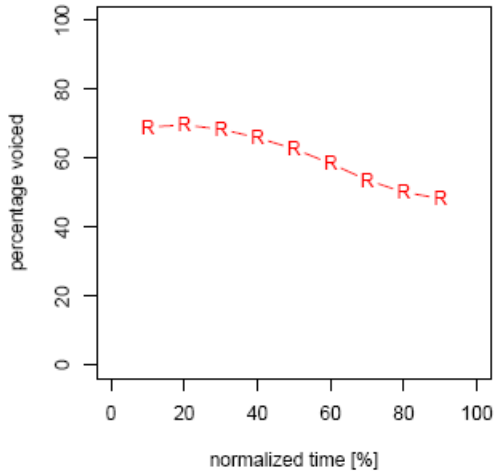


Fig. 39: Voicing profile of word-initial sonorant [R] with left voiceless obstruent context.

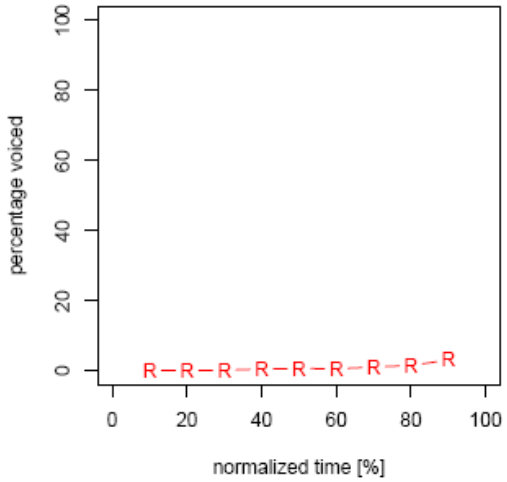


Fig. 40: Voicing profile of word-final sonorant [R] with left voiceless obstruent context.

This remarkable positional variability of voicing behavior of otherwise identical obstruent-sonorant clusters in Polish and French, as well as the non-variability of similar clusters in German and American English will be dealt with in the following chapter with regard to new phonetic category learning, exemplar transfer and other cross-linguistic influences.

CHAPTER 4

Exemplar Theory and language learning

This chapter will provide a theoretical overview of the cross-linguistic exemplar transfer mechanisms during second (and further) language acquisition. In particular, exemplar-based approaches to speech production processes will be discussed with regards to language experience and frequency, in addition to facilitation, competition and contextual effects – the latter ones based on the Context Sequence Model (CSM) (Wade et al. 2010). Finally, it will be hypothesized how newly-acquired phonemes might be influenced by the native language context and phonotactic constraints.

4.1 Exemplar transfer

4.1.1 Language experience and categorization of speech events

In the view of Exemplar Theory (Lacerda 1995; Pierrehumbert 2001) speech perception and production form a loop and are closely linked to each other, while the perceptual space is composed of speech events (percepts) stored in a memory. This perceptual space considered as a cognitive map, contains multidimensional exemplar information about its phonetic and phonological properties, where similar categories/percepts are stored closer to each other and dissimilar ones further apart. Thus, exemplars form clouds built from perceived instances and represent categories of a particular language (Pierrhumbert 2001). The most frequent exemplar clouds exhibit higher density and, being more recent, show higher activation levels (Bybee 2002). However, if they are not refreshed or reactivated, they are removed from the memory, undergoing

decay, leading even to removal of entire categories (Goldinger 1997). A new category might be formed after the perception of a sufficient number of new instances for which there is no appropriate category. Thus, frequency of occurrence and frequency of experience are crucial factors in the formation and maintenance of exemplar categories in the perceptual space. Moreover, it has been demonstrated that the role of language talent, understood as a complex phenomenon that comprises language proficiency, musical, psychological and logical skills, plays a significant role during the processes of categorizing speech events (Jilka 2009: 17-66).

According to Pallier et al. (2004: 78-91), lack of language experience can lead to complete language attrition. The authors demonstrate loss of L1 phonology by testing native speakers of French and native speakers of Korean who were adopted by French families and stopped using their first language for many years (often not being reexposed to it after their arrival to France). Behavioral tests like sentence identification (Korean vs. other languages) or word recognition and fragment detection did not show any differences between Korean adoptees and French native speakers. Moreover, functional Magnetic Resonance Imaging did not exhibit any specific activation for Korean sentences relative to unknown languages. The type of activation with response to the French sentences was, however, very similar for native speakers and Korean adoptees. Additionally, Pallier et al. (2004) describe a perceptual study with French L1 and L2 speakers who were exposed to a three-way (tense, plain and aspirated) Korean VOT contrast. This test showed no significant differences (the exception was one category of a plain-tense contrast where the adoptees demonstrated a marginally significant effect). The authors conclude that early language experiences are no guarantee for maintaining the language's phonology later in life, giving new insight into language acquisition patterns (especially concerning the critical language acquisition age periods).

In contrast, Nielsen (2007: 1961-1964) claims that knowledge about phonetic contrasts modulated by phonetic goodness seems to play a more important role than language experience. The author used an implicit imitation paradigm by comparing two types of stimuli with extended and reduced VOTs in voiceless stops. Her subjects recorded a list of words after being exposed to manipulated stimuli. Results showed significant effects on implicit phonetic imitation, which according to Nielsen (2007) is constrained by the knowledge of phonemic contrasts.

Jilka et al. (2009) have demonstrated the role of language talent in the categorization of speech events and second language acquisition. According to the author, language talent should be differentiated from the notion of language proficiency - a phenomenon that can be measured through a battery of tests comprising logical, musical, language and mathematical skills. The authors point out the additional role of linguistic abilities like phonetic skills, the so-called talent for accent and the talent for grammar during foreign language acquisition. Talent for language accent has multidimensional aspects and is said to be composed of neurocognitive flexibility in finding a way around the established L1 system and its motor pathways controlling articulatory movements, its 'language-ego' understood as pronunciation ability and empathy, and non-phonetic aspects of competence in second language acquisition (like critical age period, abilities in morphological, syntactical and semantical learning). Moreover, Jilka (2009), when describing first and second language skills, presents a view that talented second language learners often exhibit better skills already in their L1. Finally, the author states that established L1 representations might interfere with similar L2 phoneme categories, but new exemplars might also form new, high-accuracy categories.

4.1.2 Facilitation and competition

Storage of exemplars is a process of comparison on the phone, syllable and word levels, where a new token is identified and categorized in a cognitive map. Thus, the exemplar system operates through a mapping of the points in a phonetic space and their corresponding categorization labels (Pierrehumbert 2001). However, during exemplar transfer, stored exemplars undergo facilitation and competition processes resulting from frequency of occurrence and exemplar similarity within the cognitive space.

According to Pierrehumbert (2006), words in a given language which are minimally different from other real words, i.e. words which have many lexical neighbors, tend to have high-probability phonotactics. This observation leads to contradictory conclusions. The author claims that words with high-probability phonotactics generate facilitation effects by being easier and faster to recognize than words with unusual phone sequences. On the other hand, they also have many lexical neighbors which tend to compete with each other during recognition, generating inhibition effects, which slow down recognition and makes it less reliable. The author describes the following example: "ample experience with a dialect should and does facilitate recognition of that dialect. On the other hand, knowing many speakers of a dialect could make it harder to recognize one single speaker of that dialect" Pierrhumbert (2006: 526). These remarks affect Pierrhumbert's observations on the learning process. She points out that the learning of phonological categories is a bottom-up process – from the speech signal through exemplar analysis. It is assumed that learning should demonstrate relations between frequency, sample size and conceptual distance. The author posits that the task of learning a phoneme which is a direct neighbor of a well-known phoneme of an ample size in the memory, can be successful if the new token is frequent and if it is phonetically distinct from its already saved neighbor. If the size of

the token is not large enough and the item is too similar to its neighboring phoneme, it will probably not be learned (Pierrhumbert 2006).

4.2 Cross-linguistic category learning

4.2.1. Context Sequence Model

One of the approaches concerning speech production within Exemplar Theory underlines the importance of the context-dependent nature of speech, where phonetic knowledge guides speech production. Within this notion Wade and colleagues (2010: 227-239) have proposed a speech production Context Sequence Model (partially described in section 1.4.1), which operates on the local level of speech representation and models the prediction of production targets previously considered as having an abstract hierarchical representation. Within the CSM the authors propose a dimensionality composed of a covariance structure that represents a variety of phonetic cues in the language environment. The novel approach to speech production demonstrated by Wade and colleagues (2010) presents exemplars as parts of longer memory stretches of speech, which do not necessarily correspond to particular units like phonemes, syllables or words. The basic assumption of the CSM is that "selection of a stored category exemplar for production is weighted by the similarity of the exemplar's original context with the relevant neighboring sounds in the current production context" (Wade et al. 2010: 228). It is posited that speech production takes place at the segment level, for which an exemplar cloud is created and in which each token undergoes weighting through a match between the current production context and the originally produced one. Thus, production of whole utterances containing segments is modeled based on more than the unit specification, where acoustic information is completed step by step and rooted in the developing production context. Characterization of the segments is made

through the analysis of both the preceding and following contexts of currently produced utterances. The preceding (left) context is composed of acoustic information from recently produced segments, while the following (right) context is the linguistic information that will be produced in the following step.

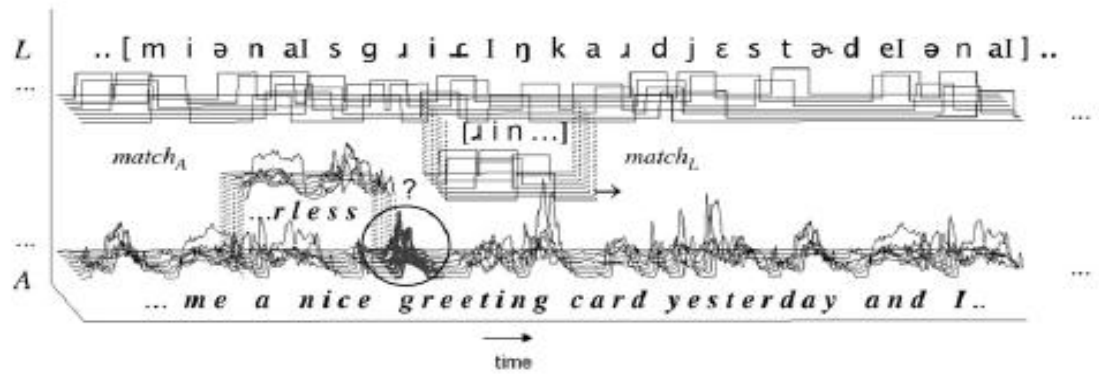


Fig. 41: Segment weighing in context-match comparison during speech production (after Wade et al. 2010: 230)

Figure 41 illustrates the context-mach comparison process, where "the acoustic information that originally preceded an exemplar is compared (match $_A$) with the current acoustic context (the sequence that has just been produced), and the names of the segments that followed that exemplar are compared (match $_L$) with the next planned segments in the current context" (Wade et al. 2010: 230). In this example the authors present a good context match between the immediately preceding part [i:g] of the acoustic sequence and the similar following (just produced) [ɹi] sequence. It is said that this sequence is very likely to be chosen for production, which is a result of context-match score exemplar weighting.

In the CSM framework it is claimed that production targets selected in a given context will be the most proper ones statistically for the language, speaker and dialect. Apart from the context-weighing simulation, the authors conducted two simulation experiments in order to

investigate how much of the context is appropriate to consider in the CSM. In the first experiment, Wade et al. (2010) estimated the importance of different acoustic contexts during the characterization of segment-level exemplars, taking into account factors like frequency of a given context, similarity with other segments from memory and the number of changes in regard to the best matching sequences. More context was added for the comparison. The studies were conducted on a professional speech corpus of standard German sentences using a random selection of frequent and infrequent tokens and computing the formula of similarity scores on them. As a result, the authors presented a subset of the match-to-context-size functions. Figure 42 demonstrates 100 randomly selected examples of each type of match. At the lowest context sizes the 'same types' exhibit better averages than the 'different-type' category. The authors point out that, at 300ms in 'same type' categories, the low-frequency context levels off; whereas in high frequency, it continues to increase up to 500ms.

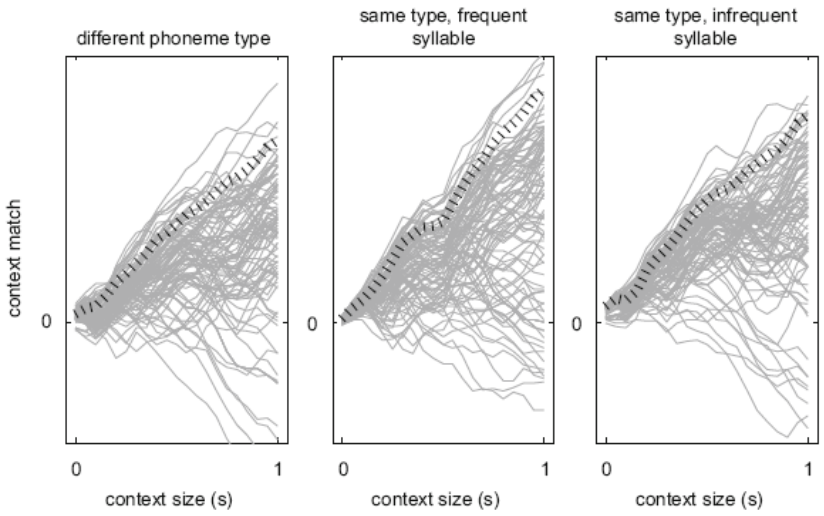


Fig.42: Match-to-context-size functions (Wade et al. 2010: 233).

Context size in the 'same type' categories can also be seen in Figure 43, where same-type advantage increases up to 0.1s and stabilizes for the infrequent contexts, while continuing to increase up to around 0.5s for frequent contexts.

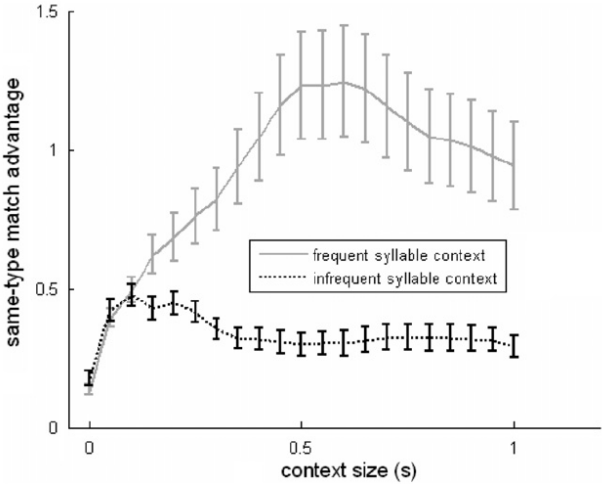


Fig. 43: Same-type match advantage at varying context sizes (Wade et al. 2010:234).

According to Wade et al. (2010), these results demonstrate the amount of context (between 0.1 and 0.5s) a context model should take into account in order to imitate human speech production.

In the second experiment the authors tested frequency-of-context effects in order to verify if "patterns of acoustic variability and exemplar selection that emerged were specific to the frequency of the contexts in which segments occurred" (Wade et al. 2010: 234). It has been shown that segments produced as parts of syllables in frequent contexts differ from their default versions (showing greater variability) to a larger extent than when they occur in infrequent contexts. As a consequence, it is claimed that the effects of syllable-level selection can be approximated by taking into account selection of local acoustic context during segment-level selection. It is also stated that more efficient segments chosen in the model's selection process which appeared to be produced as parts of more frequent syllables support the view that usage-

based accounts of speech provide more economical accounts for speech production than the models requiring discrete levels, such as the syllable.

To sum up, thanks to the Context Sequence Model (Wade et al. 2010) it has been specified how much of a token surrounding context is necessary in order to specify the acoustic shapes of phonetic categories. It has been demonstrated that 0.5s of preceding and following context provide temporal patterns of speech production, where exemplars selected for production are chosen in weighting-comparison processes. The high importance assigned to context match in the CSM makes it a real alternative to traditional models of second language acquisition, all of which concentrate on the structural unit (phone, phoneme, syllable, word) match. Further applications of the CSM with regards to second language acquisition and exemplar learning/transfer will be discussed in section 4.2.3.

4.2.2 Cross-linguistic influences in second and third language acquisition

It has been documented (Saville-Troike 2006: 1-206) that first language acquisition is an innate ability of all humans and is completed without any conscious effort. This innate capacity entails general predispositions of mastering basic phonological and grammatical operations, as well as understanding and creating new utterances until the age of about five or six. By contrast, second language acquisition capabilities are said to be based on varying factors like genetic predisposition, innate capacities used previously in L1 acquisition and the ability to acquire knowledge in a general sense during age advancing (Saville-Troike 2006). Thus, while the initial stages of L1 learning are based on innate capacities (which may or may not be available during L2 acquisition), the second language initial learning phase has its roots in L1 competence along with world knowledge and interactional skills. However, cross-linguistic influence and L1-L2

knowledge transfer are said to have positive and negative properties. The positive influence takes place when the L1 structures happen to be similar to the ones from the L2, while a clear negative transfer or interference occurs in the opposite situation (when the L1 structure is transferred to completely different L2 mechanisms). These processes take place on all levels of language representation like grammar, vocabulary and pronunciation. Saville-Troike (2006: 134-137) also claims that second language acquisition processes undergo facilitation and inhibition due to individual and social factors as well as conditions like memory capacity, analytic ability, need and desire to learn, teaching and school settings. The author claims that in the final state of L2 acquisition native linguistic competence can never be accomplished, although native-like or near-native proficiency can be reached by some learners. With regards to the phonological transfer, Saville-Troike (2006: 143) claims that "transfer from L1 to L2 phonology occurs in both perception and production, and is thus a factor in both listening and speaking. (…) Particularly at early stages of acquisition, L2 learners are likely to perceive L2 pronunciation in terms of the L1 phonemic categories which have already been established". This observation creates a direct link to the notion of exemplar transfer during language acquisition. It will be further discussed in the following section.

Interference during language acquisition also occurs on the levels of second and third acquired language. According to Cenoz (2001: 8-20), cross-linguistic influence between L2 and L3 is due to contextual, aging and recency factors. The results of his studies conducted on 90 elementary and secondary school students (Basque and Spanish native speakers learning English) have shown that cross-linguistic influences were more present in the speech of older students. Moreover, the author demonstrated that the words which were undergoing the transfer to the largest extent were the so called 'content words', i.e. nouns, verbs, numerals, adjectives, whereas

the ‘function’ words like prepositions, conjunctions, determiners and pronouns underwent lesser transfer processes. Additionally, it was claimed that observed patterns of cross-linguistic third language influences lie in the language distance, understood as structural differences between a highly-inflected language (Basque) and the Germanic/Romance languages (English/Spanish). As the author points out, “learners perceive great difficulty of transferring from a highly inflected language” (Cenoz 2001:17). Furthermore he sums up: “linguistic distance is a stronger predictor of cross-linguistic influence than L2 status, but it also indicates that language proficiency and metalinguistic development related to age affect cross-linguistic influence” (Cenoz 2001:18).

Another observation concerning cross-linguistic L2-L3 interference was made by Hammarberg (2001: 21-41), who claimed that not only recency, i.e. frequent contact with a language which enables its easy activation, but also proficiency in the L2 during L3 acquisition as well as the typological similarity of the third language and the status of L2, influence the interaction between the L2 and the L3. The author conducted studies on one speaker with a multilingual background. Over a two-year period he recorded conversations in Swedish which was – at the time of the recordings – his subject’s third language. His subject was a native speaker of English with a German L2 (near native) and additional two L2’s French and Italian (acquired earlier in the past on a basic level). As a first issue in his analysis, Hammarberg (2001) described what he calls ‘language switches’, i.e. phrases and words that were borrowed (and not morphologically or phonologically adapted to L3) by the speaker from languages other than Swedish during her first months of L3 acquisition. Based on the analysis of the recordings, he defined seven types of switches, including edition and self-correction of the elements, metalinguistic comments on the communicative situation, insertion of categories and other similar switches. Hammarberg’s (2001: 27) overall claim is that “L1 is most likely to be activated

for a language switch in those cases where the switch occurs for some pragmatic purpose, whereas L2 tends to be activated in the formulation process in L3". It was also observed that as the proficiency level of the L3 grows, the number of language switches decreases. Just as on the lexical level, influences from the L2 on the phonetic level are observed and reported to diminish with time. However, influence from the L1 becomes more evident as the L2 interferences decrease. Thus, patterns in the phonetic domain which depend on persistent articulatory settings and neuro-motor routines, set up according to L1 requirements, will continue to persist in acquired L2 and L3 articulatory patterns. Hammarberg (2001) sums up his results stating that L1 and L2 both seem to play a role in L3 usage, but that they occupy different roles. It is reported that "L1 dominates in various pragmatically functional language shifts that occur during the conversations and support the interaction of the acquisition of words and other expressions. (…) L2 has a prominent supplier role in the learner's construction of new words in L3, and also in her attempts to cope with new articulatory patterns in L3." (Hammarberg 2001: 35-36)

4.2.3. Production-perception loop

The revised version of the Motor Theory of speech perception (Liberman, A.M, Mattingly, I.G.; 1985:2) assumes that "the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations (…). To perceive an utterance, then, is to perceive a specific pattern of intended gestures." Following this view, acquisition of second, third and further languages should also depend on the perception abilities of the learner along with his or her knowledge concerning the articulatory patterns of the newly-acquired sound system. Miller and Nicely (1955: 338-352) conducted perception experiments

with the 16 English fricatives and two nasals combined with vowels (in total 200 syllables) embedded in random masking noise and frequency distortion. The authors created confusion matrices out of the participants' responses, through which they exhibited classification and discrimination of some acoustic characteristics of speech. For example, according to Miller and Nicely (1955), voicing perception is not only a matter of vocal fold vibration; it is also due to the fact that English voiceless consonants were observed to be more intense than voiced ones since for example voiceless stops contain greater amount of aspiration, which causes them to be shorter than the voiced stops. Observations of similar nature were made with nasality, affrication and duration. This representation of the perceptual space has been translated into similarity scores, which help to uodefine the perceptual distances between the perceived instances. Shepard (1972: 67-113) suggested a mathematical formula[22] for counting similarity out of a confusion matrix dataset, where two confusions between sounds are scaled by comparing the two categories representing them. This allows for comparison of the data where few confusion responses concern one category only, but where another category contains all possible confusion responses. Furthermore, the author proposed a formula for calculating the perceptual distance out of similarities, which also enables, for example, a graphical representation of the numerical confusion matrices. These experimental and mathematical methods demonstrate the complexity of the structure of the perceptual space. They are, however, examples of a unit based comparison, in which context does not play a role.

[22] The method suggested by Shepard proposes calculating similarity between category *i* and category *j* out of confusion matrices from the participants' responses by taking proportions (p) of confusions between two sounds and dividing them by the correct responses. A proportion is the number of times sound X is perceived as sound Y. Thus, a similarity of categories *i* and *j* is: Sij = P*ij* + P*ji*/ P*ii* + P*jj*.

According to Eckman (1977: 315-330) and his Markedness Differential Hypothesis (MDH) it is possible to predict the areas of language difficulty by comparing the grammar to the speaker's native language grammar and assigning features of languages typology: 'marked' or 'unmarked'. In this typology, 'marked' are the areas which occur less frequently than contrasting elements of the same category. Marked features which differ significantly from the L1 are assumed to exhibit greater difficulty (the degree of the markedness corresponds to the degree of difficulty). This means that unmarked L1 features are more likely to undergo cross-linguistic transfer, whereas the marked ones might be more difficult to learn during foreign language acquisition. Translating this proposal into an example investigation of sonorant /r/, it would be reasonable to say there are languages with only voiced liquids like English and there are languages with voiced and voiceless /r/ like Polish, but there are probably no languages with only voiceless sonorant /r/. This, according to Eckman's theory, would mean that the presence of a voiceless liquid implies the presence of the voiced one but not the other way around, which means that voiceless sonorant /r/ is more marked than its voiced counterpart. Pursuing these assumptions, cross-linguistic category transfer among the four languages investigated in this work could lead to the following language learning pattern (combined representation of the exemplar theoretic view with the MDH):

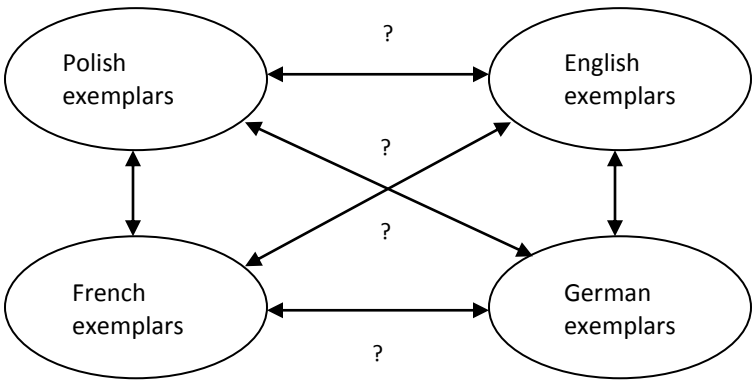


Fig. 44: Hypothesized cross-linguistic category transfer of the unmarked features.

Figure 44 shows possible transfer directions between language pairs English and German, as well as Polish and French. The pattern reflects the above described hypothesis by assuming that marked and difficult segments (like voiceless /r/) will probably be transferred between languages which both contain this sonorant feature, thus they might not be a challenge for a new learner. On the other hand, unmarked voiced /r/ (being 'easier' by definition) can probably be transferred between languages like English and German without great difficulty, because most occurrences of sonorant /r/ tend to be voiced. Moreover, it is also highly possible that voiced /r/ can be transferred among all the languages since it is claimed to be an easily learnable feature. [23]

However, in the Context Sequence Model Wade and colleagues (2010) showed that speech fragment representation used during speech production is also based on the surrounding context employed during the selection of the exemplars. The authors pointed out that multiple language acquisition is dependent on the speaker's exposure to the new L2 (and further) contexts and their distinctiveness from the speaker's native language. This process enables the storage of various phonetic categories in overlapping regions of acoustic space and their preservation for the proper foreign language production. Thus, it is claimed that successful language learning, i.e. non-native categories storage and usage, is not just dependent on similarity and speaker proficiency. This means that predicting cross-linguistic feature transfer between Polish, French, German and American English becomes a more complex task. Figure 45 illustrates the hypothetical influence of the acquired languages (one of them is always the native language) on category storage in the memory space and on production outcome. The influence is still

[23] Another problem that would have to be taken into account is the manner of articulation of all the four types of sonorants, which differs in all the languages.

hypothetical because it is still not precisely known which categories would indeed influence each other and which would remain safely saved in order to be properly used for production.

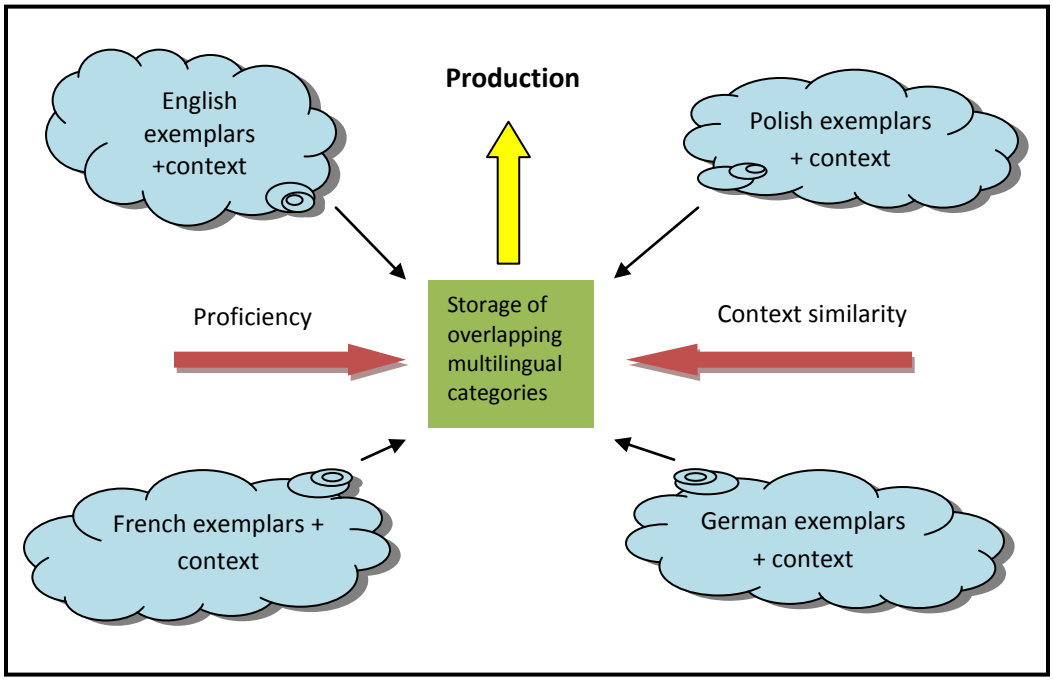


Fig.45: Hypothesized category storage for the four investigated languages.

The simplified schematized graphs presented below (Figure 46) were obtained during the investigation of voicing profiles (see previous chapter). They illustrate the same relations described above on the example taken from the probabilities of voicing in German, American English, Polish and French sonorant /r/ with left-hand voiceless obstruent context. The observed devoicing tendencies indicate clear context influences. Having found no identical item, the speaker might choose the available occurrence of only slightly devoiced word onset /r/ with the same left-hand context (voiceless obstruent) of the German inventory, the middle word occurrence of sometimes devoiced German /ʁ/ in a similar context, or he or she would select any random occurrence of an item seemingly similar to the production target.

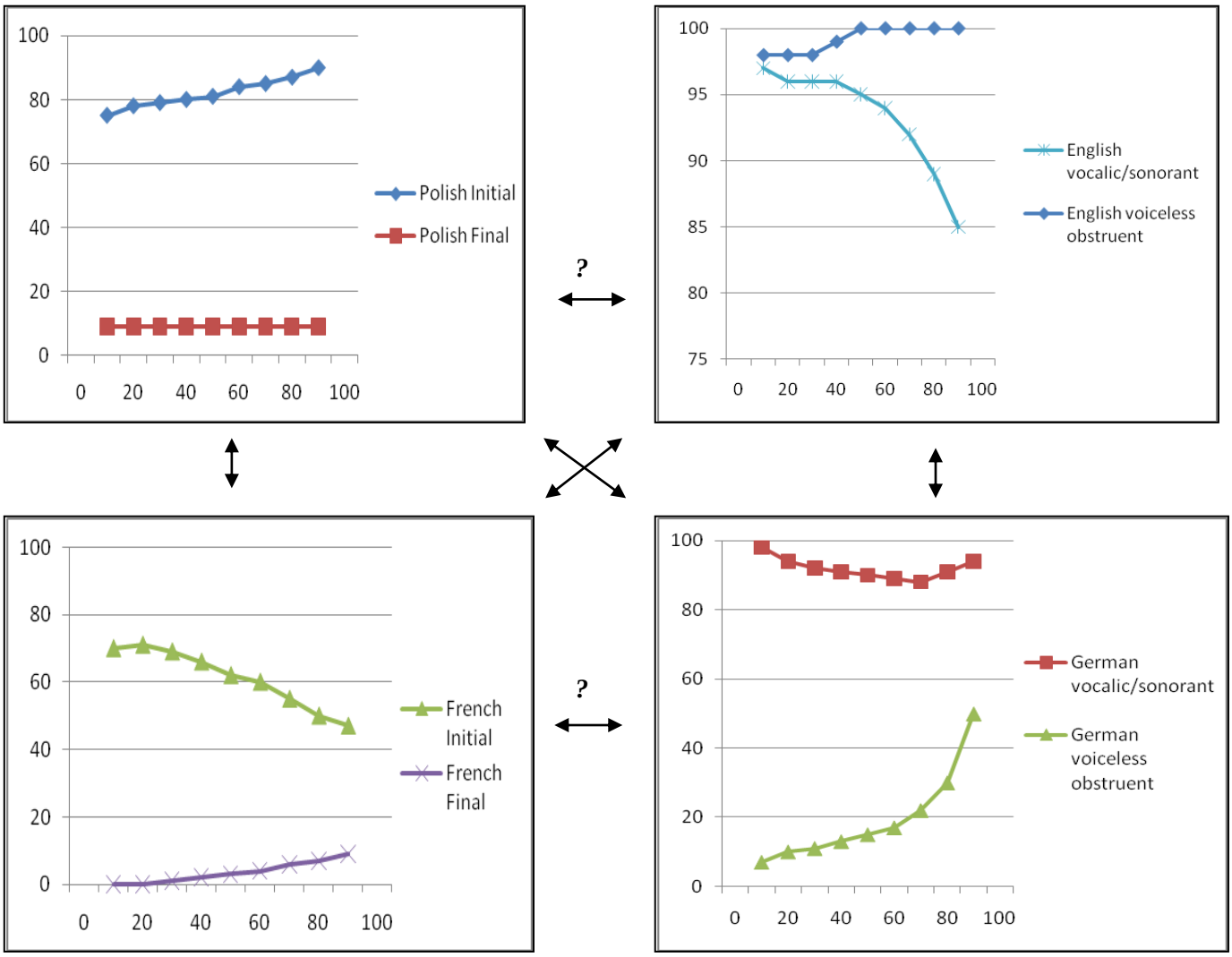


Fig 46: Hyppthesiezd context-matching between sonorant exemplars of Polish, American English, French and German based on the voicing probabilities.

The last option might cause great variability in the speaker's overall production in the foreign language which, as a consequence, increases the possibility of numerous errors. Similar erroneous exemplar matches might occur between American English and Polish/French, as well as between Polish/French and German. This is why languages with similar phonotactic constraints, similar exemplar clouds and category distributions might seem to be easier to acquire. For example, a native speaker of Polish trying to produce the above-mentioned voiceless

/r/ in French seems to have an easier matching path to deal with than the German speaker producing a Polish utterance. In this case, a Polish speaker could probably find a good match taking the available acoustic signal of a French word final voiceless /ʁ/ with that of the Polish left-hand voiceless obstruent context, like in /katr̥/. Here, the possibility of variability, and at the same time erroneous production, seems to be smaller than in the previous example (a similar effect could be found in the opposite case – a French speaker producing Polish devoiced exemplars). Hence, the error rate is predicted by context similarity and not unit similarity or unit markedness, which was the case with traditional language transfer theories. However, research on finding a good match for foreign language production during cross-language exemplar transfer is still in progress in order to clarify all the hypothetical assumptions.

4.3 Conclusion

The process of learning new phonetic categories seems to involve a variety of factors: the phonotactic structure of the non-native language, exposure to the new exemplars, their coexistence among the already-saved categories, the influence of context, degree of markedness, sensitivity during speech perception, and pronunciation/accent talent. Additionally, foreign language production depends on social factors like the background of the speaker, his or her education and attitude to acquiring knowledge (see Dogil & Reiter 2009). The study of voicing contexts presented in this dissertation constitutes a basis for further cross-linguistic investigations in multilingual language acquisition use. Particularly worth investigating are the above mentioned contextual influences during speech production, which result from acoustic and linguistic information exemplar matches in the multilingual exemplar environment. Following the view of Wade and Möbius (2010), it is believed that a key notion in understanding exemplar

storage and exemplar change during acquisition of new phonetic categories always lies in the consideration and analysis of a larger temporal context (as the authors point out, the contexts might even constitute with whole utterances) stored in the memory along the units of speech analysis. It is claimed (Wade & Möbius 2010) that this memory comprises not only lexical items and their abstract representations but also rich acoustic details acquired and stored during language experience. This attitude enables the investigation of context effects in, for example, devoicing processes and the learning of new [voice] governing rules, since it is claimed that "correlations between different types of spectral information in adjacent regions (…) are simply 'normalized for' in the recognition process" (Wade & Möbius 2010: 293). Moreover, according to the authors, selection of the exemplars for speech production is based on the similarity between the original context (all neighboring sounds) and the newly-produced one, which enables a wide range of possibilities during the analysis of multilingual production and perception.

CHAPTER 5

Gestural coordination of Polish obstruent-sonorant clusters

Polish shows remarkable differences in the voicing behavior of obstruent-sonorant clusters. In order to provide more insight into this notable variability, a series of experinents was conducted with the EMA technology. This chapter describes the articulatory profiles of Polish liquids with left voiceless stop context in word onset and coda positions as well as voicing profiles of the same context recorded in parallel. The purpose of this investigation was to examine the relation between a sonorant's devoicing and C-center effects (Browman & Goldstein 1988: 140-155; Byrd 1995: 285-306; Hermes et al. 2008: 433-436; Mücke et al. 2009: 321-338). These studies were conducted thanks to the courtesy and with the cooperation of the Institute of Linguistics at the University of Cologne.

5.1 Introduction

The investigation of the voicing profiles of Polish liquids in voiceless obstruent clusters (in onset and in coda positions) described in this dissertation has exhibited differences in voicing patterning with regard to word position and left-hand segmental context. The reason for these differences might lie in the coordination of the articulatory gesture timing with regards to the syllable structure and its components.

Browman and Goldstein (1988) analyzed two approaches with regard to the gestural organization of speech: local and global. The former was described by the authors as inter-individual coordination, while the latter was seen as large gesture formation. It has been

illustrated that English single onset consonants and consonant clusters exhibit a global organization, whose characteristics are described as the C-center Effect - stable distance of the consonants with regard to the vowel target, measured as the interval between the mean value of the onset consonantal targets and the vowel. By contrast, syllable-final consonant clusters are reported by Browman and Goldstein (1988) to exhibit a local organization of coordination, where the first consonant gesture is related to the vowel target gesture.

Nam et al. (2009: 229-328) investigated temporal factors in syllable structure during speech production. The authors proposed a dynamic model of temporal planning of speech, where "each speech unit is accociated with a planning oscillator, or clock, and the oscillators within the ensemble associated with a particular lexical item are coupled to one another in a pattern represented as a coupling graph" (Nam et al. 2009: 299). Their proposal indicates that the temporal stability of the information pattern during formation of articulatory gestures must be insured by a kind of joining factor. For this purpose the authors employed coupled oscillators, which are said to "exhibit the property of phase-locking" (Nam et al. 2009: 300). In their model gestural score time-planning is related to the oscillatory process, where each gesture is linked to a planning oscillator. Gestures are joined in pairs to one another. After the planning of a gesture begins, all of the gesture clocks (i.e. oscillators) begin to oscillate at random phases in relation to each other. Time relation is measured in oscillator cycles in which there is a formation of changes in phase of the planning of the oscillators, due to the coupling forces. Finally, a stable pattern of oscillator relative phases is formed. The authors (Nam et al. 2009) additionally propose an intrinsic mode of syllable coordination, where the in-phase mode produces the coordination of CV structures (where C is a syllable onset); VC structures are coordinated by the anti-phase mode (where C is a syllable coda). This proposal is based on the authors' assumption that in some

languages it is necessary to coordinate multiple gestures within a syllable with the usage of one of the phase modes. The hypothesis is that the onsets combine in a relatively free way, unlike the codas, which might face more restrictions. However, Golstein and colleagues (2009) demonstrate competition patterns in complex onsets C_1C_2V, where there is a leftward shift of the C_1 and rightward shift of C_2. This serves as evidence for adjustments in the planning model and C-center effects (Byrd 1995: 285-306).

The articulatory timing relations within onset syllables have also been investigated by Hermes and colleagues (2008). The authors examined word initial /s/C clusters in Italian using electromagnetic midsagittal articulography (EMA). Words containing simple and complex onsets embedded in carrier phrases were recorded by two native speakers. The authors calculated the C-center distance to the vocalic target and the rightmost consonants to the vocalic target in words with initial vocalic segment, stop + vocalic segment, /s/ + stop + vocalic segment, as in: 'Lina' (proper name), 'plina' (logatome) and 'splina' (logatome). As a result, Hermes and colleagues (2008) found a stable C-center effect for C vs. CC clusters ('lina' vs. 'plina') and a significant rightward shift of the consonant to the vowel in similar word pairs (indicating that both consonants belong to the onset). No additional rightward shift of the consonant was found in the /s/CC clusters, which led the authors to the assumption that /s/ does not belong to the articulatorily coupled onset.

5.2. Methodology

5.2.1. EMA

Speakers were recorded with a 2D Electromagnetic Articulograph, Carstens AG100, 10 channels. Sensors were placed on the vermillion border of the upper and lower lip, and on the tongue (3

sensors: 1cm, 3cm, 4cm behind the tongue tip). For analyzing coronal sounds, a sensor was used on the tongue tip and further two were attached to the dorsum to analyze articulation of vowels and velar consonants. Two additional reference sensors were placed (attached to the nose, and the upper gums to correct head movements). The data were sampled at 400 Hz, downsampled to 200 Hz, and smoothed with a low-pass filter at 40 Hz. All data were converted to Simple Signal File Format[24] (SSFF), and labeled in the EMU Speech Database System[25] (Cassidy & Harrington 2001).

Target words containing simplex onsets and codas as well as onset and coda clusters (voiceless stop + sonorant), were recorded in carrier phrases: (1) in the onset position: 'Ona mowi pranie aktualnie' ('She is saying laundry currently'), and (2) In the coda position: 'Ona powiedziala Cypr aktualnie' ('She is saying Cyprus currently'), where the underlined target word was recorded with an emphasis articulation mode. See Table 7 for a word list.

	Onset	Coda
/p/	padnij 'hit the deck'	typ 'type'
/k/	kadisz 'Kaddish'	tik 'tic'
/l/	labrys 'ax'	gil 'bullfinch'
/r/	rabin 'rabbi'	tir 'truck'
/p/+/l/	plamić 'to stain'	ZUPL (Zakładu Usług Parkowo – Leśnych) 'Park-Forest Serivce Company'
/p/+/r/	pranie 'laundry'	Cypr 'Cyprus'
/k/+/l/	klawisz 'key'	cykl 'cyckle'
/k/+/r/	krasić 'to flavor'[26]	WIKR – proper noun for a publishing company

Tab.7: Structure of target words.

[24] SSFF is a free of charge format similar to the ESPS (described in Chapter 3). The ESPS is proprietary.
[25] www.emu.sourceforge.net
[26] Cluster /t/ +/r/ has been ruled out of the recording material because of relatively small articulatory resolution of distances between the stop and the sonorant.

The labelling concerned movements in the vertical plane: consonantal and vocalic targets identified by zero-crossings in the respective velocity trace, which captured the time of maximum constriction/opening of the vocal tract during consonant and vowel articulation. For labelling lip movement, the Lip Aperture index (inter-lip distance; Byrd 2000) of upper and lower lip was used. Figure 47 illustrates the manner of landmarks labeling originally proposed by Hermes et al. (2008) and also used in this study.

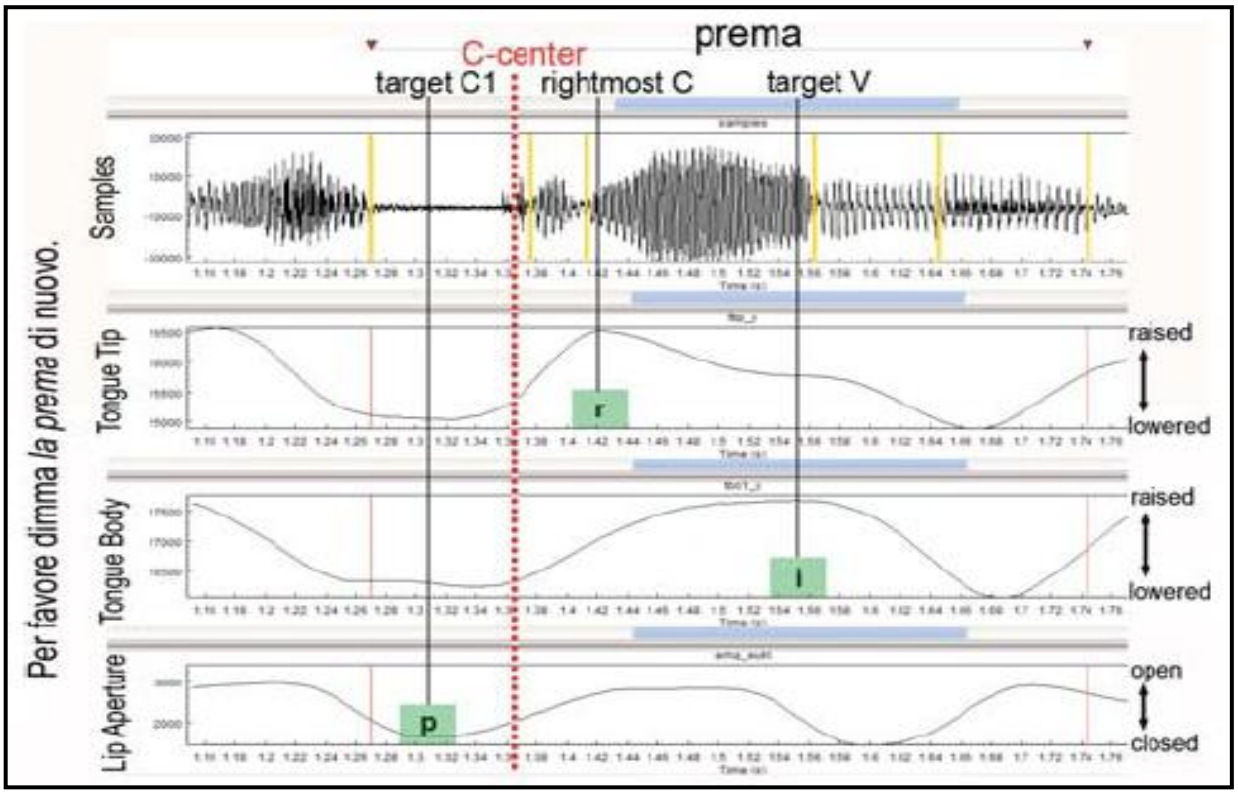


Fig. 47: Labelling example for /prema/ (Hermes et al. 2008: 434)

5.2.1.2 Measurements

The manner of the measurements conducted on Polish CV and CCV clusters stems from the results of previous studies by Hermes et al. (2008) as well as Nam et al. (2009). The authors in both studies found significant rightward and leftward shifts (Nam et al. 2009) in the consonants.

Cross-linguistic examples from Italian, Georgian and English served as a model for the investigation of the Polish articulatory coupling instances.

Following Hermes and colleagues (2008), calculation of the C-center and rightmost as well as leftmost consonant distances relative to the vocalic target were made in order to investigate temporal patterns of the simple and complex onsets. Figure 48 demonstrates leftward shift of C1 and rightward shift of C2 in the word-onset consonant clusters of Polish like in 'pranie' (laundry) and 'krasić' (to flavor). In contrast, coda clusters exhibit no C1 shift, maintaining the rightward shift of C2, like in 'Cypr' (Cyprus) and 'WIKR' (proper name).

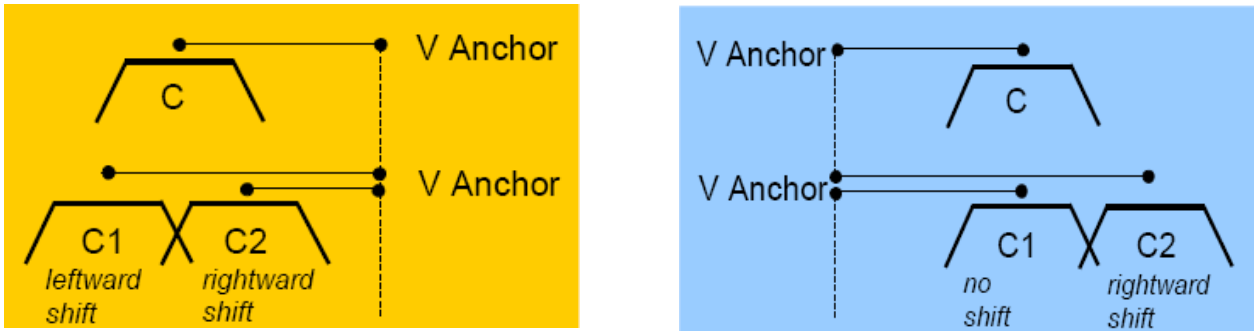


Fig. 48: C1 leftward and C2 rightward shifts of the Polish onset (orange graph) and coda (blue graph) consonant clusters.

5.2.2 Voicing Profiles

Recordings of the speech material conducted for the EMA analysis were manually labelled and further analyzed using Praat (Boersma & Weenink 2010)[27], which enabled voicing probability extraction. The script implemented in Praat analyzed pitch (Praat function: To Pitch) by using an autocorrelation method with standard voicing threshold at 0.45 (the strength of the unvoiced candidate relative to the maximum possible autocorrelation), counting voicing probabilities at

[27] For more details, see script in Appendix.

rate of 100Hz (every 10ms). In the following step, the script used Praat's 'To PointProcess' function in order to analyze the periodicity of the signal (vocal fold vibration) and counted voicing values of the single sonorants as well as the sonorants in voiceless stop-sonorant clusters. This information was put into a TextGrid file and was later implemented by a Perl script in order to count nine voicing time steps by dividing the total duration of a segment (subtraction of the maximum and minimum duration time of the investigated interval: $x_{max} - x_{min}$) into nine equidistant parts and assigning voicing values from the TextGrids to the corresponding time step. The values of the voicing status are almost always categorical (0 or 1), unless the time point of the frame happened to lie between two different voicing status values, in which case this value was obtained by interpolation rather than by a categorical decision[28]. Statistical analysis of the results was conducted with the usage of R scripts in a similar way to the scripts from the voicing investigation described in the previous chapter.

5.2.3 Hypothesis

Based on the previously described studies of Nam et al. (2009), as well as Hermes et al. (2008), it is hypothesized that word-initial CCV clusters in Polish exhibit strong bonding which is demonstrated by competitive coupling of the adjusting consonants. These are in anti-phase against each other and in-phase in relation to the vowel (Fig. 49). On the other hand, coda VCC clusters exhibit a weaker strength of sequential coupling (Fig. 50). This allows for more variation, since they are in anti-phase to one another and in relation to the vowel. Moreover, it is also

[28] Previous voicing investigation was conducted with the usage of Festival (2009) tool, since the investigated corpora contained labelling on the phone, syllable and word levels, which is required by Festival. In the case of EMA speech recordings, the raw speech material was manually analyzed by cutting out the sonorant segments, saving them as TextGrids and later counting their voicing probabilities with the Praat F0 extraction function and a simple Perl script.

hypothesized that voicing probabilities of sonorants in word-initial and word-final stop-liquid clusters might be related to the articulatory patterns in the clusters.

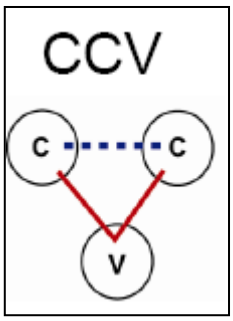
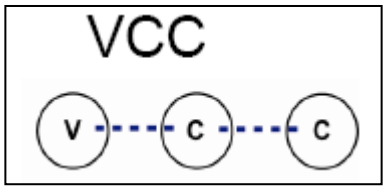


Fig. 49: Phase relation of the onset CCV clusters in Polish.



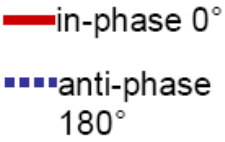


Fig. 50: Phase relation of the coda VCC clusters in Polish.

5.3 Results

5.3.1 Voicing Profiles

The voicing probabilities of liquids in onset and coda voiceless stop-liquid clusters from the speech signal of the articulatory recordings demonstrate a pattern similar to the voicing investigation presented in chapter 3 (see pages: 77-80). The graphs presented below illustrate profiles from the recordings of three Polish (standard variety) native speakers reading carrier phrases containing liquids /r/ and /l/ with left voiceless stop context in word-initial and final positions. Speaker 1 corresponds to the JSf initials in the articulatory graphs; Speaker 2, to the initials JSm; and Speaker 3 to the initials NL.

The voicing profile of /r/ sonorant of Speaker 1 in the onset position (Fig.51) exhibits slight devoicing patterns (not more than 25%); with the /kr/ cluster these patterns are stronger, reaching full voicing at around 60% of its time duration. The /pr/ cluster seems to undergo an even smaller devoicing tendency, reaching full voicing at around 20% of its time duration until its end.

The pattern changes significantly in word-final position (Fig. 52), where for Speaker 1 devoicing of the rhotic is almost complete, starting at around 20% of devoicing and reaching full voicelessness after 50% of its duration time. The voicing probabilities of the liquid /l/ show a similar trend in word-initial position, where the sonorant in both /pl/ and /kl/ clusters is fully voiced for almost the entire time of its duration (Fig. 53). Coda clusters, however, exhibit more variability, showing continuosly diminishing devoicing patterns (stronger in case of the /kl/ clusters), which lean towards 80% of devoicing (Fig. 54).

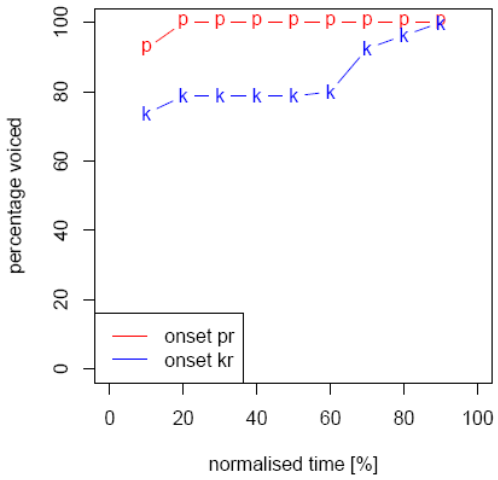


Fig. 51: Voicing probabilities of /r/ in word-initial position with the left voiceless stop context (Speaker1).

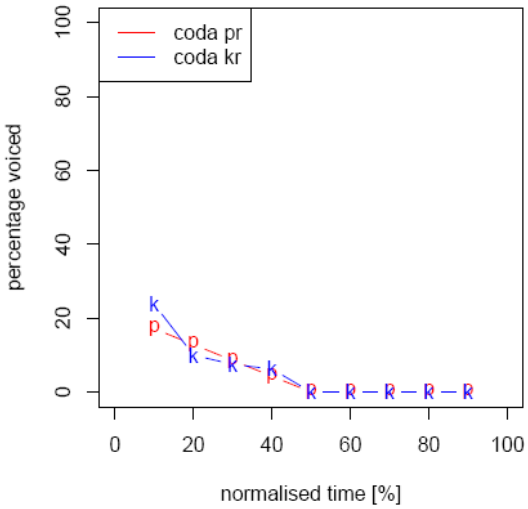


Fig. 52: Voicing probabilities of /r/ in word-final position with the left voiceless stop context (Speaker 1).

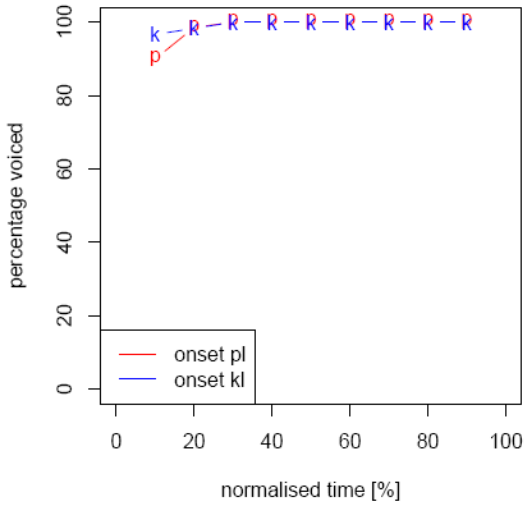


Fig. 53: Voicing probabilities of /l/ in word-initial position with left voiceless stop context (Speaker 1).

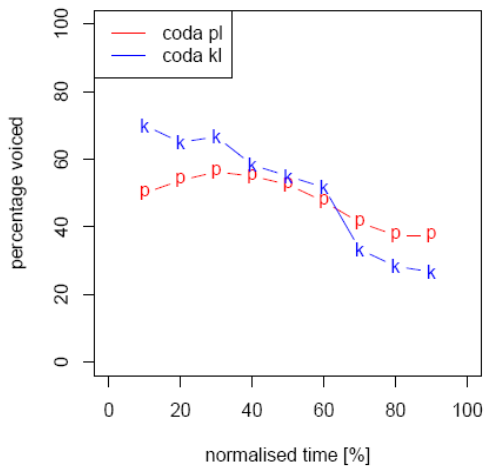


Fig. 54: Voicing probabilities of /l/ in word-final position with left voiceless stop context (Speaker 1).

The voicing probabilities of the second speaker follow the tendency of the voicing profiles from the recordings of the first speaker. Rhotics in the onset /pr/ and /kr/ clusters start their probabilities from slight devoicing (around 15%) and reach full voicing after 40% of their time duration (Fig. 55). In the coda position, it is the rhotic in the /pr/ cluster which exhibits stronger devoicing tendency (reaching 100% devoicing by the end of its duration). The sonorant in the /kr/ cluster tends to undergo systematic devoicing as well, until around 80% of devoicing probability (Fig. 56). The graphs showing the clusters containing the /l/ sonorant in word-initial position with left voiceless stop context (Fig.57) show signs of full voicing in a similar way to previous onset profiles. Coda clusters containing the same sonorant (Fig. 58) demonstrate stronger devoicing tendencies, reaching around 70% of devoicing. In /kl/ clusters the pattern is very stable, maintaining voicing probabilities at the same level through its time duration (at around 30%), whereas in /pl/ clusters there is a diminishing voicing probability, starting from 50% and reaching 70% of devoicing by the end of its time duration.

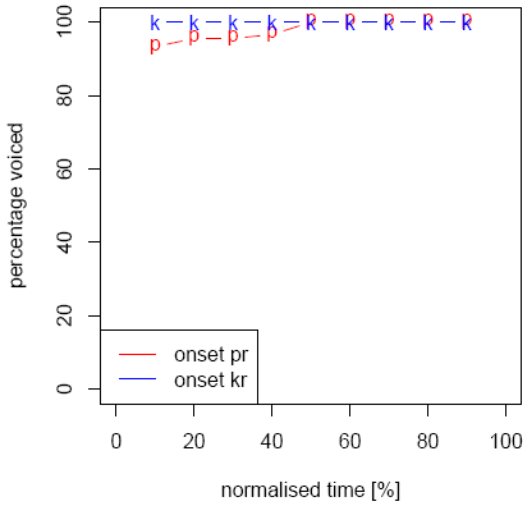


Fig. 55: Voicing probabilities of /r/ in word-initial position with left voiceless stop context (Speaker 2).

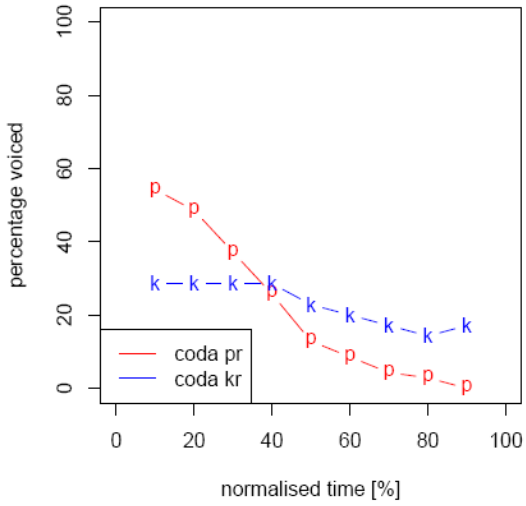


Fig. 56: Voicing probabilities of /r/ in word-final position with left voiceless stop context (Speaker 2).

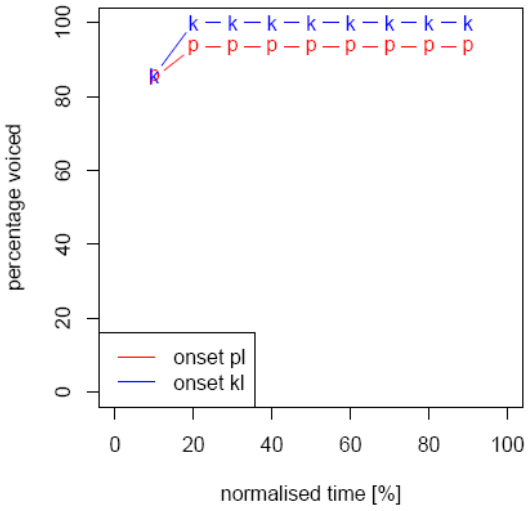


Fig. 57: Voicing probabilities of /l/ in word-initial position with left voiceless stop context (Speaker 2).

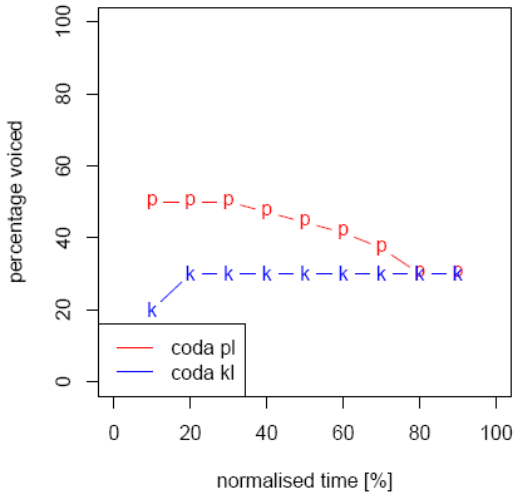


Fig. 58: Voicing probabilities of /l/ in word-final position with left voiceless stop context (Speaker 2).

The third speaker's onset voiceless stop-sonorant clusters show stronger devoicing tendencies for the /r/-clusters, when compared to the profiles of the other speakers. As demonstrated in Figure 59, the voicing probabilities of the sonorants in /pr/ and /kr/ clusters start

slightly below 80% of voicing and reach full voicing after 60% of their time duration. The coda voicing probabilities (see Fig. 60) resemble the previous patterns: where the sonorants in /pr/ clusters devoice almost completely, whereas rhotics in /kr/ clusters start their probabilities from little below 30% and reach complete devoicing after 80% of their time duration. The voicing profiles of onset /pl/ and /kl/ sonorants (see Fig. 61) exhibit more abruptnesses in their voicing probabilities, reaching a stable level of 80% of voicing (/pl/) and 100% (/kl/) after 20% of their time duration. Worth noting is the fact that initial devoicing of the rhotics in the /pr/ cluster reaches slightly more than 40%. Finally, the coda clusters (Fig. 62) demonstrate relatively minor devoicing when compared to the previous coda profiles, where the sonorants in /pl/ clusters vary between 50% and 60% of voicing, while sonorants in /kl/ clusters start their voicing probabilitiy from slightly more than 20%, reaching 50% of voicing by the end of their time duration.

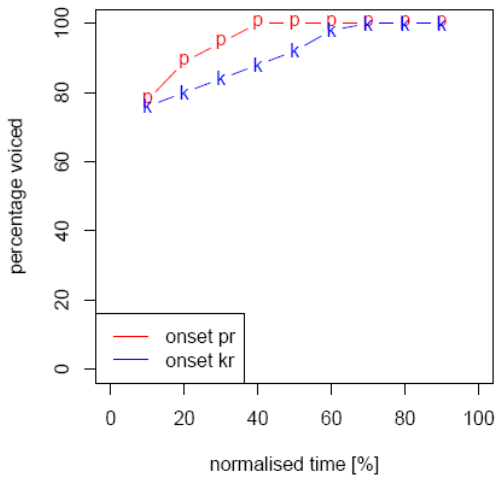


Fig. 59: Voicing probabilities of /r/ in word-initial position with left voiceless stop context (Speaker 3).

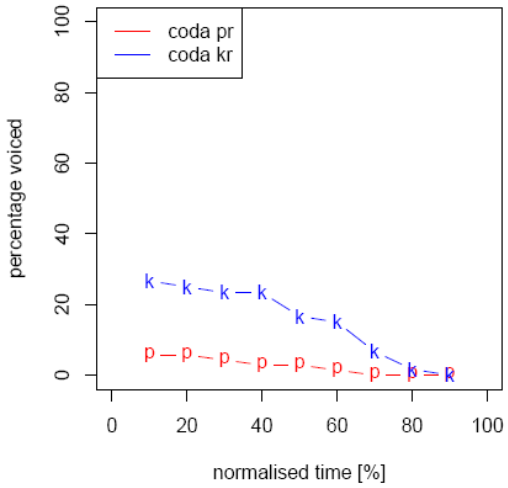


Fig. 60: Voicing probabilities of /r/ in word-final position with left voiceless stop context (Speaker 3).

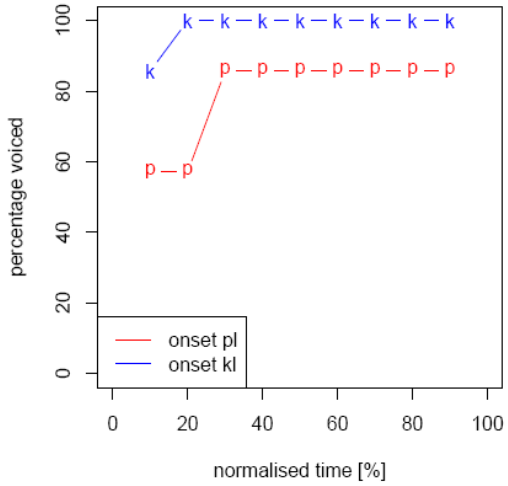


Fig. 61: Voicing probabilities of /l/ in word-initial position with left voiceless stop context (Speaker 3).

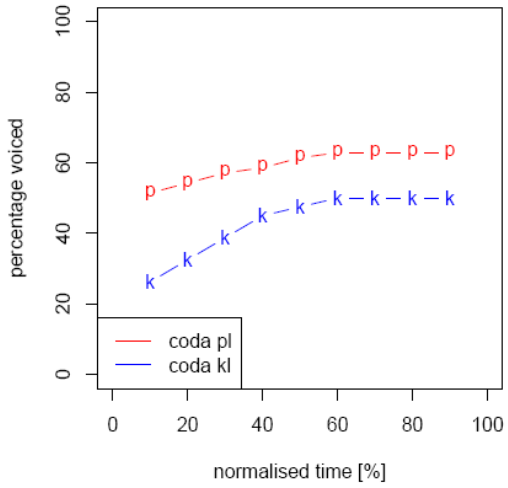


Fig. 62: Voicing probabilities of /l/ in word-final position with left-voiceless stop context (Speaker 3).

5.3.2 Articulatory profiles

5.3.2.1 Gestural coordination of C2

With regard to the previously described studies (Hermes et al. 2008; Nam et al. 2009), the articulatory investigation of Polish sonorants in clusters included the examination of the gestural coordination of the consonant adjacent to the vowel in sequences $C_1\mathbf{C_2}V$ and $V\mathbf{C_1}C_2$. In the figures, gray bars indicate shifts of single consonants in words like 'rabin' (rabbi), while black bars indicate stop-sonorant clusters in words like 'krasić' (to flavor). These are presented in the order Speaker 1 to Speaker 3, with the mean value at the end on the time axis (in ms).

Figure 63 illustrates a rightward shift of the C_2 in the onset of $C_1\mathbf{C_2}V$ cluster when compared to the CV onset sequences. Based on these results, it can be seen that the rightmost consonant in the stop-sonorant clusters undergoes a rightward shift with regard to the vowel

target compared to a single consonant. On the other hand, the C_1 in the stop-sonorant coda V$\mathbf{C_1}C_2$ cluster (Fig.64) shifts leftwards to make room for the added consonant.

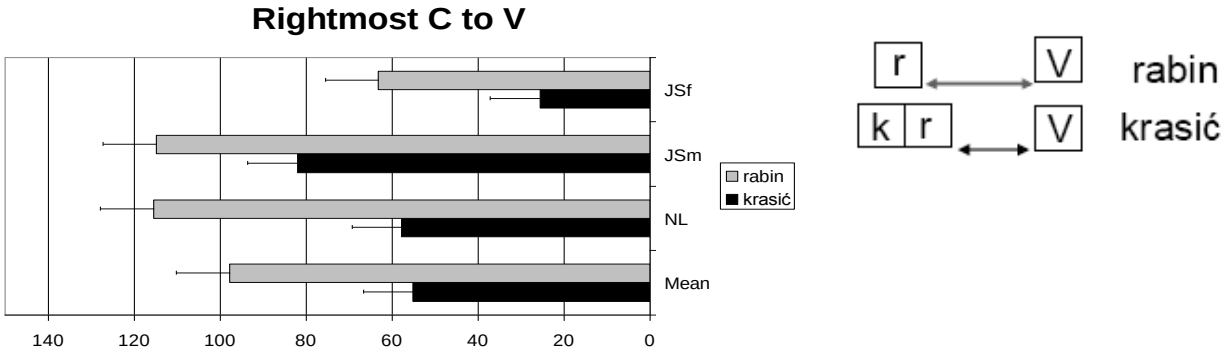


Fig.63: Gestural coordination of the rightmost consonant with regard to the adjacent vowel (onset position).

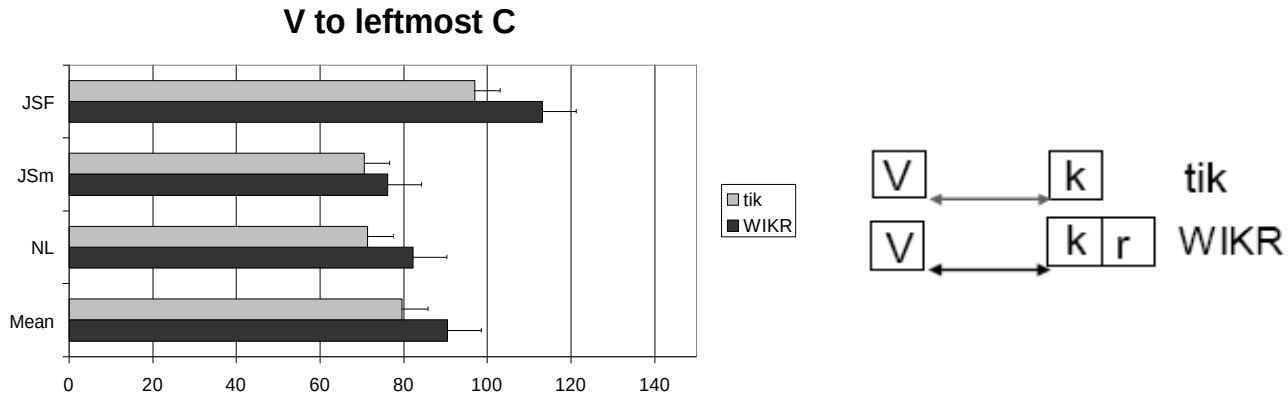


Fig.64: Gestural coordination of the leftmost consonant with regard to the adjacent vowel (coda position).

5.3.2.2 Gestural coordination between C1 and C2

The evidence of the rightward and leftward consonant shifts in onset and coda clusters leads to the question of whether these movements also differ with regard to the various stop consonants and if this coordination leads to the C-center effect.

In word-initial position there is a significant leftward shift of the /p/ consonant in CCV sequences (see Fig.65), like in ‘pranie’ (laundry). The observed tendency is stronger in the recording of the third speaker, whereas for speakers 1 and 2 the extent of the shift is similarly smaller. Rhotics in the onset position, as single consonants and consonants in clusters, demonstrate a rightward shift with regard to the vowel target. In the coda positions (see Fig. 66), words contaiting /pr/ sequences exhibit leftward shifts of the stop towards the vowel target and a rightward shift of the rhotic. The movement of the stop has the strongest effect in the recording of the third speaker, with only slight shifts for the first and second ones. The rightward shifts of word-final /r/, as in ‘Cypr’ (Cyprus), exhibit much stronger tendencies to an almost equal extent for all speakers.

Word-initial /kr/ segments as in ‘krasić’ (to flavor) show similar patterns to the words containing the /p/ consonant, where the leftward shift of the stop is slightly smaller than in the /p/ examples. Once again, this was strongest for the third speaker (see Fig. 67). The rightward shifts of the rhotics maintain the same level for the speakers. In word-final position, there is a significantly stronger rightward shift of the rhotics. The stops exhibit rightward shifts towards the rhotic and not towards the vowel target (see Fig.68).

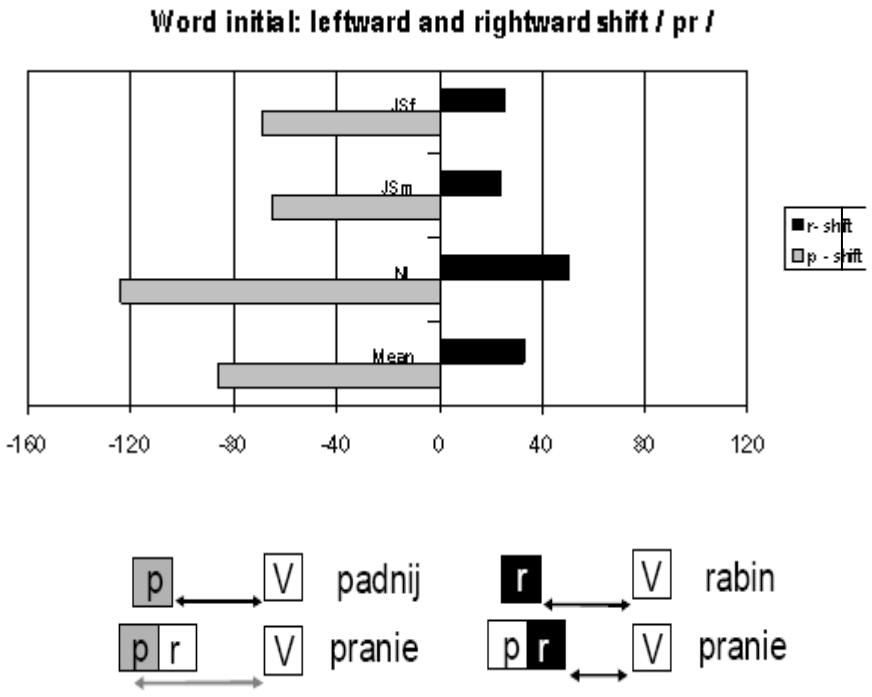


Fig. 65: Gestural coordination between the C1 and C2 in /pr/ onset clusters.

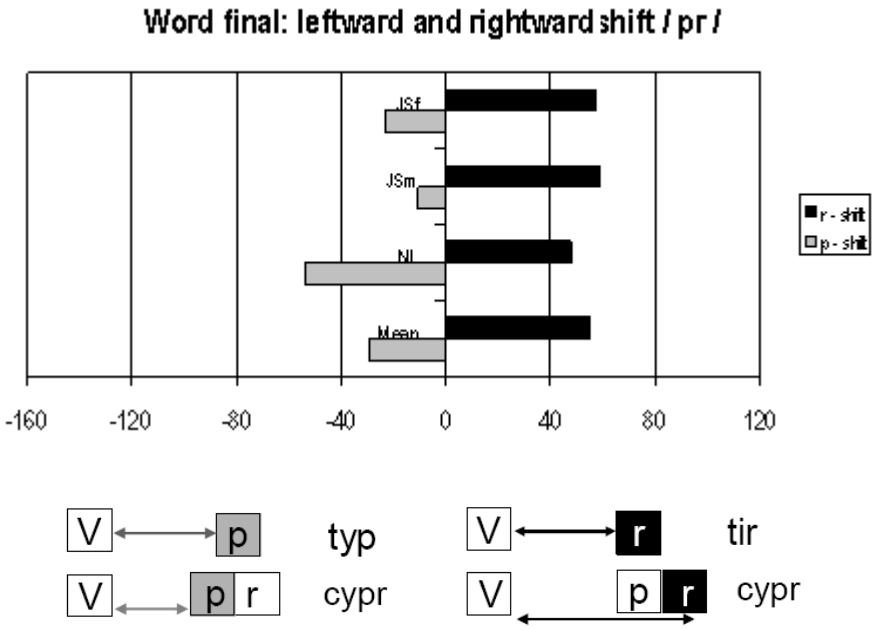


Fig. 66: Gestural coordination between the C1 and C2 in /pr/ coda clusters.

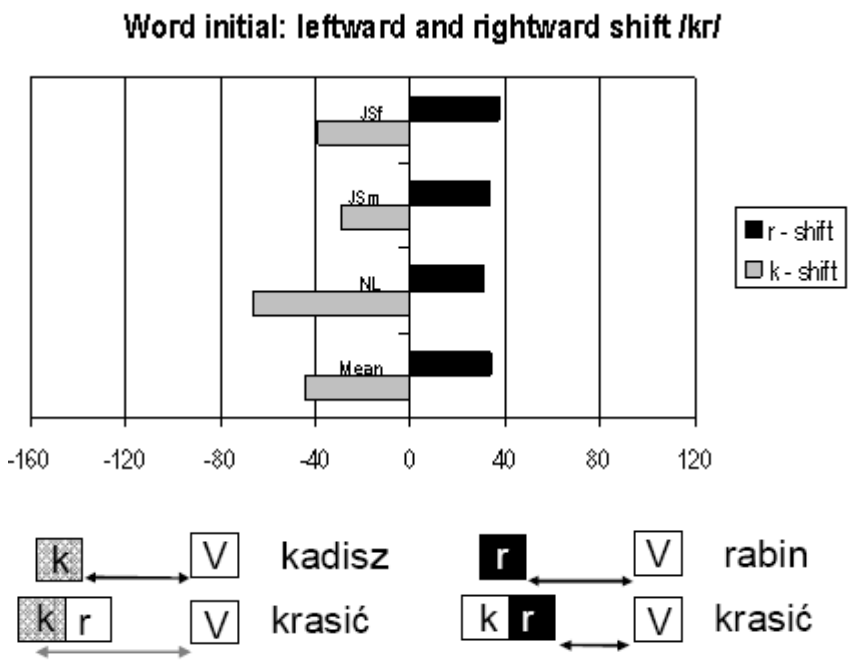


Fig. 67: Gestural coordination between the C1 and C2 in /kr/ onset clusters.

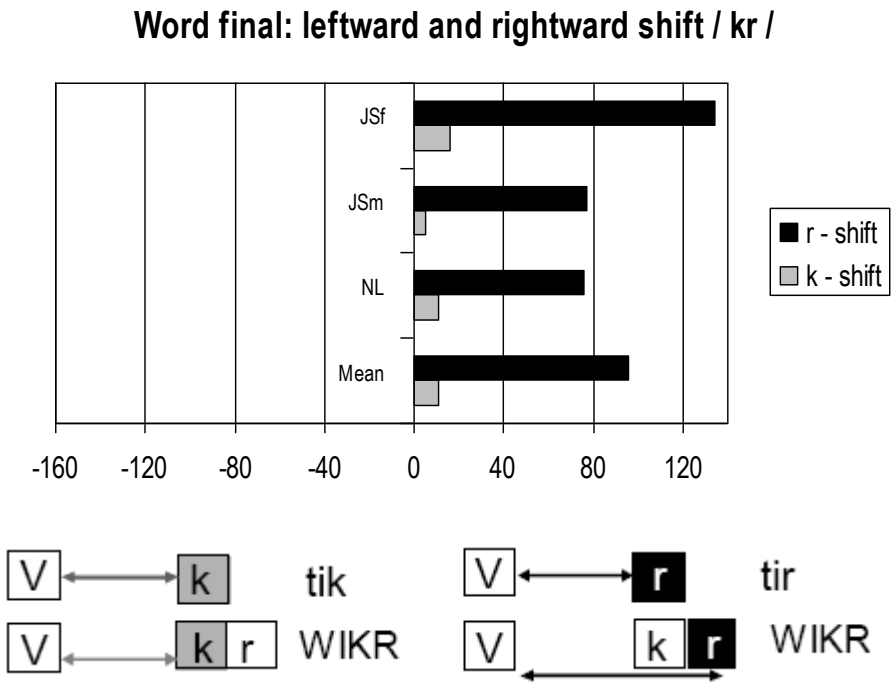


Fig. 68: Gestural coordination between the C1 and C2 in /kr/ coda clusters.

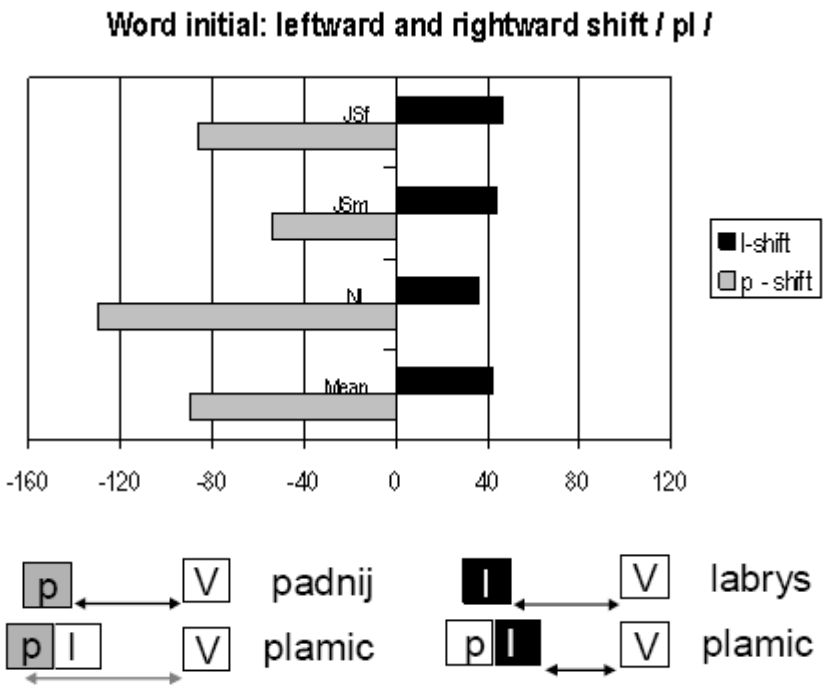


Fig. 69: Gestural coordination between the C1 and C2 in /pl/ onset clusters.

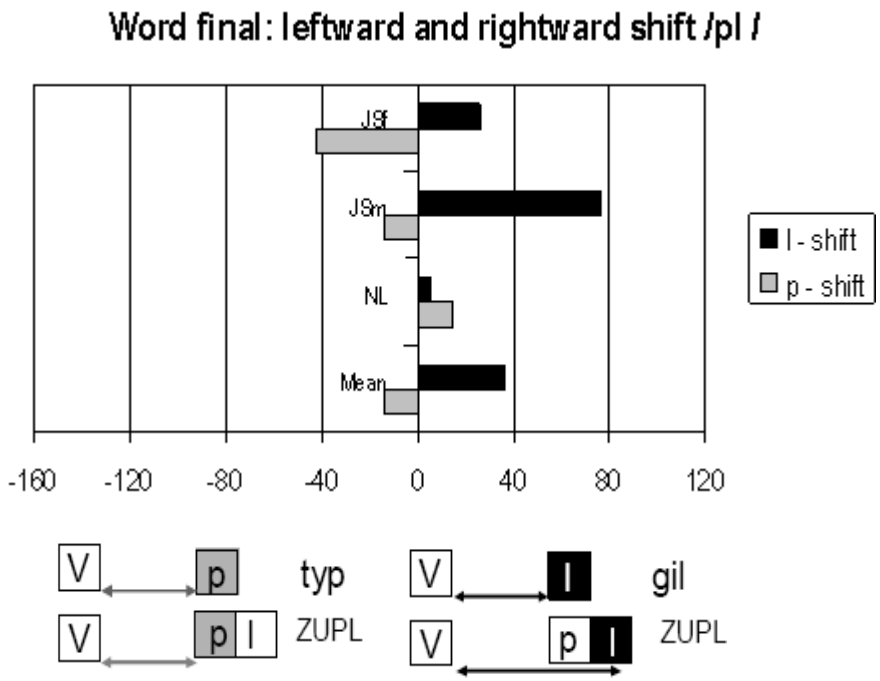


Fig. 70: Gestural coordination between the C1 and C2 in /pl/ coda clusters.

The articulatory profiles of the /p/ in /pl/ sequences in word-initial position exhibit large leftward shifts of the stops, particularly for the third speaker (see Fig.69). Liquids in clusters, however, demonstrate rightward shifts to make room for the added consonant, where the extent of the movement is similar for all three speakers. Coordination between C1 and C2 in word-final position shows irregular patterns (see Fig. 70). Stops from the recordings of the first two speakers exhibit leftward shifts towards the vowel target, whereas the /p/'s from the recordings of the third speaker shift rightwards. Liquids show similar to one another shifting tendencies, having the largest extent in the case of the second speaker.

The results of the articulatory investigation of Polish sonorants in stop-sonorant clusters demonstrate C-center-like coordination in word-initial position, where in the C_1C_2V, the first consonant exhibits a stable pattern in the leftward shifts, whereas the second consonant always shifts towards the vowel target making room for the added consonant. In word-final position, the second consonant in the VC_1C_2 sequence demonstrates a stable rightward shift pattern. Stops, however, vary in their movement directions and display both leftward and rightward shifts.

An ANOVA table presented below (Tab. 9) illustrates the differences in results among all speakers, all stop-sonorant cluster types and in onset and coda positions. It is the word position of the consonant clusters which appears to exhibit the most significant role in the articulatory sonorant specification.

The findings presented in this chapter exhibit in-phase coupling relations of Polish liquids in word-initial positions and out-of-phase configurations in word-final positions. These articulatory patterns contribute to a better understanding of voicing behavior in Polish, providing a possible explanation for devoicing tendencies in word-final positions, which seem to be due to three factors: the contextual surrounding, [voice] unlicensing and presence/lack of articulatory

bounding. The influence of the configuration of articulatory gestures is an area ripe for further investigation in other languages (e.g. French) in order to strengthen this hypothesis.

	Sum sq	Mean sq	F value	Pr (>F)	
Position	10.3765	10.3765	133.4211	**< 2e-16**	*******
Sonorant	0.2721	0.0907	1.1663	0.32477	
Speaker	0.6895	0.3447	4.4327	0.01352	*
Position vs. sonorant	0.7671	0.2557	3.2879	0.02252	*
Position vs. speaker	0.0215	0.0108	0.1384	0.87083	
Sonorant vs. speaker	1.1305	0.1884	2.4227	0.02914	*
Position vs. Sonorant vs. Speaker	0.9259	0.1543	1.9842	0.07152	.

Signif. codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Tab. 9: ANOVA results for the acoustic-articulatory EMA study.

CHAPTER 6

Conclusions & Discussion

6.1 Voicing profiles

The results of the cross-linguistic voicing investigation presented in Chapter 4 show diversity in the behavior of sonorants in German, Polish, American English and French. However, it has been observed that two main factors influence voicing probabilities in these languages, namely the acoustic/articulatory context and word position variation. Thanks to the automatic method of voicing extraction originally proposed by Möbius (2004), it has been possible to perform voicing analysis on large speech databases and incorporate a wide range of phonetic and structural features relating to the environment of the sonorants using the IMS Festival speech synthesis system tool (IMS Festival 2003). It has been observed that temporal frame-by-frame voicing resolution provides a new kind of analysis that enables a more exact investigation of consonants, including their phonotactic and structural relations within the syllable and the underlying articulatory gesture.

The strongest predisposition for devoicing was found in Polish and French liquids, where [r] and [R] devoiced almost completely due to their word-final position and the preceding voiceless segment. Results from German (Möbius 2004) and American English showed a relatively lower number of devoiced exemplars, displaying slight differences cross-linguistically (see Fig. 46).

Results of an analysis concerning German sonorants (Möbius 2004), which served as the seminal example for the current study, provided clear evidence for the role the left-hand

segmental context plays in the voicing probabilities of the sonorants. It has been demonstrated that their devoicing takes place only after voiceless obstruents – a result consistent for all sonorants, in particular liquids. Evidence was also provided for the role of the syllable boundary. Sonorants tend to be voiced to a larger extent when separated from the left-hand voiceless segment by a syllable boundary.

The investigation of voicing profiles of Polish sonorants shows the influence of the left acoustic and right linguistic context, as predicted by the CSM (Wade et al. 2010). While a majority of sonorants remained voiced, it is the phonetic and phonological behavior of [r] which remains the focus of this study. As predicted by Gussmann (1992, 2007) the trill exposed to left-hand voiceless context demonstrated devoicing tendencies, but only to a minor extent (a little below 80% of the corpus exemplars). On the other hand, [r] with the same context but in word-final position exhibited full devoicing throughout its time duration. These results provide further evidence for the [voice] licensing mechanisms in Polish proposed by Gussmann (1992, 2007), where voicing of the sonorant is only licensed within a syllable and thus a sonorant in a word-final obstruent-sonorant undergoes desyllabification. Similarly Lombardi (1991, 1995) discusses this phenomenon, explaining devoicing of Polish word-final sonorants in clusters as being the result of the Final Extrametricality and not the adjacency of the Laryngeal Nodes proposed by Rubach (1996). Moreover, articulatory analysis of the clusters shows that it could be gestural coupling of the sonorant which is responsible for its tendencies to devoice.

The results of the investigation of American English sonorants exhibit two tendencies. The laboratory news recordings annotated both automatically and manually show similar patterns in sonorant behavior. Segments [w], [m], [l] and [r] with left voiceless obstruent context start their voicing from a high level between 85% and 97% of exemplars, reaching full voicing

towards the end of their duration. The second pattern observed demonstrates [ŋ] and [r] with left vocalic/sonorant context as being almost fully voiced (between 95% and 100% of the exemplars), devoicing slightly towards the end of their duration (until not less than 85%). It is worth noting that English sonorants do not occur in word-final obstruent-sonorant clusters, a condition which has only been examined in their initial and medial word positions. Thus, it seems to be the left-hand voiceless obstruent context that causes initial devoicing of [w], [m], [r] and [l].

The voicing probabilities of American English sonorants from the radio news part of the Boston University Corpus exhibit a similar behavior for sonorants [w] and [m], which also devoice in their initial part when occurring with the left-hand voiceless obstruent context. Devoicing of the sonorants occurring with the left vocalic/obstruent context seems to increase towards the end of their duration, with the strongest tendency observed for [ŋ].

The results of the French sonorant voicing investigation demonstrate stronger devoicing patterns. Due to its phonotactic constraints, French exhibits devoicing tendencies similar to the Polish ones. It has been observed that sonorants with the left-hand vocalic/sonorant context display final voicing tendencies. An exception is [R], which tends to devoice significantly. Following a voiceless obstruent, all sonorants tend to devoice initially, however only the liquids devoice remarkably. [R] in particular shows large devoicing probabilities which, like the Polish trill, differ with regard to word position: up to 40% of devoicing for initial ones, with nearly 100% of corpus exemplars in final position.

According to previous analyses of phonetic features and their specification (among them Iverson & Salmons 1995; Lombardi 1991, 1995; Gussman 1992, 2007; Browman & Goldstein 1992), languages employ the feature [voice] to distinguish between voiced and voiceless consonants, and the feature [spread glottis] to differentiate the amount of glottal opening during the

production of the consonants. Polish and French use [voice] contrasts with regard to the obstruent and sonorant consonants, whereas Germanic languages like English and German contrast them using [spread glottis] (Iverson & Salmons 1995). Findings concerning the voicing analysis of the four languages show feature licensing in a straightforward way. Similar patterns of sonorant devoicing in Polish and French suggest phonological licensing of [voice], according to which a sonorant segment must belong to a syllable structure in order to be licensed for voicing. The findings concerning the voicing probabilities in word-final positions with left-hand voiceless context for the Polish and French liquids [r] and [R] illustrate the positional dependencies of their licensing capabilities. Following an idea put forward by Gussmann (1992, 2007) and Lombardi (1991, 1995), it is not the adjacency of laryngeal nodes but rather the final extrametricality that leads to both sonorants tending to devoice entirely, being excluded from the word-final syllable, since the left-hand presence of voiceless obstruents causes only minor devoicing of the sonorants in word-initial positions, unlike complete devoicing at the end of the word. Thus, it is the change in word position that controls the voicing probabilities of Polish and French sonorants (with the same left-hand voiceless contexts). On the other hand, the results of voicing profiles of German and American English sonorants confirm the spreading of the feature [spread glottis] from the voiceless obstruent to the following sonorant in the initial part of their duration (Iverson & Salmons 1995). As stated by Möbius (2004), the left-hand segmental context is the most important factor in the devoicing of German sonorants. The same pattern has been observed in the laboratory and in the radio news parts of the corpus of American English, where sonorants (particularly [w] and [m]) preceded by a voiceless obstruent segment tended to devoice until almost 50% of their time duration, reaching full voicing towards the end.

In the studies conducted on German sonorants (Möbius 2004) it was also observed that a syllable boundary intervening between the voiceless obstruent and a following sonorant plays a crucial role. Obstruent-sonorant clusters separated by a syllable boundary exhibit significantly smaller devoicing tendencies in sonorant segments, unlike the ones occurring in a single syllable with left-hand voiceless obstruent. Similar conditions applied to American English sonorants (laboratory news part of the corpus), which display (see Fig. 71) slightly stronger devoicing tendencies of the sonorant [m] when separated by a syllable boundary from the left-hand voiceless obstruent in mid-word position (like in the word 'pacman' [pæk.mən]), than when they occur in one syllable like in the word 'smart' [smʌɹt] (see Fig. 32). However, voicing probabilities of the other sonorants seem to remain independent of syllable-boundary.

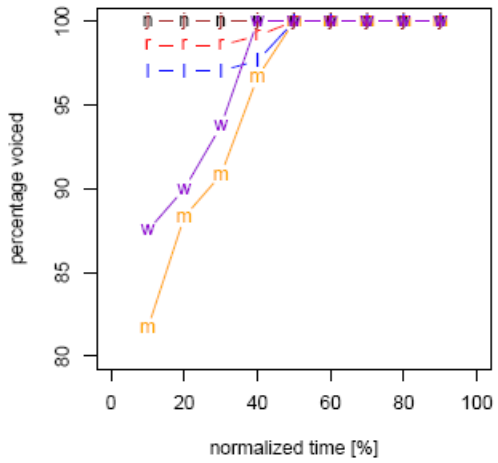


Fig. 71: Voicing profile of American English sonorants (lab news corpus) in word-medial positions with left voiceless obstruent context separated by a syllable boundary.

6.2 Articulatory strengthening

The voicing probabilities of American English sonorants and the initial devoicing of [m] and [w] occurring with the left voiceless context appear to be a consequence of articulatory strengthening

and coarticulation effects. According to Recasens (1989), sounds produced with smaller vocal tract regions like palatals should allow smaller coarticulation effects than the ones produced with larger vocal tract constrictions, like labials and dentals. Furthermore, it has been hypothesized (Stevens 1972, 2010) that the relations between acoustic and articulatory values of distinctive phonetic features have quantal properties, i.e. they exhibit nonlinear relations when a particular articulatory dimension is manipulated with regards to the acoustic parameter. Thus, the resistance to coarticulation increases for sounds with narrow quantal regions and diminishes for sounds with large quantal areas.

Figure 72 shows the representation of formant placement during the production of a sound with a large quantal area, like that in [ɑ]. Sounds with larger vocal tract regions allow larger coarticulatory effects (yellow space) due to the manner of their articulation, which Fant (1970) and Stevens (1997) have illustrated with a two-tube model (see Fig. 73). By contrast, sounds with smaller quantal regions (yellow space) and larger vocal constrictions (see Fig. 74), like palatals, are modeled with three-tube models (Fig. 75). Following these ideas, devoicing of bilabial [w] and [m] preceded by a voiceless obstruent in American English would appear to be the result of lesser coarticulatory resistance, which allows these segments to undergo initial devoicing more than other sonorants in the same context. As stated by Stevens and Keyser (2010: 15), "(…) quantal acoustic/articulatory relation underlies each distinctive feature, and consequently each feature can be said to be based on a defining articulatory range and defining acoustic attribute."

Recasens et al. (1996: 165-185) reported coarticulatory effects for German non-velarized /l/. They compared German articulation patterns with Catalan velarized /l/ and found that the latter realization is subjected to less articulatory control, whereas the former allows more alveolar and palatal coarticulation. Their conclusions are based on the notion that German and Catalan

differ in the degree of velarization, which the authors compare to English dark /l/ and clean /l/ realizations.

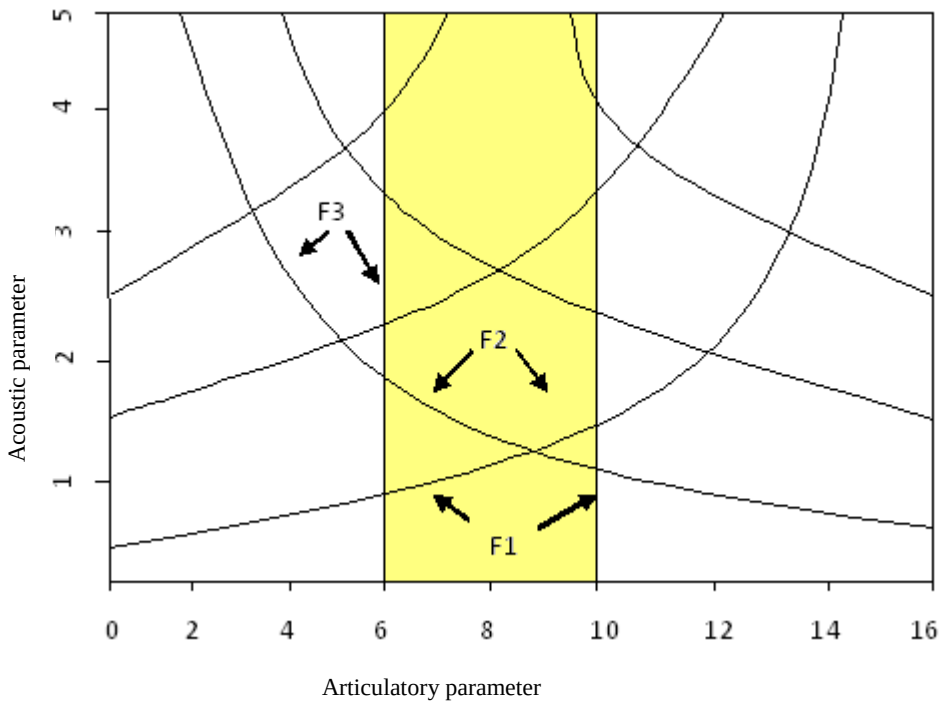


Fig.72: Representation of the sounds with large quantal areas (yellow space) less resistant to coarticulation (figure after Johnson 1997).

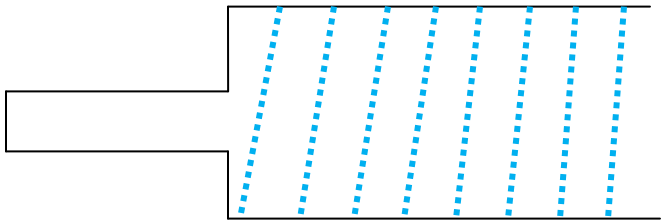


Fig.73: Two-tube model of a vocal tract formation during production of [ɑ]. Dashed lines show the region where the frequency is determined, i.e. front cavity.

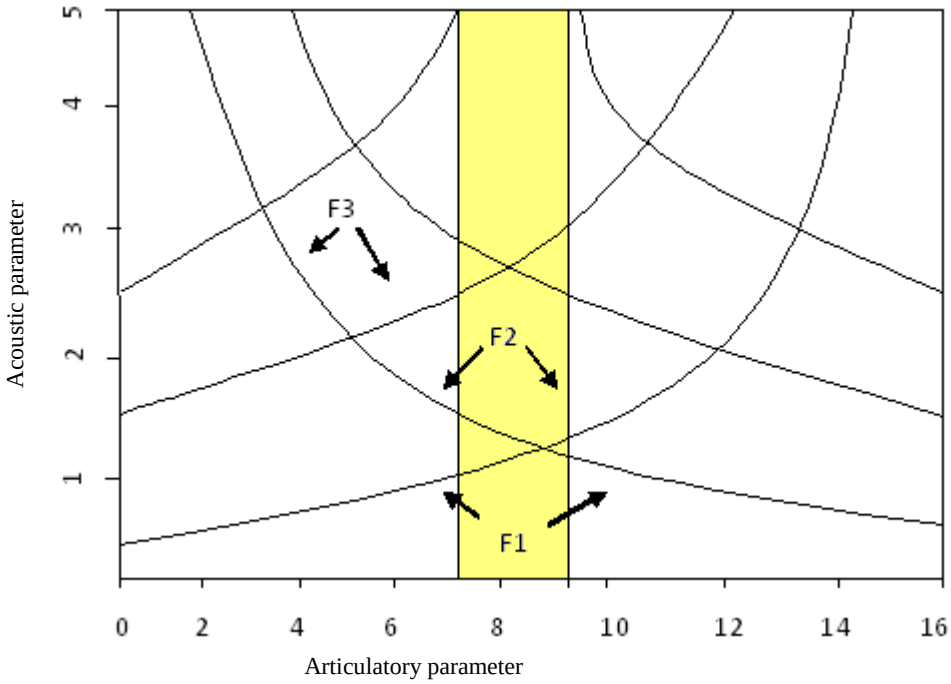


Fig.74: Representation of the sounds with narrow quantal areas (yellow space) more resistant to coarticulation (figure after Johnson 1997).

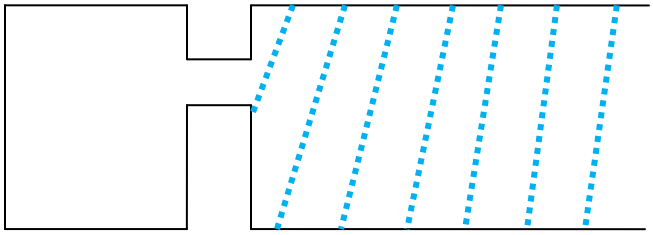


Fig.75: Three-tube model of vocal tract formation during production of palatals. Dashed lines show the region where the frequency is determined, i.e. front cavity.

"Differences in degree of dorsopalatal coarticulation between the two varieties of /l/ are mostly associated with the sensitivity levels to tongue dorsum raising effects exerted by adjacent /i/ in /ili/ sequence. A trend towards more context-independent articulatory configuration for velarized

/l/ is consistent with the former realization showing less dorsopalatal contact than the latter in the sequence /ulu/. Both consonantal varieties show no dorsopalatal contact in the sequence /ala/."(Recasens et al. 1996: 16)

Sproat and Fujimura (1993: 291-312) investigated articulatory differences between clear and dark /l/ in American English. The authors analyzed realizations of the two allophones by presenting acoustic and X-ray microbeam data. Their results indicate that the darker /l/ variant exhibits greater retraction and lowering of the tongue dorsum. Moreover, in the darker variant, dorsal retraction and lowering extremum relative to the advancement extremum occurs earlier than in the clear /l/. As a result, the authors claim that the two lateral allophones of English should not be treated as categorically distinct phonological entities. Additionally, the authors state that production of the dark and clear /l/ variants involves two lingual gestures: apical and dorsal, with application dependent on the syllable position. The dorsal (also called 'vocalic') gesture occurs in the nucleus of the syllable, whereas the apical (also called 'consonantal') appears in the margin of the syllable (Sproat and Fujimura 1993).

Figure 76 presents voicing probability of clear /l/ exemplars[29] from the Boston University Corpus. Up to a point of 60% of its time duration, the 'clear' allophone tends to devoice from around 68%, reaching 100% towards the second half of its duration. By contrast, voicing probability of the dark /l/ presented in Figure 77 demonstrates much smaller devoicing probabilities, which start shortly after 60% of the time duration, not reaching more than 20% of final devoicing.

[29] The exemplars from the corpus were extracted based on the Clark and Yallop (2007: 96) description of clear and dark /l/ differentiation.

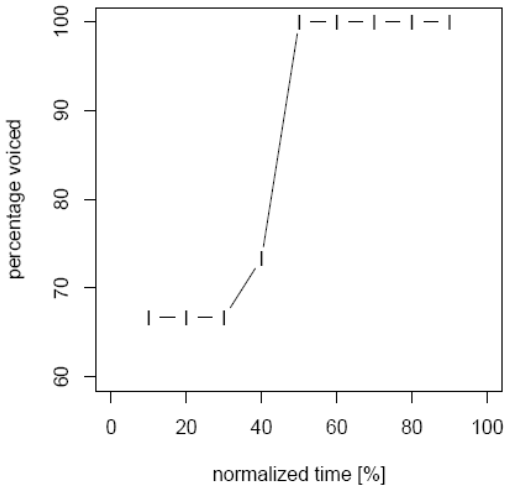


Fig.76: Voicing profile of the American English clear /l/.

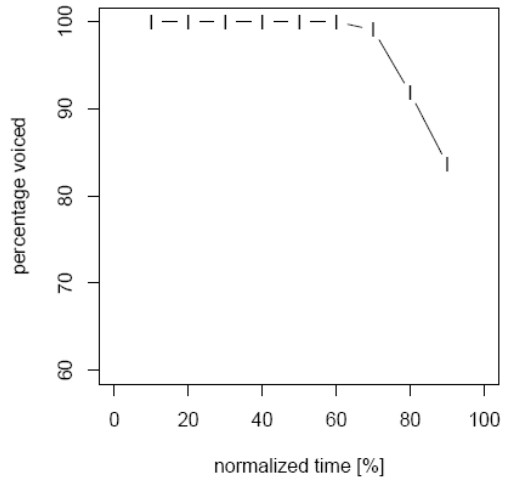


Fig.77: Voicing profile of the American English dark /l/.

Thus clear /l/ seems to be less coarticulatory resistant due to lesser vocal tract constrictions, while dark /l/ exhibits stronger coarticulatory resistance, which finds its repercussions in smaller devoicing tendencies. These issues would also benefit from further investigation with large speech materials.

6.3 Feature [voice] & context specification

Recent work by Kingston, Lahiri and Diehl (2010) has shed a new light onto feature licensing, according to which distinctive phonetic features do not have any particular phonetic definitions but vary in terms of phonetic realization *as a function of context*. The authors claim that the distinctive features are better characterized by their contextual phonetic behavior and historic development than by their inherent descriptive values. In their proposal, the focus is on the analysis of laryngeal contrasts for the feature [voice] instead of [spread glottis] in the Germanic languages. Kingston and colleagues (2010) outlined three arguments for this kind of analysis. First, "distinctive features do not have essential phonetic definitions. (…) it is nonetheless still possible to define features phonetically, because variable acoustic properties integrate perceptually" (Kingston, Diehl & Lahiti 2010: 56). They argue that differences in the pronunciation of stops contrasting for [voice] vary across contexts and languages. "(…) no context provides better phonetic evidence than any other for deciding which distinctive feature the stops contrast for" (Kingston, Diehl & Lahiti, 2010: 7). The authors also claim that speakers are more likely to independently control many articulations which might be the consequences of the [voice] contrast than to produce a singly controlled articulation. It is also stated that F0 differences might be intentionally produced in order to co-vary with the production of voicing differences so as to preserve perceptual integration. Kingston, Diehl and Lahiti (2010) claim that it is the integration of co-varying acoustic consequences of a context and manners of articulation that defines the phonetic contrast. As their second argument, the authors argue that historical changes of the present-day Germanic languages can be described in a better way using [voice] rather than [spread glottis]. This statement is supported by the fact that English and German have different diachronic roots in the present-day form of their /p, t, k/ stops, as the English forms did

not undergo the Old High German or Second Consonant Shift, in contrast to the German forms. The authors claim that this sound change caused almost all voiceless stops in Old High German to undergo affrication or spiranization, which means that the present German /p, t, k/ do not correspond to the English ones. The third and final argument of Kingston et al. (2010) is that the phonetic and phonological patterns of Dutch, English and German are similar, assuming that Dutch is one of the West Germanic languages that undoubtedly contrasts its obstruents [voice]. This indicates the same contrast model for English and German, in which the laryngeal contrast identity still remains under discussion.

Another work on phonological features was recently presented by Hawkins (2010: 60-89). The author states that "speech perception (…) relies on a good match between memorized experience and current sensation: when sensation meshes with expectations, listeners believe they perceive 'real' linguistic objects in spite of possibly severe variation and degradation in the acoustic signal" (Hawkins 2010: 60). In her study, the author analyzed acoustic-phonetic correlates of [voice] in stop consonants, classified as voiced or voiceless. She claimed that the phonological contrast [+/- voice] does not depend on physical properties (i.e. vocal fold vibration, periodicity), instead acoustic-phonetic correlates can only be considered in the context of perceptual processing of "complex auditory patterns which extend over a syllable or more" (Hawkins 2010: 64). She also argues that during speech perception the perceptual system focuses more on the properties of the speech signal than on the extent to which this signal resembles a clear speech-like occurrence. As an example, Hawkins (2010) describes the way in which synthetic speech works, where the acoustic patterns are understood by the listener as long as they preserve critical properties/cues. In distinguishing the stimulus and its context, the author points out the relevance of amplitude rate and its decay during [voice] categorization on the boundary

between the sonorant-obstruent occurrences. It is claimed that the difference in the abruptness of the amplitude at the sonorant offset before a voiced/voiceless obstruent determines cue perception: a slower rate is associated with [+ voice] (like in the word 'led', unlike 'let' with [- voice] where the amplitude decay is more abrupt). The [-voice] obstruent codas, therefore, have shorter preceding sonorants, while [+voice] obstruents have the opposite. Thus, the author claims that the phonological status of a coda obstruent influences the status of the entire syllable. Hawkins (2010) hypothesizes that the cue argument is the subjective phonological feature perception, determined by a memory-sensation relation, which is said to be relevant in further audio-visual processing.

It has also been claimed (Alexiadou 2010) that the process of specification (a model of which has been developed in the SFB 732[30]) of an underspecified representation allows for a disambiguation of the constraints and conditions provided by various-level contexts across languages. "Ambiguation is a property of a natural language which normally involves a choice between two (or more) specific meanings that make sense in a particular context" (Alexiadou 2010: 19). In normal speech, phonological representations rendered in kinematic activity are trained by the processes of specification and underspecification (Dogil 2010: 343-379). In the view of the incremental context specification and Exemplar Theory, a model of speech representation has been proposed (Dogil 2010) in which the phonetic and phonological regions emerge as clouds in the perceptual space of the speakers. As presented in Figure 78, speech exemplars are not only composed of the realization of concrete tokens, but also contain internally acquired realizations of the analysis-by-synthesis processes. It has been posited (Dogil 2010) that

[30] SBF 732 – Sonderforschungsbereich 732: Incremental Specification in Context, financed by the German National Science Foundation (DFG).

the hearer begins his/her still underspecified phonemic/prosodic category analysis after having begun the process of the already fully specified acoustic analysis of the most relevant landmarks and pivots (see Stevens 2002, 2005). If there is a degree of uncertainty regarding feature values, a process of re-analysis of the input takes place, in which the local and non-local contextual information is taken into account. Finally, the underspecified feature representation is matched with the entries of the lexicon. This completes the incremental process of underspecification. In the next step, fully specified categories are reached by an internal analysis-by-synthesis, and the underspecified category is internally processed by the hearer. It is claimed (Dogil 2010) that the relevant landmarks and contextual information are used in the process of specification, in which the available context (possibly richer than the context steming from the analysis process) comprises vast data-like information on prosodic voice quality along with tonal and temporal cues as well as data on syllable and prosodic structure, syllabic stress and discourse structure (Dogil 2010). A further development is the successfully implemented Context Sequence Model (Wade et al. 2010), which specifies a number of assumptions from the SFB Model, particularly context information (left acoustic and right linguistic contexts).

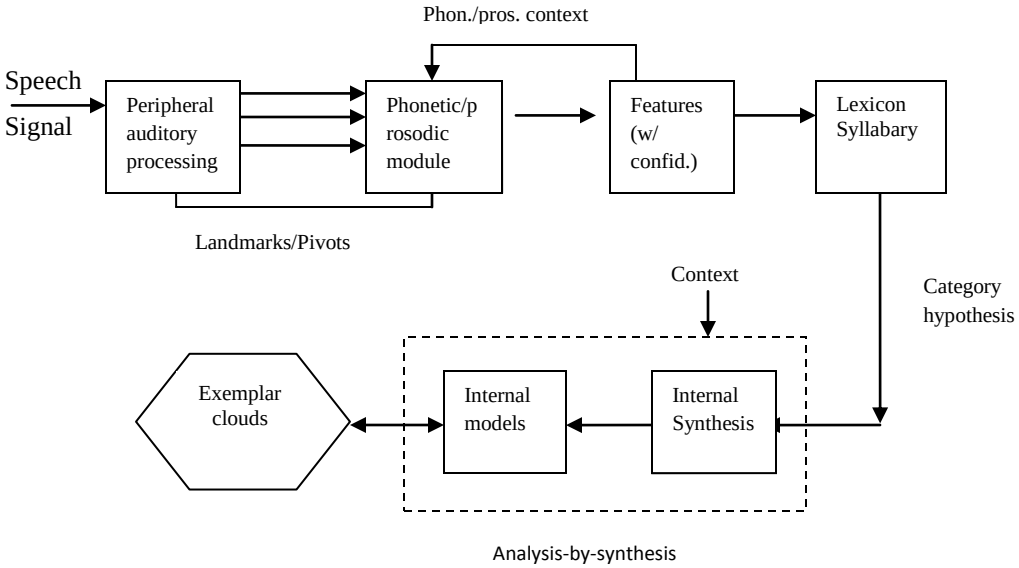


Fig.78. Incremental Specification in Context Model (Dogil 2010: 356)

6.4 Articulatory Profiles

Articulatory studies conducted on Polish sonorants in voiceless stop-sonorant clusters have shown that Polish allows for complex onsets due to the C-center type coordination found in word-initial clusters. Their word-final counterparts, by contrast, exhibit lesser coupling tendencies with different time patternings of the first consonant in the VCC clusters.

These tendencies have been confirmed by previous studies on Italian onset CV and CCV clusters, where Hermes et al. (2008) observed a rightward shift tendency, indicating significant changes for two out of three target word sets. According to the authors, the consonant shifts considerably towards the vowel in words like /lina/ and /plina/ by 72ms on average, in /rima/ and /prima/ by 60ms on average, but in /rema/ and /prema/ by 51ms on average. Thus, in order to make room for the added consonant, the rightmost consonant shifts rightwards. This serves as evidence for the C-center effect in non-sibilant onset clusters in Italian, where the distances from the C-center to the vowel remain stable, whereas distances between the rightmost consonant and a vowel decrease, demonstrating the formation of complex onsets (CCV).

Nam et al. (2009), who conducted studies using a coupling oscillator model, investigated the timing of rightward and leftward consonant shifts, seeking the C-center type of coordination in English single word-initial consonants (/p/, /s/ and /l/), as well as CC clusters (/sp/ and /pl/). The authors observed both rightward and leftward shifts in the onset /sp/ and /pl/ clusters (which is said to provide proof of the existence of competitive, multiply-linked onsets), which were produced with a lack of consistency regarding speaker/accent factors, apart from the leftward /pl/ asymmetries. After having employed a quantitative model for the leftward, weak asymmetry in the /sp/ clusters and strong asymmetry in the /pl/ clusters, Nam et al. (2009) stated that "the English data support a view that patterns of coupling that seem to be regular aspects of

the phonological knowledge of a language (in that they generalize across individual talkers) can be well represented qualitatively (discretely) in terms of the set of the edges that define the coupling graph. Individual differences in coordination patterns can then be modeled by quantitative variation in the coupling strength values of the graph's edges" (Nam et al. 2009: 303). Additionally, the authors investigated onset CC clusters in Georgian, which allows two to three stop consonants in word-initial position. Like English onset clusters, Georgian exhibits the competitive, multiply-linked structure of word-initial syllables, whose existence has been hypothesized by the authors using the coupled oscillator model. Based on the results from the EMA recordings, Goldtein et al. (2009) demonstrated a 19ms rightward shift in /t$^{s'}$k'/ clusters and a 7ms rightward shift in /k'b/ sequences. The direction of the shifts was said to be consistent across the speakers. The authors conclude that results from Georgian recordings illustrate a tendency in which all onset consonants display in-phase coupling relations with the vowel (in the /t$^{s'}$k'/ examples). In the /k'b/ sequences it has been observed that it is the closure of the initial consonant that exhibits a relation to the vowel target.

Another study by Nam (2007: 483-506) illustrates the coupling of articulatory gestures with regards to syllable structure and word position. The author shows the timing asymmetries between the onset and coda clusters, pointing out that clusters containing stops can be further specified into closure and release gestures. The author also specifies kinds of gestural coupling in the onset position with relation to the vowel target as synchronous, multiple and competitive, whereas in the coda position they are sequential, single and non-competitive. Nam (2007) points out that there are various temporal asymmetries, like the C-center effect, between onset and coda positions, in which it is the onset-placed consonants that exhibit more stability and intergestural phasing. The author proposes a model in which gestural scores are the output of the intergestural

level, which is understood as 'a dynamics of planning', i.e. it describes the relations between the activation of the gestures used in an utterance, their shapes and the duration of each gesture. In the model, timing of the consonant gestures in relation to the vowel target is modulated when multiple C-gestures are inserted in the onset. According to the author, the overall timing of the consonant gestures in the onset is preserved in its relation to the vowel target (a regularity which is not observed in the coda position). The author also proposes a moraic interpretation of the intervals between articulatory gestures, claiming that "the emergence of sequential coordination between a pair of gestures in a gestural score, which is the output of the intergestural timing model, is potentially equivalent to a single mora" (Nam 2007: 500), where the onsets exhibit weightless tendencies in that they allow the formation of distinctive coupling structures. In conclusion, the author states that it is the intergestural coordination in the onset and its weightless nature that provides stability in gesture phasing, a process which does not take place in the coda consonants.

Honorof and Browman (1995: 552-555) conducted articulatory and acoustic studies on American English accented monosyllables in one-, two- and three-consonant pre- and post-vocalic consonant clusters. The findings presented by the authors support an idea of Browman and Goldstein (1988: 140-155) in that they demonstrate greater and more stable C-center organization of pre-vocalic (word-initial) consonants, while consonants in post-vocalic position (word-final) exhibit a left-edge[31] articulatory organization, which is shown to be more stable than the C-center one.

[31] Honorof & Browman (1995) explain left and right edge organization following Browman and Goldstein (1988): "(…) judging by patterns of articulatory stability, the C-center of a pre-vocalic consonant or consonant cluster is more tightly coordinated with the vowel than is either the left-edge (LE) of the first pre-vocalic consonant plateau or the right edge (RE) of the last one." Honorof & Browman (1995: 552)

Studies presented in this section show a general tendency in which articulatory coordination within syllables differ with regard to the word position of the investigated segments. As pointed out by Hermes et al. (2008), Mücke et al. (2009) and Goldstein (2009), word-initial consonant clusters in Italian, English and Georgian exhibit a gestural time patterning similar to those in Polish, where there is clear evidence for the C-center-like coordination of C1 (which shows leftward shifts) and C2 (rightward shifts). On the other hand, as has been shown in other studies (Browman & Goldstein 1988; Honoref & Browman 1995; Nam 2007), there is no clear C-center patterning in word-final consonant clusters, which allow more variability in relative timing – a tendency which has also been observed in the investigation of Polish coda VCC sequences.

6.5 Discussion

The results of the voicing and articulatory investigations presented in this dissertation demonstrate complex dependencies of [voice] licensing in sonorants in Polish, German, American English and French. It is claimed that the described devoicing patterns result from *context specifications* and vary due to *contextual changes*. Thus, it seems reasonable to claim that descriptions concerning [voice] contrasts should not take the context factor for granted.

From the perspective of the investigation of voicing probabilities, it has been demonstrated that Polish and French rhotics devoice fully when they occur word-finally in the left-hand voiceless obstruent context. Moreover, smaller but significant devoicing tendencies of rhotics and other sonorants in word-initial and medial positions with left-hand voiceless obstruent context have been observed. By contrast, sonorants occurring in word-initial, medial and final positions in the intervocalic contexts exhibited no significant devoicing tendencies. Due to similar phonotactic constraints, Polish and French word-final voiceless obstruent-sonorant clusters show similar devoicing patterns, which have been analyzed through the processes of

desyllabification of the word-final sonorant and Final Extrametricality as described by Gussmann (1992, 2007) and Lombardi (1995). The voicing profiles of German and American English sonorants demonstrate the strongest variability in the left-hand voiceless obstruent context in word-initial and word-medial positions, exhibiting significantly smaller devoicing tendencies in vocalic contexts. Moreover, the devoicing of German sonorants increases when voiceless obstruent-sonorant clusters belong to one syllable (i.e. obstruent and sonorant are not separated by a syllable boundary). In the investigation concerning American English sonorants, various patterns were observed, but all share the same tendency of devoicing in the left-hand voiceless obstruent context due to the transfer of [spread glottis] from the obstruent to the sonorant, or to put it in another way, the lack of [voice] spreading from the sonorant to the obstruent. The preservation of voicing due to articulatory strength was also presented using the example of clear and dark /l/.

In Chapter 4, the general question concerning the behavior of the voicing probabilities of sonorants was narrowed down to the investigation of the correlation between voicing and articulatory profiles in Polish voiceless stop-sonorant onset and coda clusters. Results from the acoustic recordings conducted together with the EMA study demonstrated similar tendencies to those from the Polish voicing investigation described in the third chapter, where devoicing patterns were found to be mostly present in word-final sonorants with left-hand voiceless obstruent context. Articulatory profiles obtained in parallel in the same EMA study illustrate C-center like coordination in voiceless stop-sonorant C_1C_2V clusters in word-initial positions, where C_1 shifts leftwards and C_2 rightwards. In their word-final VC_1C_2 counterparts, no C-center cluster organization has been observed. Despite the rightward shifts of C_2, no clear pattern for C_1 has been found.

These results seem to support the contextual specification of voicing governing Polish sonorants. In both profiles it is the coda position in which voicing probabilities and articulatory coupling probabilities exhibit the greatest irregularity. From the articulatory perspective, word-initial obstruent-sonorant clusters show the clearest association with the vowel target, which does not exist in the coda position, where it is only the second coda consonant (the sonorant) that demonstrates stable latencies with regard to the vowel. From a phonological perspective, the presence of the voiceless segment in the coda stop-sonorant cluster determines its devoicing tendencies due to phonological unlicensing for [voice].

APPENDIX

1. Transcription tables for the four investigated corpuses:

Boston Radio Corpus Transcription	SAMPA	IPA
AA	A	ɑ
AXR	@r	ɚ
IY	i	i
IH	I	I
EH	e	e
ER	3`	ɜr
AE	{	æ
AH	V	ʌ
OW	O	ɔ
UH	U	ʊ
UW and UX	u	u
AX	@	ə
	@`	ɘ
EY	eI	eI
AY	aI	aI
OY	OI	ɔI
AW	aU	aʊ
PCL P	p	p
BCL B	b	b
TCL T	t	t
DCL D	d	d
TCL CH	tS	tʃ
DCL JH	dZ	dʒ
KCL K	k	k
GCL G	g	ɡ
F	f	f
V	v	v
TH	T	θ
DH	D	ð
S	s	s
Z	z	z
SH	S	ʃ
ZH	Z	ʒ
HH	h	h
M	m	m
N	n	n
NG	N	ŋ
L	l	l
EL	tl	tl
R	r	r
W	w	w
Y	j	j
EM	Em	ɛm
EN	En	ɛn
W	W	ʍ
x	x	x
q	q	q
dx	dx	dx
jh	dZ	dʒ
ax-h	@	ə

BOSS Polish Corpus Transcription (blf – SAMPA modified for Polish)	IPA
i	i
y	ɨ
e	ɛ
a	a
o	ɔ
u	u
@	ə
p	p
b	b
t	t
d	d
k	k
g	ɡ
c	ts
J	gj
f	f
m	m
n	n
v	v
s	s
z	z
S	ʃ
Z	ʒ
s`	ɕ
z`	ʑ
x	x
t^s	ts
d^z	dz
t^S	tʃ
d^Z	dʒ
t^s`	ɕ
d^z`	dʑ
n`	ŋ
l	l
r	r
w	w
j	j
w~	ã
j~	ẽ

French SVOX Corpus Transcription (SAMPA)	IPA
i	i
e	e
E	ɛ
a	a
A	ɑ
O	ɔ
o	o
y	y
2	ø
9	œ
@	ə
e~	ẽ
a~	ã
o~	õ
9~	œ̃
p	p
b	b
t	t
d	d
k	k
g	ɡ
f	f
v	v
s	s
z	z
S	ʃ
Z	ʒ
j	j
m	m
n	n
J	ɲ
N	ŋ
l	l
R	R
w	w
H	ɥ
j	j

German MS Corpus Transcription (SAMPA)	IPA
i:	iː
I	ɪ
y:	yː
Y	ʏ
e:	eː
E:	ɛː
E	ɛ
2:	øː
9	œ
u:	uː
U	ʊ
o:	oː
O	ɔ
a:	aː
a	a
@	ə
6	ɐ
aI	aɪ
aU	aʊ
OY	ɔʏ
p	p
t	t
k	k
b	b
d	d
g	ɡ
?	ʔ
f	f
s	s
v	v
z	z
S	ʃ
Z	ʒ
C	ç
x	x
h	h
m	m
n	n
N	ŋ
l	l
r	r
j	j

2. Scripts:

a) Perl script(s) for generating .phones/.syllables/.word files necessary for the Festival tree-structure building

```
#!/usr/bin/env perl

        use strict;

my $blf_dir = "....";
my $phones_dir = "....";
        for my $blf_file (<$blf_dir/*.blf>) {
             (my $phones_file = $blf_file) =~ s/$blf_dir(.+)blf$/$phones(syllable or
             word)_dir$1phones(syllable or word)/;
        open(IN, "<$blf_file") or die "Cannot open $blf_file: $!\n";
             open(OUT, ">$phones(syllable or word)_file") or die "Cannot open $phones(syllable or word)_file:
             $!\n";
        print OUT "#\n";
        while (<IN>) {
        my @input = split(/ /);
        if ( ($input[0] == 0 ) & ($input[1] =~ /^\#$/)) {next;}
        my $time_stamp = $input[0]/16000;
        my $phone = $input[1];
        chomp $phone(syllable or word);

        $phone(syllable or word) =~ s/e~{1,2}/j~/g;
        $phone(syllable or word) =~ s/(-?2)|(5)|['<_:"%&]|((?<![wj])~)//g;
        if ($phone =~ /(^\s*$)|(\?)/) {
        $phone = "[nosegment]";
        }
        print OUT "$time_stamp 121 $phone(syllable or word)\n";
        $phone = $input[1];
        }
        close(IN);
        close(OUT);
      }
```

b) Praat script for obtaining voicing information from the articulatory recordings

```
        sound_directory$ = "/home/…./….wavs/"
        textgrid_directory$ = "/home/…./…../labels/"
        vuv_directory$ = "/home/…../…../vuv/"
        result_file$ = "/home/…./foo.txt"
        filedelete 'result_file$'
        Create Strings as file list... list 'textgrid_directory$'*.TextGrid
        numberOfFiles = Get number of strings
        for ifile to numberOfFiles
        filename$ = Get string... ifile
        Read from file... 'textgrid_directory$''filename$'
        textgrid_name$ = selected$ ("TextGrid", 1)

        soundname$ = replace$("'textgrid_name$'", "label", "POL", 1)
        Read from file... 'sound_directory$''soundname$'.wav

        To Pitch... 0 75 600
        To PointProcess
        To TextGrid (vuv)... 0.02 0.01
        vuv_name$ = "'textgrid_name$'_vuv"
        Rename... 'vuv_name$'
        Write to text file... 'vuv_directory$''vuv_name$'.TextGrid

        select Pitch 'soundname$'
        plus PointProcess 'soundname$'
        plus Sound 'soundname$'
        Remove

        select TextGrid 'textgrid_name$'
        trans_intervals = Get number of intervals... 1

        for interval to trans_intervals

        select TextGrid 'textgrid_name$'
        label$ = Get label of interval... 1 interval
```

```
        if label$ <> ""
        interval_start = Get starting point... 1 interval
        interval_end = Get end point... 1 interval
        interval_duration = interval_end - interval_start

        select TextGrid 'vuv_name$'
        Extract part... interval_start interval_end Yes

        vuv_intervals = Get number of intervals... 1

        if vuv_intervals == 2
        vuv_label_1$ = Get label of interval... 1 1
        vuv_label_2$ = Get label of interval... 1 2
        if vuv_label_1$ == "U"
        change_point = Get end point... 1 1
        fileappend "'result_file$'" 'soundname$''tab$''change_point:3''tab$''vuv_label_1$' ->
        'vuv_label_2$''newline$'
        endif
        endif
        Remove
        endif
        endfor
        select TextGrid 'vuv_name$'
        plus TextGrid 'textgrid_name$'
        Remove
        select Strings list
        endfor
        select Strings list
        Remove
```

c) Perl scripts for changing the svox format to the Festival-like-input-table:

```
        #!/usr/bin/perl -w
#definition of variables

my $i=0;
my $j=0;
#my $k=0;
my $number_letter;
my $index;
my @letter;               #each letter will be put one after the othe in this array
my @letter_index;         #the index of the line where the letter was
                          #found is put in this array -> $letter[$i] is at
                           #line $letter_index[$i] in the file
my $space= "     ";
my $number_steps = 9;
my $average_step = "low";

# build the tables:

my @VC_table = ("p","t","k","b","d","g","s","z","S","Z","f","v","m","n","N","l","j","J","R","w");
my @CTP_table = ("s","s","s","s","s","s","f","f","f","f","f","f","n","n","n","r","r","r","r","r");
my @CPL_table = ("l","a","v","l","a","v","a","a","p","p","ld","ld","l","a","v","a","p","p","a","v");
my @CVX_table = ("-","-","-","+","+","+","-","+","-","+","-","+","+","+","+","+","+","+","+","+");
my @Sonorants = ("R","m","n","j","w","l","J","N");
my @Vowels = ("e","E","i","a","A","2","9","u","O","o","e~","a~","o~","9~","@","E/","A/","&/","O/","U~/");

#open the input file
open (MYFILE_input, "nk0co0fr-FR22desc.dat") || die "ERROR - Could not open nk0co0fr-FR22desc.dat\n";

#open output files
open (MYFILE_output, ">>file_output.txt") || die "ERROR - Could not open file_output.txt\n";
open (MYFILE_error_report, ">>Report_Error.txt") || die "ERROR - Could not open Report_Error.txt\n";
#print the 1st line of the output:

# print pp_name  p_name   name     n_name   nn_name
print MYFILE_output "pp_name".$space."p_name".$space."name".$space."n_name".$space."nn_name".$space;

# print pp_cvox p_cvox  name_cvox n_cvox   nn_cvox
print MYFILE_output "pp_cvx".$space."p_cvx".$space."name_cvx".$space."n_cvx".$space."nn_cvx".$space;

# print pp_ctp p_ctp  name_ctp n_ctp       nn_ctp
print MYFILE_output "pp_ctp".$space."p_ctp".$space."name_ctp".$space."n_ctp".$space."nn_ctp".$space;

# print pp_ccpl p_ccpl  name_cpl n_ccpl    nn_ccpl
print MYFILE_output "pp_cpl".$space."p_cpl".$space."name_cpl".$space."n_cpl".$space."nn_cpl".$space;
```

```
#print the steps of name
for ($i=0;$i<$number_steps;$i++)
{
        print MYFILE_output "name_step_".$i.$space;
}

# print avg_pp_name       avg_p_name       avg_n_name       avg_nn_name
print MYFILE_output "avg_pp_name".$space."avg_p_name".$space."avg_n_name".$space."avg_nn_name".$space;

#print cluster_type word_structure
print MYFILE_output "cluster_type".$space."word_structure".$space."prev".$space."syllable";
print MYFILE_output "\r\n";

#read the input file, and put each line in the array @file
my @file= <MYFILE_input>;

#get the number of lines = size of @file
my $n_lines = @file;

#print the number of lines
print $n_lines;

########################################
#go through each line and catch the letter#
########################################

$index=0;
for($i=0;$i<$n_lines;$i++)
{
        #if $file[$i] starts with " then it's a letter -> put this letter in $letter[$index] +
        #$letter_index[$index] = $i + ($index++)

        my $start = substr($file[$i], 0, 1);
        if($start =~ /"/)
        {
                my @line_split = split(/"/,$file[$i]);
                $letter[$index]= $line_split[1];
                $letter_index[$index]= $i;
                $index++;
        }
}

#put the number of letters found in $number_letter
$number_letter=$index;

#prepare the variables for the next steps
my $first_index= 0;
my $is_begin=0;

#############################################################
#go through the array @letter to find out the wished clusters#
#############################################################

$i= $first_index;
while($i<$number_letter)
{
        # variable giving the case of cluster - initialised to "no_cluster"
        my $case= "no_cluster";

        #possible cases:
        # "1a"
        # "2a"
        # "3a"
        # "1b-CC"
        # "1b-CCC"
        # "2b-CC"
        # "2b-CCC"
        # "3b-CC"
        # "3b-CCC"
        # "1b-3b-CC" possible case?
        # "1b-3b-CCC" possible case?
        # "no_cluster"

#if a cluster is found get its informations
##1 - At the beginning of the word:
#a)single consonnant {R, m, n, N, j, w, l , J} + {e,E,A,a,O,o,u,y,2,9,@,e~,a~,o~,9~,E/,A/,&/,O/,U~/}
#b)consonants in clusters (consonant + consonant (CC) || consonant + consonant + consonant (CCC)) with
consonant in { TBD }
#2 - In the middle of the word:
```

```
#a)single consonnant (preceeded and followed by a vowel) {e,E,A,a,O,o,u,y,2,9,@,e~,a~,o~,9~,E/,A/,&/,O/,U~/} +
{R, m, n, N, j, w, l ,
#J} + {e,E,A,a,O,o,u,y,2,9,@,e~,a~,o~,9~,E/,A/,&/,O/,U~/}
#b)consonants in clusters (consonant + consonant (CC) || consonant + consonant + consonant (CCC)) with
consonant in { TBD }
##3 - At the end of the word:
#a)single consonnant {e,E,A,a,O,o,u,y,2,9,@,e~,a~,o~,9~,E/,A/,&/,O/,U~/} + {R, m, n, N, j, w, l , J}
#b)consonants in clusters ((consonant + consonant (CC)) || consonant + consonant + consonant (CCC)) with
consonant in { TBD }
##HINT : see how to handle CCC for example,that they are not counted 3 times or in CC too... (jump to next
consonnant outside the
#found cluster)

#save $i as first index of the eventual cluster
$first_index= $i;

# go throught the letter
# catch a consonnant

if(is_consonnant($letter[$i]))
{
        # check if the consonnant is at the beginning, middle or end of a word
        if((get_sentence_mark($i - 1) eq "/") || (get_sentence_mark($i - 1) eq "|") || ($letter[$i - 1] eq
"/"))
        {
                #is at beginning
                # if next letter is consonnant
                if(is_consonnant($letter[$i+1]))
                {
                        if((get_sentence_mark($i + 1) eq "/") || (get_sentence_mark($i +1) eq "|"))
                        {
                                # next letter is at end of word

                                # case (1b + 3b) (begin) CC(end) (possible case?)
                                $case = "1b-3b-CC";
                                # jump +2 letter
                                $i++;
                                $i++;
                        }
                        else
                        {
                                # if 2nd next letter is consonnant
                                if(is_consonnant($letter[$i+2]))
                                {
                                        if((get_sentence_mark($i + 2) eq "/") || (get_sentence_mark($i +2) eq
                                "|"))
                                        {
                                                # 2nd next letter is at end of word
                                                # case (1b + 3b) (beginning) CCC (end) (possible case?)
                                                $case = "1b-3b-CCC";
                                                #jump +3 letter
                                                $i++;
                                                $i++;
                                                $i++;
                                        }
                                        else
                                        {
                                                #else
                                                # case 1b CCC
                                                $case = "1b-CCC";
                                                # jump +3 letter
                                                $i++;
                                                $i++;
                                                $i++;
                                        }
                                }
                                else
                                {
                                        # else
                                        # case 1b CC
                                        $case = "1b-CC";
                                        #jump +2 letter
                                        $i++;
                                        $i++;
                                }
                        }
                }
                else
                {
```

```
                # else //next letter is vowel
                # case 1a
                $case = "1a";
                #jump +2 letter
                $i++;
                $i++;
        }
}
else
{
        if((get_sentence_mark($i) eq "/") || (get_sentence_mark($i) eq "|"))
        {
                #is at the end
                # case 3a // the preceeding letter must be a vowel cause if not it should be included
        in a CC or a CCC.
                $case = "3a";
                #jump +1 letter
                $i++;
        }
        else
        {
                #is in the middle
                # if next letter is consonnant
                if      (is_consonnant($letter[$i+1]))
                {
                        if((get_sentence_mark($i + 1) eq "/") || (get_sentence_mark($i +1) eq "|"))
                        {
                                #next letter is at end of word
                                # case 3b  CC(end)
                                $case = "3b-CC";
                                #jump +2 letter
                                $i++;
                                $i++;
                        }
                        else
                        {
                                #if 2nd next letter is consonnant
                                if      (is_consonnant($letter[$i+2]))
                                {
                                        # if 2nd next letter is at end of word
                                        if((get_sentence_mark($i + 2) eq "/") ||
                                (get_sentence_mark($i +2) eq "|"))
                                        {
                                                # case 3b CCC (end) (possible case?)
                                                $case = "3b-CCC";
                                                #jump +3 letter
                                                $i++;
                                                $i++;
                                                $i++;
                                        }
                                        else
                                        {
                                                #else
                                                # case 2b CCC (middle)
                                                $case = "2b-CCC";
                                                # jump +3 letter
                                                $i++;
                                                $i++;
                                                $i++;
                                        }
                                }
                                else
                                {
                                        # else
                                        # case 2b CC (middle)
                                        $case = "2b-CC";
                                        #jump +2 letter
                                        $i++;
                                        $i++;
                                }
                        }
                }
                else
                {
                        # else // next letter is vowel
                        # case 2a // the preceeding letter must be a vowel cause if not it should be
                included in a CC or a CCC.
                        $case = "2a";
                        #jump +2 letter
```

```
                                $i++;
                                $i++;
                        }
                }
        }
}

# if C at beginning
# if next letter is consonnant
# if next letter is at end of word
# case (1b + 3b) (begin) CC(end) (possible case?)
# jump +2 letter
# else if 2nd next letter is consonnant
# if 2nd next letter is at end of word
# case (1b + 3b) (beginning) CCC (end) (possible case?)
#jump +3 letter
#else
# case 1b CCC
# jump +3 letter
# else
# case 1b CC
#jump +2 letter
# else //next letter is vowel
# case 1a
# if C in the middle
# if next letter is consonnant
# if next letter is at end of word
# case 3b  CC(end)
# jump +2 letter
# else if 2nd next letter is consonnant
# if 2nd next letter is at end of word
# case 3b CCC (end) (possible case?)
#jump +3 letter
#else
# case 2b CCC (middle)
# jump +3 letter
# else
# case 2b CC (middle)
#jump +2 letter
# else // next letter is vowel
# case 2a // the preceeding letter must be a vowel cause if not it should be included in a CC or a CCC.
# if C at the end
# case 3a // the preceeding letter must be a vowel cause if not it should be included in a CC or a CCC.
# for a single consonnant cluster, the C must be in {R, m, n, N, j, w, l, J}
# for CC and CCC verify that at least 1 of the C are in {R, m, n, N, j, w, l , J}


# $case must be different from "no_cluster"
# &&
# for a single consonnant cluster, the C must be in {R, m, n, N, j, w, l , J}
# &&
# for CC and CCC verify that at least 1 of the C are in {R, m, n, N, j, w, l , J}
# checked by sub $



if(is_no_cluster($case,$first_index))
{
        #no cluster was found -> jump to next
        if($case eq "no_cluster")
        {
                $i++;
        }
}
else
{
        my @order = get_position_sonorant($case,$first_index);
}


## Beginnig of a word:
## the previous Letter is /
## OR
## the previous Sentence mark is / OR |
$is_begin=0;              #variable to check if the cluster is at the beginning
if ($case =~ /1/)
{
        $is_begin=1;
}
```

```
## End of a word:
# the Sentence_mark is / OR |
$is_end=0;                #variable to check if the cluster is at the end
if ($case =~ /3/)
{
$is_end=1;
}

## word_structure:
# beginning
# middle
# final

my $word_structure ="";
$word_structure = get_word_structure($is_begin, $is_end);

#cluster_type in:
# 1C - single consonant, ex. aRa, eRu ...
# 2C - consonat + consonant, ex. tr, pr, tl ...
# 3C - consonant + consonant + consonant, ex. str, spl ...

my $sep_syllable = "";
$sep_syllable = is_sep_syllable($case,$first_index);

#get the 9 steps
my @steps = get_steps($letter_index[$order[2]]);

 #pp_namep_name   name     n_name   nn_name  pp_cvox p_cvox   n_cvox    nn_cvox  step1    ...      step9    c_type
             c_plac   word_structure   cluster_type

# print pp_name  p_name   name     n_name   nn_name
print MYFILE_output
get_letter($letter[$order[0]]).$space.get_letter($letter[$order[1]]).$space.get_letter($letter[$order[2]]).$spa
ce.get_letter($letter[$order[3]]).$space.get_letter($letter[$order[4]]).$space;

# print pp_cvox p_cvox  name_cvx n_cvox    nn_cvox
print MYFILE_output
get_cvx($order[0]).$space.get_cvx($order[1]).$space.get_cvx($order[2]).$space.get_cvx($order[3]).$space.get_cvx
($order[4]).$space;

# print pp_ctp p_ctp  name_ctp n_ctp       nn_ctp
print MYFILE_output
get_ctp($order[0]).$space.get_ctp($order[1]).$space.get_ctp($order[2]).$space.get_ctp($order[3]).$space.get_ctp
($order[4]).$space;

# print pp_cpl p_cpl  name_cpl n_cpl       nn_cpl
print MYFILE_output
get_cpl($order[0]).$space.get_cpl($order[1]).$space.get_cpl($order[2]).$space.get_cpl($order[3]).$space.get_cpl
($order[4]).$space;

#print the steps of name
for ($j=0;$j<$number_steps;$j++)
{
        print MYFILE_output set_average_step($steps[$j]).$space;
}
#print the avg of the other letters
print MYFILE_output
set_average_step(get_average_steps($order[0])).$space.set_average_step(get_average_steps($order[1])).$space.set
_average_step(get_average_steps($order[3])).$space.set_average_step(get_average_steps($order[4])).$space;

# print cluster_type word_structure
print MYFILE_output get_cluster_type($case).$space.$word_structure.$space;

#print if cluster made with consonnant from previous letter
print MYFILE_output get_prev($case,$order[2]).$space;

# debug print if the cluster is in 1 or 2 syllables
print MYFILE_output $sep_syllable.$space;

# debug print line of the name letter found
#print MYFILE_output "line: ".($letter_index[$order[2]]+1);

print MYFILE_output "\r\n";


#pp.name p.name name n.name pp_cvox p_cvox  n_cvox* voice_10 voice_20 voice_30 (...)** c_type c_place
syll_structure cluster_type
#s       t      r    e       1        1        0        s      a       single        3C
```

```
#-       t       r     a       1       1       0       s       a       final       2C
#-       -       r     o       1       1       0       s       a       single       1C
#-       o       r     -       1       1       0       s       a       final        VC


close MYFILE_error_report;
close MYFILE_output;
close MYFILE_input;

sub get_pp_name
{
        my $index_l = $_[0];      #get the index of the letter
        if($index_l<2)
        {
                #there is no letter as pp_name
                return "-";
        }
        else
        {
                #get the letter twice before
                return $letter[$i-2];
        }
}

sub get_p_name
{
        my $index_l = $_[0];      #get the index of the letter
        if($index_l<1)
        {
                #there is no letter as p_name
                return "-";
        }
        else
        {
                #get the letter once before
                return $letter[$i-1];
        }
}

sub get_sentence_mark
{
        my $index_l = $_[0];      #get the index of the letter
        if ($index_l > 0)
        {
                my $line_b = $file[$letter_index[$index_l] + 2];
                return $line_b[1];
        }
        else
        {
                return "out_of_range";
        }
}



sub get_ctp
{
        my $index_l = $_[0];      #get the index of the letter
        my $letter_b = $letter[$index_l];
        my $l=0;
        my $position= -1;
        for($l=0;$l<@VC_table;$l++)
        {
                if($VC_table[$l] eq $letter_b)
                {
                        $position = $l;
                }
        }
        if($position == -1)
        {
                #letter not found
                return "-",
        }
        else
        {
                return $CTP_table[$position];
        }
}
```

```
sub get_cpl
{
	my $index_l = $_[0];	#get the index of the letter
	my $letter_b = $letter[$index_l]
	my $l=0;
	my $position= -1;
	for($l=0;$l<@VC_table;$l++)
	{
		if($VC_table[$l] eq $letter_b)
		{
			$position = $l;
		}
	}
	if($position == -1)
	{
		#letter not found
		return "-",
	}
	else
	{
		return $CPL_table[$position];
	}
}

sub get_cvx
{
	my $index_l = $_[0];	#get the index of the letter
	my $letter_b = $letter[$index_l];
	my $l=0;
	my $position= -1;
	for($l=0;$l<@VC_table;$l++)
	{
		if($VC_table[$l] eq $letter_b)
		{
			$position = $l;
		}
	}
	if($position == -1)
	{
		#letter not found
		return "-",
	}
	else
	{
		return $CVX_table[$position];
	}
}

sub get_cluster_type
{
	my $caseb = $_[0];	#get the case of the cluster
	if ($caseb =~ /a/)
	{
		return "1c";
	}
	else
	{
		if ($caseb =~ /CCC/)
		{
			return "3c";
		}
		else
		{
			if ($caseb =~ /CC/)
			{
				return "2c";
			}
			else
			{
				return "-";
				print MYFILE_error_report "Case not found: ".$caseb."\r\n";
			}
		}
	}
}

sub is_no_cluster
{
```

```
	my $caseb = $_[0];
	my $indexb = $_[1];
	if($caseb eq "no_cluster")
	{
		return 1;
	}
	else
	{
		if (($caseb =~ /a/)&&(!(is_sonorant($letter[$indexb])))) ###not in {R, m, n, N, j, w, l , J}))
		{
			return 1;
		}
		else
		{
			if ($caseb =~ /CCC/)
			{
				if((!(is_sonorant($letter[$indexb])))&&(!(is_sonorant($letter[$indexb+1])))&&
				(!(is_sonorant($letter[$indexb +2]))))## not in {R, m, n, N,  j, w, l , J}))
				{
					return 1;
				}
				else
				{
					return 0;
				}
			}
			else
			{
				if ($caseb =~ /CC/)
				{
					if((!(is_sonorant($letter[$indexb])))&&(!(is_sonorant($letter[$indexb
					+1])))) ### not in {R, m, n, N, j, w, l , J}))
					{
						return 1;
					}
					else
					{
						return 0;
					}
				}
				else
				{
					return 0;
				}
			}
		}
	}
}

sub get_position_sonorant
{
	my $caseb = $_[0];
	my $first_indexb = $_[1];
	my @orderb = (-1,-1,-1,-1,-1);
	if ($caseb =~ /a/)
	{
		#case 1 consonant
		if (is_sonorant($letter[$first_indexb])) ### in {R, m, n, N, j, w, l , J})
		{
			if ($caseb =~ /1/)
			{
				#case 1a
				@orderb = (-1,-1,$first_indexb,$first_indexb + 1,-1);
			}
			else
			{
				if ($caseb =~ /2/)
				{
					#case 2a
					@orderb = (-1,$first_indexb-1,$first_indexb,$first_indexb + 1,-1);
				}
				else
				{
					if ($caseb =~ /3/)
					{
						#case 3a
						@orderb = (-1,$first_indexb -1,$first_indexb,-1,-1);
					}
					else
```

```
                                        {
                                                #error, the case is not in 1a, b, 1c
                                                print MYFILE_error_report "Cluster error: ".$caseb." - at
                                        ligne ".$letter_index[$first_indexb]."\. Not in 1a, 2a, 3a."."\r\n";
                                        }
                                }
                        }
                }
                else
                {
                        #error, the cluster has no {R, m, n, N, j, w, l , J}
                        print MYFILE_error_report "Cluster error: ".$caseb." - at ligne
                ".$letter_index[$first_indexb]."\. It has no sonorant."."\r\n";
                }
        }
        else
        {
                if ($caseb =~ /CCC/)
                {
                        if (is_sonorant($letter[$first_indexb])) ###in {R, m, n, N,  j, w, l , J})
                        {
                                #sonorant at 1st position
                                @orderb = (-1,-1,$first_indexb,$first_indexb + 1,$first_indexb + 2);

                        }
                        else
                        {
                                if (is_sonorant($letter[$first_indexb + 1])) ###in {R, m, n, N, j, w, l , J})
                                {
                                        #sonorant at 2nd position
                                        @orderb = (-1,$first_indexb,$first_indexb+1,$first_indexb + 2,-1);
                                }
                                else
                                {
                                        if (is_sonorant($letter[$first_indexb + 2])) ###in {R, m, n, N, j, w,
                                        l , J})
                                        {
                                                #sonorant at 3rd position
                                                @orderb = ($first_indexb,$first_indexb+1,$first_indexb+2,-
                                        1,-1);
                                        }
                                        else
                                        {
                                                #error, the cluster has no {R, m, n, N, j, w, l , J}
                                                print MYFILE_error_report "Cluster error: ".$caseb." - at
                                        ligne ".$letter_index[$first_indexb]."\. It has no sonorant."."\r\n";
                                        }
                                }
                        }
                }
                else
                {
                        if ($caseb =~ /CC/)
                        {
                                if (is_sonorant($letter[$first_indexb])) ###in {R, m, n, N, j, w, l , J})
                                {
                                        #sonorant at 1st position
                                        @orderb = (-1,-1,$first_indexb,$first_indexb + 1,-1);

                                }
                                else
                                {
                                        if (is_sonorant($letter[$first_indexb + 1])) ### in {R, m, n, N, j,
                                        w, l , J})
                                        {
                                                #sonorant at 2nd position
                                                @orderb = (-1,$first_indexb,$first_indexb+1,-1,-1);

                                        }
                                        else
                                        {
                                                #error, the cluster has no {R, m, n, N, j, w, l , J}
                                                print MYFILE_error_report "Cluster error: ".$caseb." - at
                                        ligne ".$letter_index[$first_indexb]."\. It has no sonorant."."\r\n";
                                        }
                                }
                        }
                        else
                        {
```

```
                                #error, the cluster is not known
                                print MYFILE_error_report "Cluster error: ".$caseb." - at ligne
                        ".$letter_index[$first_indexb].". Cluster not known."."\r\n";
                        }
                }
        }
        return @orderb;
}

sub get_word_structure
{
        my $is_beginb= $_[0];
        my $is_endb= $_[1];
        #report an error if is_begin and is_end at the same time, and decide for is_begin
        my $word_structureb="";
        if( $is_beginb && $is_endb)
        {
                $word_structureb = "beginning";
                print MYFILE_error_report "Cluster is beginning and final at Line
        ".$letter_index[$first_index]."\r\n";
        }
        else
        {
                if($is_beginb)
                {
                        $word_structureb = "beginning";
                }
                else
                {
                        if($is_endb)
                        {
                                $word_structureb = "final";
                        }
                        else
                        {
                                $word_structureb = "middle";
                        }
                }
        }
        return $word_structureb;
}

sub is_consonnant
{
        my $letterb= $_[0];
        my $l=0;
        my $is_consonnant=0;
        for($l=0;$l<@VC_table;$l++)
        {
                if ($letterb eq $VC_table[$l])
                {
                        $is_consonnant = 1;
                }
        }
        return $is_consonnant;
}



sub is_sonorant
{
        my $letterb= $_[0];
        my $l=0;
        my $is_sonorant=0;
        for($l=0;$l<@Sonorants;$l++)
        {
                if ($letterb eq $Sonorants[$l])
                {
                        $is_sonorant = 1;
                }
        }
        return $is_sonorant;
}

sub is_vowel
{
        my $letterb= $_[0];
        my $l=0;
```

```
        my $is_vowel=0;
        for($l=0;$l<@Vowels;$l++)
        {
                if ($letterb eq $Vowels[$l])
                {
                        $is_vowel = 1;
                }
        }
        return $is_vowel;
}

sub is_sep_syllable
{
        my $caseb= $_[0];
        my $indexb= $_[1];
        my $sep = "one_Syllable";
        if ($caseb =~ /CCC/)
        {
                #1st letter end syllable or #2nd letter end syllable
                if((get_sentence_mark($indexb) eq "-") || (get_sentence_mark($indexb + 1) eq "-"))
                {
                $sep = "two_Syllable";
        }
        }
        else
        {
                if ($caseb =~ /CC/)
                {
                        #1st letter end syllable
                        if((get_sentence_mark($indexb) eq "-"))
                        {
                                $sep = "two_Syllable";
                        }
                }
        }
        return $sep;
}

sub get_steps
{
        my $lineb = $_[0];
        my $k=0;
        my $l=0;

        my @stepsb;

        #split the line of the letter by spaces and take the 2nd element which is the number of points
        my @line_splitb = split(/ /,$file[$lineb]);
        my $nbr_points = $line_splitb[1];

        #get the points:
        my @line_points = split(/ /,$file[$lineb+1]);
        #to change
        #my @pt($nbr_points);
        my @pt;


        if(($nbr_points==0)||($nbr_points eq ""))
        {
                for($k=0;$k<$number_steps;$k++)
                {
                        $stepsb[$k]= "-";
                }
                print MYFILE_error_report "No steps at line: ".$lineb.$space;
        }
        else
        {
                for($k=0;$k<$nbr_points;$k++)
                {
                        $pt[$k] = $line_points[2*$k+1];
                }

                #$pt[$nbr_points] = 0;
                #my $coeff = (100 / $nbr_points);
                #coeff = 100/n
                #$number_steps = $coeff * ($q + $r)
                #where r < 1

                #DIV ($number_steps/coeff) = q
```

```
		my $r = $nbr_points % $number_steps;
		#Rest (10/coeff) / coeff = r
		my $q = ($nbr_points - $r)/$number_steps;
		#q = nbr of enteger points entiers in nbr_steps
		#r = rest / last point non integer /

		if($q>=1)
		{
			my $last_point = 0;
			my $rest= 0;
			for($k=0;$k<$number_steps;$k++)
			{
				$stepsb[$k]= ($number_steps - $rest) * $pt[$last_point];
				if($q>1)
				{
					for($l=($last_point + 1);$l<$last_point + $q;$l++)
					{
						$stepsb[$k] += $number_steps * $pt[$l];
					}
				}
				if(($last_point + $q) < $nbr_points)
				{
					if(($r + $rest)<$number_steps)
					{
						$stepsb[$k] += ($r + $rest) * $pt[$last_point + $q];
						$last_point += $q;
						$rest += $rest;
					}
					else	# ($r + $rest >= $number_steps)
					{
						$stepsb[$k] += $number_steps * $pt[$last_point + $q];
						$rest += $r - $number_steps;
						$last_point += $q +1;
						$stepsb[$k] += $rest * $pt[$last_point];
						}
				}
				$stepsb[$k] = $stepsb[$k] / $nbr_points;
			}
		}
		else	#$q=0
		{
			#put in @weight the weight of each point
			my @weight;
			for($k=0;$k<$nbr_points;$k++)
			{
				$weight[$k]=$number_steps;
			}

			my $index=0;
			for($k=0;$k<$number_steps;$k++)
			{
				if($weight[$index]>=$nbr_points)
				{
					#only the index goes inside
					$stepsb[$k] = $pt[$index];
					$weight[$index] = $weight[$index] - $nbr_points;
				}
				else
				{
					#the index is not enough, need index+1
					$stepsb[$k] = ((($weight[$index]/$number_steps) * $pt[$index]) +
				((($nbr_points - $weight[$index])/$number_steps) * $pt[$index+1])) *
				($number_steps/$nbr_points);
					$weight[$index + 1] = $weight[$index +1] - ($nbr_points -
				$weight[$index]);
					$weight[$index] = 0;
					$index++;
				}
			}
		}
	}
		return @stepsb;
}

sub get_average_steps
{
	my $indexb= $_[0];
	my $lineb = $letter_index[$indexb];
	my $letterb = $letter[$indexb];
```

```
	my $k=0;

	if (($lineb == -1)||(get_letter($letterb) eq "-"))
	{
		return "-";
	}
	else
	{
		#split the line of the letter by spaces and take the 2nd element which is the number of points
		my @line_splitb = split(/ /,$file[$lineb]);
		my $nbr_points = $line_splitb[1];
		if(($nbr_points==0)||($nbr_points eq "")||($line_splitb[1] eq "/"))
		{
			return "-";
		}
		else
		{
			#get the points:
			my @line_points = split(/ /,$file[$lineb+1]);
			my $average=0;
			for($k=0;$k<$nbr_points;$k++)
			{
				$average += $line_points[2*$k+1];
			}
			$average = $average / $nbr_points;
			return $average;
		}
	}
}

sub set_average_step
{
	my $stepb = $_[0];
	if ($stepb eq "-")
	{
		return "-";
	}
	else
	{
		if (average_step eq "low")
		{
			if ($stepb <= 0.5)
			{
				return 0;
			}
			else
			{
				return 1;
			}
		}
		else	#average_step eq "high"
		{
			if ($stepb < 0.5)
			{
				return 0;
			}
			else
			{
				return 1;
			}
		}
	}
}

sub get_letter
{
	my $letterb = $_[0];
	if ($letterb eq "/")
	{
		return "-";
	}
	else
	{
	return $letterb;
	}
}

sub get_prev
{
```

```
        my $caseb = $_[0];
        my $indexb = $_[1];
        if((($caseb=~ /2a/)||($caseb=~ /3a/))&&(is_consonnant($letter[$indexb-1])))
        {
        return "prev_3c";
        }
        else
        {
        return "-";
        }
}
```

3. Perl script for converting Praat voicing EMA results into 'Festival-like' format with 9 equidistant voicing steps.

```
#!/usr/bin/perl -w

#########################################################
# script: ema_voicingPerl_v2.pl                         #
#                                                       #
# description: conversion of ema Polish voicing         #
#########################################################


#parameters
my $tab  = "          ";
my $i=0;
my $j=0;

#open the input file
open (MYFILE_input, "speaker1_non_emph.txt") || die "ERROR -  Could not open speaker1_non_emph.txt \n";
#open output file
open (MYFILE_output, ">>output_v2.txt") || die "ERROR -  Could not open output.txt \n";

#print the columns names
print MYFILE_output "name voice_1 voice_2 voice_3 voice_4 voice_5 voice_6 voice_7 voice_8 voice_9 position
sonorant\n";



#read the lines of the input file one after the other (read line by line)
my @file= <MYFILE_input>;
my $size_file = @file;

for ($i=0; $i < $size_file; $i++ )
{
           #substitute tab by space
           $file[$i]=~ s/        / /g;

           #remove the \r and \n
           $file[$i]=~ s/\r//g;
           $file[$i]=~ s/\n//g;

           #divide the line $i in 4 blocks:

           #1st split by spaces
           my @line_split = split(/ +/,$file[$i]);

           #then reduce into only 4 blocks
           #create a new array that will contain only 4 blocks
           my @line_split_corrected;

           #put the 1st block into the 1st corrected block
           $line_split_corrected[0]=$line_split[0];

           #check if more than 4 blocks were created
           if(@line_split>4)
           {
                      #concatenate the 2nd block with the next ones except the 2 last blocks
                      $line_split_corrected[1]="";

                      for($j=1;$j<(@line_split-2);$j++)
                      {
                                 if($j==1)
                                 {
                                            $line_split_corrected[1].=$line_split[$j];
```

```
                                        }
                                        else
                                        {
                                                    #add a space between 2 parts
                                                    $line_split_corrected[1].= " ".$line_split[$j];
                                        }
                                }
                                #the 2 last blocks are copied in the corrected 2 last blocks
                                $line_split_corrected[2]= $line_split[@line_split-2];
                                $line_split_corrected[3]= $line_split[@line_split-1];
                    }
                    else
                    {
                                #in case already 4 blocks were created, just copy the blocks as they are
                                $line_split_corrected[1]=$line_split[1];
                                $line_split_corrected[2]=$line_split[2];
                                $line_split_corrected[3]=$line_split[3];
                    }

                    #modify the output of block 1 with the sub function build_1
                    $line_split_corrected[1]= build_1($line_split_corrected[1]);

                    #print the corresponding output of the 4 blocks in the output file
                    for($j=0;$j<4;$j++)
                    {
                                print MYFILE_output $line_split_corrected[$j];

                                #add a tabulation after if it's not the last block of the line:
                                if($j<3)
                                {
                                            print MYFILE_output $tab;
                                }
                    }

                    #go to the next line if it was not the last line
                    if ($i < ($size_file-1))
                    {
                                print MYFILE_output "\n";
                    }


}

#close the files

close MYFILE_input;
close MYFILE_output;



################################################################################
# sub function: build_1                                                        #
# description: change the string from the input to the the string with the 9 steps      #
################################################################################

sub build_1
{
            #get the input parameter
            my $string_in = $_[0];

            my $number_change =0;
            my @change_times;
            my @letters;
            my @steps;
            my $k=0;


            #split first by the commas (each change is splited by ", " in the input file)
            my @blocks_comma = split(/, /,$string_in);


            for($k=0;$k<@blocks_comma;$k++)
            {
                        # for each splited block, split the spaces (to get the successive parts of information of
one change)
                        my @blocks_space = split(/ +/,$blocks_comma[$k]);

                        if(@blocks_space == 1)
                        {
```

```
                    #only 1 letter is found (no change)
                    $letters[0]=$blocks_comma[$k];
          }
          else
          {
                    #several changes found (increase the parameter giving the number of changes)
                    $number_change++;
                    #put the % in the array position 0:
                    $blocks_space[0]=~ s/%//g;
                    $change_times[$k]=$blocks_space[0];
                    #put the 1st letter in the array position 1:
                    $letters[$k]=$blocks_space[1];
          }
}

$change_times[$number_change]=-1;

#fill the @steps array
if ($number_change==0)
{
          #always the same value, so every steps have the same value
          for($k=0;$k<9;$k++)
          {
                    if($letters[0] eq "V")
                    {
                              $steps[$k]=1;
                    }
                    else
                    {
                              if($letters[0] eq "U")
                              {
                                        $steps[$k]=0;
                              }
                              else
                              {
                                        $steps[$k]="Error - unknown letter";
                              }
                    }
          }
}
else
{
          my $changes_done=0;
          my $letter_value=0;
          for($k=0;$k<9;$k++)
          {
                    my $stop = 1;

                    #get the letter

                    if ($changes_done<$number_change)
                    {
                              if($letters[$changes_done] eq "V")
                              {
                                        $letter_value=1;
                                        $steps[$k]=1;
                              }
                              else
                              {
                                        if($letters[$changes_done] eq "U")
                                        {
                                                  $letter_value=0;
                                                  $steps[$k]=0;
                                        }
                                        else
                                        {
                                                  $steps[$k]="Error - unknown letter";
                                                  $stop = 0;
                                        }
                              }
                    }
                    else
                    {
                              $steps[$k]=$letter_value;
                    }


                    # check if the change is done at this step
```

```
                                my $low_boundary_k = (10*($k+1)) - 5;
                                if($k==0)
                                {
                                        $low_boundary_k = 0;
                                }


                                my $high_boundary_k = (10*($k+1)) + 5;
                                if($k==8)
                                {
                                        $high_boundary_k = 100;
                                }


                                if(($low_boundary_k   <=   $change_times[$changes_done])&&($high_boundary_k   >
$change_times[$changes_done])&&($stop))
                                {
                                        my $low_boundary = $low_boundary_k;
                                        my $in_bounds = 1;
                                        $steps[$k]=0;

                                        while($in_bounds)
                                        {
                                                $steps[$k]+=        ($change_times[$changes_done]       -
$low_boundary) * $letter_value;

                                                $low_boundary = $change_times[$changes_done];

                                                $changes_done++;

                                                $letter_value = 1-$letter_value;

                                                if($changes_done<$number_change)
                                                {
                                                        if(($low_boundary                                <=
$change_times[$changes_done])&&($high_boundary_k > $change_times[$changes_done]))
                                                        {
                                                                $in_bounds = 1;
                                                        }
                                                        else
                                                        {
                                                                $in_bounds = 0;
                                                        }
                                                }
                                                else
                                                {
                                                        $in_bounds = 0;
                                                }
                                        }

                                        $steps[$k]+= ($high_boundary_k - $low_boundary) * $letter_value;
                                        $steps[$k] = $steps[$k]/($high_boundary_k - $low_boundary_k);

                                        $steps[$k] = sprintf("%.1f", $steps[$k]);
                                        $steps[$k]=~ s/.0//;

                                }
                        }
                }

                my $sub_string = "";

                for ($k=0;$k<9;$k++)
                {
                        $sub_string .= $steps[$k];
                        if($k<8)
                        {
                                $sub_string .= " ";
                        }
                }

                return $sub_string;
}
```

References:

Abramson, A. S. Lisker L. (1968): Voice Timing: Cross-Language Experiments in Identification and Discrimination. Journal of the Acoustical Society of America, Vol. 44, p. 377.

Alexiadou, A. (2006): Incremental Specification in Context. Grant Application for SFB 732, Universität Stuttgart and DFG (German Science Foundation).

Alexiadou, A. (2010): Incremental Specification in Context. Grant Application for SFB 732, Universität Stuttgart and DFG (German Science Foundation).

Bachan, J. (2006): Close Copy Speech Synthesis for Perception Testing and Annotation Validation. Adam Mickiewicz University, Poznań.

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

Boersma, P., Weenink, D. (2010): Praat: http://www.fon.hum.uva.nl/praat/.

Browman, C., Goldstein, L. (1988): Some Notes on Syllable Structure in Articulatory Phonology. Phonetica 45, p. 140-155.

Browman, C. P. Goldstein L. (1989): Articulatory gestures as phonological units. Phonology, Vol. 6, p. 201–251.

Browman, C., Goldstein, L. (2000): Competing constraints on intergestural coordination and self-organization of phonological structures. Bulletin de la Communication Parlée, Vol. 5, p. 25-34.

Bybee, J. (2002): Word frequency and context of use in the lexical diffusion of phonetically conditioned change. Language Variation and Change, Vol. 14, p. 261-290.

Bybee, J. (2006): From usage to grammar: the mind's response to repetition. Language Variation and Change, Vol. 82(4), p. 711-733.

Byrd, D. (1995): C-centers revisited. Phonetica, Vol. 52, p. 285-306.

Chafcouloff, M. (1980) : Les caractéristiqes acoustiques de [j, ɥ, w, l, r] en Français. Travaux de L'Institut de Phonétique d'Aix. Vol. 7, p.7-56.

Cenoz, J. (2001): The Effect of Linguistic Distance, L2 Status and Age on Cross-linguistic Influence in Third Language Acquisition. In Cenoz, J., Hufeisen, B., Jessner, U. (eds.): Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives. Cromwell Press Ltd. Great Britain.

Cho, Y.Y. (1990): Typology of voicing assimilation. The proceedings of the Ninth West Coast Conference in Formal Linguistics. p. 141-156.

Chomsky, N. Halle M. (1968): The sound pattern of English. New York: Harper and Row.

Clark, J. and Yallop, C. (2007): An Introduction to Phonetics and Phonology. Oxford: Blackwell Publishing.

Dell, F. (1995): Consonant clusters and phonological syllables in French. Lingua Vol. 95, p. 5 - 26.

Delvaux, V. Huet K. Piccaluga M. Harmegnies B. (2008): Perceptually driven VOT lenghtening in initial stops by French-L1 English L2-lerners. 8th International Seminar on Speech Production, p. 149-152

Delattre, P. (1971): Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English. Phonetica, Vol. 23, p. 129-155.

Demenko, G. Bachan J. Möbius B. Klessa K. Szymański M. Grocholewski S. (2008): Development and Evaluation of Polish Speech Corpus for Unit Selection Speech Synthesis Systems. Proceedings of Interspeech 2008, Brisbane, Australia, p. 1650-1653.

Demenko, G. Möbius B. Klessa K. (2008): The design of Polish Speech Corpus for Unit Selection Speech Synthesis. Language Technology Vol. 11, p. 85-101.

Dogil, G., Möbius, B. (2001): Toward a perception based model of the production of prosody. Journal of the Acoustical Society of America 110, p. 2737.

Dogil, G. (2010). Hard-wired phonology: limits and latitude of phonological variation in pathological speech. In: C. Fougeron, B. Kühnert, M. D'Imperio, N. Valleè (eds.): Laboratory Phonology, Vol. 10, p. 343-380. Mouton de Gruyer. Berlin, NY.

Duez, D. Legou T. Viallet F. (2008): Final Lenghtening in Parkinsonian French speech. Clinical Linguistics and Phonetics, Vol. 23/11, p. 781-793.

Dukiewicz, L. Sawicka I. (1995): Fonetyka i fonologia. Wydawnictwo Instytutu Jezyka Polskiego PAN. Kraków.

Eckmann, F.R. (1977): Markedness and the contrastive analysis hypothesis. Language Learning, Vol. 27, p. 315-30.

Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Narayanan, S., Alwan, A. (2000): Acoustic modeling of American English /r/. Journal of Acoustical Society of America, Vol. 108, p. 343-356.

Fant, G. (1970): Acoustic Theory of Speech Production. Mouton. The Hague.

Festival (2009): http://www.cstr.ed.ac.uk/projects/festival/.

Fisher, B. (1997): ftp://jaguar.ncsl.nist.gov/pub/

Goldinger, S. (1997): Words and Voices: Episodic Theory of Lexical Access. Psychological Review, Vol. 105(2), p. 251-279.

Goldstein, L., Browman C. P. (1986): Representation of vocicing contrasts using articultory. Journal of Phonetics, Vol. 14, p. 339–342.

Guenther, F. H. (2006): Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production. In: Brain and Language, Vol. 96, p. 280–301.

Guenther, F. H, Espy-Wilson, C. Y. ,Boyce, S. E., Matthies, M. L., Zandipour, M., Perkell, J. S. (1999): Articulatory tradeoffs reduce acoustic variability during American English /r/ production. In: Journal of Acoustical Society of America, Vol. 105/5, p. 2854–2865.

Gussmann, E. (1992): Resyllabification and Delinking: The Case of Polish Voicing. Linguistic Inquiry, Vol. 23, p. 29–56.

Gussmann, E. (2007): The Phonology of Polish: Oxford University Press.

Hall, T. A. (1993): The phonology of German /R/. In: Phonology, Vol. 10, p. 83–105.

Halle, M., Stevens, K. (1971): Feature Geometry and Feature. Spreading. Lingustic Inquiry, Vol. 26, p. 1 - 46.

Hammarberg, B. (2001). Roles of L1 and L2 in L3 Production and Acquisition. In Cenoz, J., Hufeisen, B., Jessner, U. (eds.): Cross-linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives. Cromwell Press Ltd. Great Britain.

Harris, J. Gussman E. (2002): Word-final onsets. UCL Working Papers in Linguistics, Vol. 14, p. 1–42.

Hawkins, S. (2010): Phonological features, auditory objects and illusions. Journal of Phonetics. Vol. 38, Issue 1, p. 60-89.

Hermes, A., Grice, M., D. Mücke & H. Niemann (2008): Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. Proceedings of the 8th International Seminar on Speech Production, Strasbourgh, France, p. 433-436.

Honorof, D.N., Browman, C.P. (1995): The Center or Edge: How are consonant clusters organized with respect to the vowel? In K. Elenius & P. Branderup (Eds.). Proceedings of the 13th ICPhS, Stockholm, Sweden, p. 552-555.

Hoonhorst, I., Colin, C., Markessis, E., Radeau, M., Deltenre, P., Serniclaes, W. (2009): Some Aspects of Speech and the Brain. The N100 component: An electrophysiological cue of voicing perception. Fuchs, S. Lœvenbruck H. Pape D. Perrier P. (Hg.). Frankfurt am Main: Peter Lang.

Iverson, G., Sang-Cheol Ahn (2007): English voicing in dimensional theory. Language Sciences, Vol. 29, p. 247-269.

Iverson, K., Salmons J.C. (1995): Aspiration and Larnygeal Representation in Germanic. Phonology, Vol.12, No.3, p. 369-396.

Jakobson, R., Halle M. (1956): *Fundamentals of Language*. The Hague: Mouton

Jessen, M. (1998): Phonetics and Phonology of Tense and Lax Obstruents in German. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Jessen, M. (2000): Phonetic implementation of the distinctive auditory features [voice] and [tense]. AIMS, Vol. 6, No.4, p. 11–64.

Jilka, M. (2009): Talent and proficiency in language. In Dogil, G., Susanne M. Reiterer (eds.): Language Talent and Brain Activity, p. 17-66. Mouton de Gruyter.

Johnson, K. (1997): Acoustic & Auditory Phonetics. Blackwell Publishing.

Johnson, K. (1997): Speech perception without speaker normalization: An exemplar model. In: Johnson, K., Mullenix, *J.* (eds.), Talker Variability in Speech Processing. Academic Press, p. 145-165.

Kahn, D. (1968): Syllable-based generalizations in English Phonology. Doctoral disseratation. City University of New York.

Kallestinova, E. (2004): Voice and Aspiration of Stops in Turkish. In: Voicing. Special Issue, Folia Linguistica XXXVIII/1-2, Dogil. G. (eds.).

Keating, P. (1980): A phonetic study of a voicing contrast in Polish. PhD Thesis. Brown University.

Keating, P. (1984): Phonetic and Phonological Representation of Stop Consonant Voicing. In: Language, Vol. 60/2, p. 286–319.

Keating, P. (2003): Phonetic and other Influences on Voicing Contrasts. Proceedings of the 6th International Seminar on Speech Production, Macquarie University, S. Palethorpe and M. Tabain (eds), p. 119-124.

Keating, P.; Mikoś, M. J. Ganong W. F., III (1981): A cross-linguistic study of range of voice onset time in the perception of initial stop voicing. In: Research Laboratory of Electronics. MIT, p. 36–521.

Keating, P. MacEachern P. Shryock A. Dominguez S. (1994): A Manual for Phonetic ´Transcription: Segmentation and Labeling Words in Spontaneous Speech. In: UCL Working Papers in Linguistics No. 88, p. 91-120.

Kingston, J. Diehl R. L. (1994): Phonetic Knowledge. Language, Vol. 70, p. 419-454.

Kingston, J. Lahiri, A., Diehl, R. (2010): Voice. University of Massachusetts, Amherst, Oxford and University of Texas at Austin.

Kingston, J., Diehl, R. L., Kirk, C. J., Castleman, W. A. (2008): On the internal perceptual structure of distinctive features: The [voice] contrast. In: Journal of Phonetics, Vol. 36, p. 28–54.

Klabbers, E., Stöber, K., Veldhuis, R., Wagner, P., Breuer, S. (2001): Speech Synthesis Development Made Easy: The Bonn Open Synthesis System. Proceedings of Eurospeech, Aalborg.

Kreidler, Ch W. (2004): The prononciation of English. A course book. USA: Blackwell Publishing.

Lacerda, F. (1995): The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. Proceeding of the 13th International Congress of Phonetic Sciences (Stockholm). Vol. 2, p. 140-174.

Ladefoged, P., Maddieson, I. (1996): The sounds of the world's languages. Blackwell Publishing, Oxford.

Lamel, L.F., Kassel, R.H., Seneff, S. (1986): Speech Database Development: Design and Analysis of the Acoustic-Phonetic Corpus. Proceedings DARPA Speech Recognition Workshop, Report No. SAIC-86/1546, p. 100-109.

Léon, P. (2009): Phonétisme et prononciations du français. Paris: Armand Colin.

Liberman, A. M., Mattingly, I. G. (1985). The motory theory of speech revised. Cognition, Vol. 21, p. 1-36.

Lipski, S. (2006): Neural correlates of fricative contrasts across language boundaries. PhD Dissertation. IMS, Universität Stuttgart.

Lisker, L., Abramson, A.S. (1964): A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. Word, Vol. 20, No. 3, p. 384-422.

Lisker L., Abramson, A. S. (1967): Discriminability along the Voicing Continuum: Cross-Language Tests. Prague: Proceedings of the 6th International Congress of Phonetic Sciences, p. 569–573

Lombardi, L. (1991): Laryngeal Features and Laryngeal Neutralization. Ph.D. dissertation. Amherst. University of Massachusetts.

Lombardi, L. (1995): Restrictions on direction of voicing assimilation: and OT account. University of Maryland Working Papers in Linguistics 3, Vol. 3, p. 89–115.

Lombardi, L. (1995): Laryngeal Neutralization and Syllable Wellformedness. Natural Language and Linguistic Theory, Vol. 13, No. 1, p. 39-74.

Maekawa, K. Kikuchi H. (2004): Corpus-based analysis of vowel devoicing in spontaneous Japanese: and interim report. In: van de Weijer, J. Nanjo K. Nishihara T. (eds.) Voicing in Japanese. Mouton de Gruyter, Berlin.

Moosmüller, S., Ringen, C. (2004): Voice and Aspiration in Austrian German Plosives. Folia Linguistica, Vol. 38 (1-2), p. 43-61.

Meunier, C. (1989): Une approche acoustique des groups consonantique: étude des phases de transision. Travaux de l'Institut de Phonétique d'Aix, Vol. 13, p.11-29.

Miller, G.A., Nicely, P.E. (1955): An analysis of perceptual confiusions among some English consonants. Journal of the Acoustical Society of America, Vol. 27, p. 338-52.

Möbius, B., Dogil, G. (2002). Phonemic and postural effects on the production of prosody. In Bernard Bel and Isabelle Marlien (eds.), Proceedings of the Speech Prosody 2002 Conference (Aix-en-Provence, Laboratoire Parole et Langage), p. 523-526.

Möbius, B. (2004): Corpus-based investigations on the phonetics of consonant voicing. In: Folia Linguistica, Vol. 38 (1-2), p. 5–26.

Mortreux, S. (2008): English Coronal Consonants Produced by L2 French Lerners - An articulatory and Acoustic Study. 8th International Seminar on Speech Production, p. 145-148.

Mücke, D., Grice, M., Becker, J., Hermes, A. (2009): Sources of variation in tonal alignment: evidence from acoustic and kinematic data. Journal of Phonetics 37 (3), p. 321–338.

Nam, H. (2007): Syllable-level intergestural timing model: Split-gesture dynamics focusing on positional asymmetry and moraic structure. In J. Cole & J. I. Hualde (eds.), Laboratory Phonology 9 (Phonology and Phonetics), p.483-506. Berlin, New York: Walter de Gruyter.

Nam, H., Goldstein, L., & Saltzman, E. (2009): Self-organization of syllable structure: a coupled oscillator model. In F. Pellegrino, E. Marisco, & I. Chitoran (eds.). Approaches to phonological complexity, p. 299-328.

Nielsen, K. Y. (2007): Implicit Phonemic Imitiations are constrained by phonemic contrasts. Proceedings of Interspeech 2008 (Brisbane), p. 1961-1964.

Nosofsky, R.M. (1988): Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality. Journal of Experimental Psychology: Learning, Memory and Cognition, Vol. 14, No. 4, p. 700-708.

Ostendorf, M., Price, P.J., Shattuck-Hufnagel, S. (1995): The Boston University Radio News Corpus. Linguistic Data Consortium.

Pallier, C., Ventureyera, V., Yoo, H.-Y. (2004). The loss of first language phonetic perception in adopted Koreans. Speech Communication, Vol. 17, p.78-91.

Perl, programming language (2009): http://www.perl.org/

Petrova, O., Szentgyörgyi, S. (2004). /v/ and Voice Assimilation in Hungarian and Russian. In: Voicing. Special Issue, Folia Linguistica XXXVIII/1-2, Dogil. G. (eds.).

Poon, P. G. & Mateer, C.A. (1985): A study of VOT in Nepali Consonants. Phonetica, Vol. 42, p. 39-47.

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

Pierrehumbert, J. (2006): The next toolkit. Journal of Phonetics, Vol. 34, p. 516-530.

Pierrhumner, J. (2003). Probabilistic phonology: Discrimination and robustness. In: Bod, R., Hay, J., Jannedy, S. (eds.). Probability Theory in Lingusitics. The MIT Press, p. 177-228.

Pinker, S. (1995): The language instinct. Penguin Press.

Piroth. H.G., Janker P. M. (2004): Speaker-dependent differences in voicing and devoicing of German obstruents. In: Journal of Phonetics, Vol. 32, p. 81–109.

Prince, A. and Smolensky, P. (1993): Optimality Theory: Constraint Interaction in Generative Grammar. Rutgers University Center for Cognitive Science and Computer Science Department, University of Colorado at Boulder.

Python (2009): http://www.python.org/

R project. (2001): The R Project for Statistical Computing. Online verfügbar unter http://www.R-project.org/.

Rapp, S. (1995): Automatic Phonemic Transcription and Linguistic Annotation from Known Text with Hidden Markov Models: An Aligner for German. Proceedings of ELSNET Goes East and IMACS Workshop 'Integration of Language and Speech in Academia and Industry' (Moscow, Russia).

Recasens, D. (1989): Long range coarticulation effects for tongue dorsum contact in VCVCV sequences, Speech Communication, Vol. 8, p. 293-307.

Recasens, D., Fontdevila, J., Pallarés, M.D., (1996): Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized /l/. Journal of Phonetics, Vol. 24, p. 165-185.

Rice, K. (2005): Sequential voicing, postnasal voicing, and Lyman's Law revisited. In: van de Weijer, J. Nanjo K. Nishihara T. (eds.) Voicing in Japanese. Mouton de Gruyter, Berlin.

Rubach, J. (1996): Nonsyllabic Analysis of Voice Assimilation in Polish. In: Linguistic Inquiry, Vol. 27, p. 69–100.

Rubach, J. (2008): Prevocalic Faithfulness. Phonology, Vol. 25/3., p. 433-468.

Saltzmann, E., Byrd, D. (2003): Speech Production. In: The Handbook of Brain Theory and Neural Networks. Arbib, M.A. (eds). The MIT Press. Cambridge, Massachuesetts; London, England.

Sampa for Polish: http://www.phon.ucl.ac.uk/home/sampa/polish.html

Saville-Troike, M. (2006): Introducing Second Language Acquisition. Cambridge University Press.

Schneider, K., Lintfert, B., Dogil, G., Möbius, B. (2006): "Phonetic grounding of prosodic categories". In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter, and Johannes Schließer (eds.), Methods in Empirical Prosody Research (De Gruyter, Berlin), p. 335-361

Schweitzer, A., Braunschweiler, N., Klankert T., Möbius, B., Säuberlich B. (2003): Restricted Unlimited Domain Synthesis. Proceedings of the European Conference on Speech Comunication and Technology (Geneva), p. 1321-1324.

Schweitzer, A., Möbius, B. (2004): Exemplar-based production of prosody: Evidence from segment and syllable durations. Proceedings of the speech prosody 2004 conference (Nara), p. 459-462.

Schweitzer, K., Walsh, M., Möbius, B., Riester, A., Schweitzer, A., Schütze, H. (2009): Frequency matters: Pitch accents and information status. Proceedings of EACL (Athens), p. 728-736.

Shepard, R.N. (1972): Psychological representation of speech sounds. In David E. E. and Denes, B. (eds.). Human Communication: A unified view. New York: McGraw-Hill, p. 67-113.

Shih, C., Möbius, B., Narasimhan, B. (Hg.) (1999): Contextual effects on consonant voicing profiles: a cross-linguistic study. San Francisco: Proceedings of the 14th ICPhS Conference.

Shimizu, K. (1989): A Cross-Language Study of Voicing Contrasts of Stops. Studia Phonologica XXIII, p. 1-13.

Sieczkowska, J., Möbius, B., Dogil, G. (2010): Specification in Context – Devoicing Processes in Polish, French, American English and German Sonorants. Proceedings of Interspeech, Japan, p. 1549-1552.

SmartKom (2003): Das Leitprojekt SmartKom: Dialogische Mensch-Technik-Interaktion durch koordinierte Analyse und Generierung multipler Modalitäten. [http://smartkom.dfki.de/start.html].

Sproat, R., Fujimura, O. (1993): Allophone variation in English /l/ and its implications for phonetic implementation. Journal of Phonetics, Vol. 21, p. 291-312.

Steriade, D. (1997): Phonetics in Phonology: The Case of Laryngeal Neutralization. PhD Thesis. UCLA.

Stevens, K. N. (1989): On the quantal nature of speech. In: Journal of Phonetics, Vol. 17, p. 3–45.

Stevens, K. N. (2002): Towards a model of lexical access based on acoustic landmarks and distinctive features. Journal of the Acoustical Society of America, Vol. 111, p. 1872-1891.

Stevens, K. N. (2005): The Handbook of Speech Perception. Features in Speech Perception and Lexical Access. Pisoni, D. B. Remez R. E. Blackwell Publishing.

Stevens, K.N. (2010): Quantal theory, enhancement and overlap. Journal of Phonetics, Vol. 38, p. 10-19.

Tranel, B. (1987): The sounds of French. An Introducion. Cambridge: Cambridge University Press.

The International Phonetic Alphabet (IPA) (2009): http://www.arts.gla.ac.uk/ipa/ipa.html

Tsuchida, A., Cohn A.C., Kumada M. (2000): Sonorant devoicing and the phonetic realization of [spread glottis] in English. Working Papers of the Cornell Phonetics Laboratory, Vol 13, p.167-181.

van den Berg, J. (1958): Myoelastic-aerodynamic theory of voice production. Journal of Speech and Hearing Research, Vol.1, p. 227-44.

Wade, T. (2008): Detailed phonetic memory for multi-level and part-word sequences. Laboratory Phonology, Vol. 11, p.151-152.

Wade, T., Möbius, B. (2010): Detailed phonetic memory for multi-word and part-word sequences. Mouton de Gruyter, Laboratory Phonology, Vol.1, Nr.2, p. 283-294.

Wade, T., Dogil, G., Schütze, H., Walsh, M., Möbius, B. (2010): Syllable frequency effects in a context-sensitive segment production model. Journal of Phonetics, Vol. 38, p. 227-239.

Walsh, M., Schütze, H., Möbius, B., Schweitzer, A. (2007): An Exemplar-Theoretic account of syllable frequency effects. Proceedings of the XVI ICPhS, Saarbrücken, p. 481-484.

Walsh, M., Wade, T., Möbius, B., Schütze, H. (2010): Multilevel exemplar theory. Cognitive Science , p. 537-582.

Westbury, J. R. Keating P. (1986): On the naturalness of stop consonant voicing. In: Journal of Linguistics, Vol. 22, p. 146–166.

Wetzels, Leo W., Mascaro, J. (2001): The Typology of Voicing and Devoicing. Language, Vol.77, No.2, p.207-244.

Wiese, R. (1996): The Phonology of German. Oxford: Clarendon Press.

Wilson, C. Y. Boyce S. E. Jackson M. Narayanan S. Alwan A. (2000): Acoustic modeling of American English /r/. Journal of Acoustical Society of America, Vol. 108/1, p. 343–356.