Relation between acoustic and articulatory dimensions of speech sounds

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig-Maximilians-Universität

München

vorgelegt von

Eugen Klein

aus

Russkaja-Poljana

2020

Referent: Prof. Dr. Phil Hoole

Korreferentin: Prof. Dr. Marianne Pouplier

Tag der mündlichen Prüfung: 10.07.2020

To my wife Elena and my children Constantin and Anthea

Contents

Acknowledgments

First and foremost, I would like to thank Jana Brunner for making this dissertation possible by providing me with a fruitful and focused research environment at the Department of German Studies and Linguistics of the Humboldt-University of Berlin. Jana supported me relentlessly, especially in the beginning of my endeavor, helping me to flesh out a concrete research agenda by regular chats, extensive discussions, and words of encouragements. She also endorsed me to attend meetings and conferences where I was able to present and discuss my work with other researchers. As a principal investigator, Jana was always supportive of my ideas and let me develop and pursue my own intuitions. On the other hand, she provided me with her extensive experience and guidance during the unnerving and lengthy publication process. As a parent, I very much appreciated Jana's understanding nature which made it possible for me to flexibly adjust my time schedule and to work from home. This became more and more necessary throughout the project, especially as my second child was born.

I am very grateful to my supervisor Phil Hoole at the Institute of Phonetics and Speech Processing of the Ludwig-Maximilians-University Munich. Phil always supported me along the way and responded instantly when I got stuck with any practical, methodological, or technical questions. Phil's broad and profound expertise in speech production and experimental research inspired me to strive for concise investigations and thorough analyses. Along with Jana, Phil was a key figure in making this dissertation possible. I also thank Phil for agreeing to supervise my dissertation and inviting me several times to Munich to present my work to fellow phoneticians.

I also would like to thank Tine Mooshammer, who welcomed me as a new member of the phonetics group at the Humboldt-University of Berlin and let me set up and conduct my experiments at her laboratory. I am thankful to Marianne Pouplier who instantly agreed to review my dissertation and was available for any additional questions.

I thank Miriam Oschkinat, Yulia Guseva, and Megumi Terada who supported my work

throughout different stages of the project by helping me to conduct experiments and annotate data. Special thanks goes to Megumi who saved a lot of my time and spared me additional stress by preparing and wrapping up of EMA recording sessions. As a result, we were able to complete the recordings for the last experiment within five weeks which we both were very happy about. Thanks for being a firm and competent assistant!

Overall, I spent about four wonderful years as a researcher and a PhD student. During this time, I was able to learn a lot of different things about speech, research, teaching, and how to approach challenging technical and organizational issues. Beside my actual investigation, I had the luxury to stray into adjacent topics and explore my interests in programming and machine learning. Being payed to visit some of the most popular venues in speech science and to meet some of the world's top researchers was an unforgettable experience and helped me to settle (at least for now) on the topic of speech and language technology for the next chapter of my life. As much as I enjoyed this time, I feel that at this point am ready to move on.

When something that played a special role in one's life comes to an end, there is often a feeling of sadness and a sense of void. However, this time, I realized that what prevailed was a feeling of joy. For that, I would like to thank my loving wife Elena and my two wonderful children who always managed to distract me from any work-related stress with their foolishness and their laughter. Constantin was six months old when I started my position as a research associate. Anthea was born about one year before I finished the work on the project. Although this caused some really exhausting weeks (if not months) in my life, it was also the best thing that could happen to me. My family made the last four years appear like a brief moment, and every time I was about to become fed up with the experiments, analyses and reviewers, there was always something more important that I could get upset about. Although getting research accepted for publication might cause boundless excitement, it doesn't compare to witnessing the first steps of your children.

I was lucky to finish my research and submit my dissertation shortly before the beginning of the lockdown due to COVID-19 pandemic. One of my personal silver linings in this terrible situation was the fact that I could defend my thesis online. That made it possible for many of my family members to join my defense. Although it might be sometimes difficult to convey to your family what exactly you are working on as a PhD researcher, I definitely felt their support and feelings of pride during and after my defense. For their kind words, I would like to thank every one of them.

Zusammenfassung

Sprecher[1] produzieren Sprachlaute, indem sie einen kontrollierten Luftstrom vorbei an ihren Stimmlippen und durch eine artikulatorische Konfiguration führen, was letztendlich in einem bestimmten akustischen Ergebnis mündet. In diesem Sinne können Sprachlaute als Relationen zwischen der artikulatorischen und der akustischen Dimension verstanden werden. Diese allgemeine Vorstellung wird durch die Ergebnisse der Neuroforschung gestützt, die darauf hindeuten, dass sensorische Repräsentationen von Sprachlauten sowohl im auditiven als auch somatosensorischen Cortex gespeichert werden und sich durch neuronale auditiv-somatosensorische Zuordnungen auszeichnen (Hickok, Houde, & Rong, 2011; Tourville & Guenther, 2011). Das übergeordnete Ziel der vorliegenden Dissertation ist es, unser Verständnis von der Funktionsweise dieser Relationen zu verbessern.

Neben der Neuroforschung stützen sich moderne Sprachproduktionstheorien auf Verhaltensexperimente, die zeigen, dass Sprecher somatosensorische und auditive Feedbacksignale nutzen, um Fehler in ihrer eigenen Sprachproduktion auszugleichen. Dies wurde mithilfe von oral-artikulatorischen und auditiven Perturbationsstudien gezeigt. In solchen Experimenten müssen Probanden unter erschwerten Bedingungen kurze Wörter oder einzelne Silben vorsprechen, z.B. während ihre Artikulationsbewegungen blockiert werden (z.B. Hamlet & Stone, 1976; Fowler & Turvey, 1980; Abbs & Gracco, 1984; Savariaux, Perrier, & Orliaguet, 1995; Tremblay, Shiller, & Ostry, 2003) oder ihr auditives Feedback in Echtzeit manipuliert wird (z.B. Jones & Munhall, 2000; Houde & Jordan, 1998; Shiller, Sato, Gracco, & Baum, 2009). Um die eigene Verständlichkeit aufrecht zu erhalten, müssen Sprecher in solchen Situationen die von ihren sensorischen Feedbackkanälen übertragenen Fehler mit passenden artikulatorischen Korrekturbewegungen koordinieren. Das heißt, sie müssen ihre Sprachpro-

[1] Aus Gründen der Lesbarkeit wurde im Text die männliche Form gewählt, nichtsdestoweniger beziehen sich die Angaben auf Angehörige beider Geschlechter.

duktion an die erschwerten Bedingungen anpassen. Diese Fähigkeit wird allgemein unter dem Begriff motorische Äquivalenz gefasst (eine Übersicht dazu findet sich in Perrier & Fuchs, 2015).

Da die Rolle des akustischen und somatosensorischen Feedbacks in den meisten Studien separat untersucht wurde, ist es nicht vollständig klar, in wieweit beide Feedbacksignale bei der Sprachproduktion vom Sprecher berücksichtigt werden. In jüngerer Zeit stellten Lametti, Nasir, and Ostry (2012) die Hypothese auf, dass Sprecher individuelle Präferenzen in Bezug auf den sensorischen Feedbackkanal aufweisen, den sie vorwiegend zur Überwachung ihrer eigenen Sprachproduktion nutzen. In ihrer Studie untersuchten die Autoren die Reaktionen der Probanden auf eine gleichzeitige somatosensorische Kiefer- sowie eine auditive F1-Perturbation. Lametti et al. (2012) stellten fest, dass Sprecher, die ihre Kieferposition verändert hatten, um für die somatosensorische Perturbation zu kompensieren, die F1-Frequenz während der auditiven Perturbation unverändert ließen und umgekehrt.

Während F1 in der Studie von Lametti et al. (2012) nach oben perturbiert wurde, und so einen mit der zugleich nach unten wirkenden Kieferperturbation kompatiblen Fehler verursachte, können Perturbationen zu inkompatiblen Informationen im Sprachproduktionssystem führen. Das kann z.B. auftreten, wenn das akustische Feedback einen Fehler signalisiert, das somatosensorische Feedback aber erfolgreiches Erreichen des artikulatorischen Ziels rückmeldet. Einige Autoren schlugen die Hypothese vor, dass eine solche Inkongruenz zwischen Feedbacksignalen ein möglicher Grund für unvollständige Kompensation sein könnte (z.B. Katseff, Houde, & Johnson, 2012). Diese Hypothese wurde jedoch durch einige empirische Befunde in Frage gestellt. Beispielsweise zeigte Feng, Gracco, and Max (2011), dass Sprecher nur dann für schließende Perturbation des Kiefers kompensierten, wenn dies zu einer messbaren F1-Absenkung führt. Wenn aber zugleich die F1-Frequenz im auditiven Feedback der Sprecher erhöht wird, um dem intendierten akustischen Output zu entsprechen, kompensieren sie nicht länger für die Kieferperturbation.

Die obigen Befunde beschränken sich hauptsächlich auf F1-Perturbationen in halb-offenen und offenen Vokalen. Verallgemeinerungen, die auf diesen Ergebnissen basieren, können daher schwierig sein, da neuere Untersuchungen darauf hindeuten, dass der Einfluss des somatosensorischen Feedbacks bei der Artikulation von verschiedenen Vokalphonemen unterschiedlich stark sein kann. Insbesondere wird angenommen, dass die Kompensation im Falle geschlossener Vokale wie /i/ schwächer ist im Vergleich zu offenen Vokalen, da die ersteren sich durch einen stärkeren physischen Kontakt zwischen aktiven und passiven Artikulatoren auszeichnen (siehe Mitsuya, MacDonald, Munhall, & Purcell, 2015).

Die Hypothese, dass in einigen Fällen die somatosensorische Dimension bei der Produktion eines Sprachlauts die dominantere Rolle spielt, lässt sich auf die allgemeine Unterscheidung zwischen Vokalen und Konsonanten zurückführen, die in der Sprachproduktionsforschung häufig gemacht wird (z.B. Guenther, Hampson, & Johnson, 1998). Dieser Unterscheidung liegt die Idee zu Grunde, dass Konsonanten eher in der artikulatorischen Dimension definiert sind, Vokale aber in der auditiven. Die Rolle des auditiven Feedbacks für die Konsonantenproduktion ist daher umstritten. Die Beantwortung dieser Frage wird dadurch erschwert, dass auditive Perturbation von Konsonanten aufgrund technischer Schwierigkeiten eingeschränkt ist. Uns sind nur zwei Untersuchungen von Shiller et al. (2009) und von Casserly (2011) bekannt, die auditive Perturbation von Frikativen untersuchten.

In beiden Studien mussten die Probanden Sibilanten produzieren, während ihr akustisches Spektrum so in Echtzeit perturbiert wurde, dass sich der spektrale Schwerpunkt (COG) im auditiven Feedback der Sprecher verringerte oder erhöhte. Die Autoren berichteten von drei unterschiedlichen Verhaltensmustern ihrer Probanden als Reaktion auf die angewandte Perturbation: keine Veränderung, Erhöhung des COG oder Absenkung des COG. Diese Befunde stehen im starken Kontrast zu den zumeist konsistenten kompensatorischen Anpassungen, die während der auditiven Perturbation von Vokalen beobachtet werden. Wenn wir uns jedoch oral-artikulatorischen Perturbationsstudien zuwenden, beobachten wir analoge Unterschiede bezüglich der kompensatorischen Variabilität von Frikativen und Vokalen.

Die Ergebnisse dieser Studien legen nahe, dass die erfolgreiche Kompensation von Frikativen durch hohe artikulatorische Anforderungen gekennzeichnet ist, da um den entsprechenden Sprachlaut genau zu produzieren eine Reihe von artikulatorischen Parametern (Verschlussstelle, Zungenrille sowie Kieferhöhe) kontrolliert werden müssen (Hamlet & Stone, 1978; Honda, Fujino, & Kaburagi, 2002; Brunner, Hoole, & Perrier, 2011). In dieser Hinsicht scheinen sich die Frikative erheblich von den Vokalen zu unterscheiden. Beispielsweise stellte McFarland and Baum (1995) fest, dass die Sprecher nach 15-minütigem Sprechen mit einem Beißblock in der Lage waren, die spektralen Eigenschaften von Vokalen, jedoch nicht von Frikativen, fast vollständig wiederherzustellen. Diese Beobachtungen können die kompensatorische Variabilität erklären, die in auditiven Perturbationsstudien von Frikativen beobachtet wurde und implizieren zugleich, dass es bei der Untersuchung von Kompensation bei Frikativen hilfreich sein könnte, zusätzliche Parameter neben dem COG zu betrachten.

Eines der Ziele der vorliegenden Untersuchung ist es, eine formale Analyse zu entwickeln, anhand derer möglich wird, das Ausmaß der akustischen Anpassungen bei Frikativen unabhängig von bestimmten Maßen (z.B. COG) zu beurteilen. Mit einem solchen analytischen Instrument wird es uns darüber hinaus möglich, kompensatorische Effekte in der akustischen und artikulatorischen Dimension zu untersuchen, indem wir das Ausmaß der Anpassungen anhand von spektralen sowie räumlichen Signalen berechnen können, die zeitgleich während auditiver Perturbation aufgezeichnet werden. Dadurch soll unser Verständnis der Wechselwirkung zwischen artikulatorischen und akustischen Dimensionen bei der Produktion von Sprachlauten verbessert werden. Eine Antwort auf die Frage, wie Sprecher kompensatorische Anpassungen als Reaktion auf auditive Perturbationen vornehmen, ist nämlich alles andere als trivial, da der Grad der Anpassungseffekte aufgrund der motorischen Äquivalenz in beiden Dimensionen unterschiedlich groß sein kann.

Im ersten Kapitel untersuchten wir den Einfluss eines stärkeren linguo-palatalen Kontakts auf die Fähigkeit der Sprecher, mehrere Kompensationsstrategien gleichzeitig anzuwenden. Während des Experiments produzierten die Probanden den zentralen, geschlossenen, ungerundeten Vokal /ɨ/, während seine F2-Frequenz in Abhängigkeit vom vorhergehenden Konsonanten (/d/ oder /g/) in entgegengesetzte Richtungen perturbiert wurde. Die bidirektionale Perturbation sollte die Probanden dazu ermutigen, zwei unterschiedliche Kompensationsstrategien anzuwenden, um den Vokal in /dɨ/ und /gɨ/ zu produzieren. Die Sprecher mussten für die Perturbation unter stark somatosensorisch eingeschränkten Bedingungen kompensieren, da sie den linguo-palatalen Kontakt im Zielvokal stets beibehalten mussten, während sie die Verschlussstelle entlang der anterior-posterioren Achse verschoben. Die beiden unterschiedlichen Konsonantenkontexte wurden so gewählt, dass die erforderlichen kompensatorischen Anpassungen entweder mit der üblichen koartikulatorischen Relation zwischen /dɨ/ und /gɨ/ vereinbar oder nicht vereinbar waren. Zweiunddreißig russische MuttersprachlerInnen (25 Frauen, 7 Männer) nahmen an der Studie teil.

Bei der Untersuchung des durchschnittlichen Kompensationsverhaltens stellten wir fest, dass die Sprecher zwei Anpassungsstrategien anwendeten, auch wenn diese von den koartikulatorischen Relationen ihrer unperturbierten Sprachproduktion abwichen. Eine detailliertere Analyse der individuellen Anpassungsstrategien ergab, dass 72 Prozent der Sprecher in der Lage waren, zwei unterschiedliche Produktionsstrategien für den Ziellaut zu entwickeln. Etwa die Hälfte dieser Sprecher, entwickelten ein symmetrisches Kompensationsmuster, bei dem für beide Perturbationsrichtungen im gleichen Maße kompensiert wurde, während die übrigen Sprecher ein asymmetrisches Kompensationsmuster aufwiesen, beim dem für die

Aufwärtsperturbation stärker kompensiert und die Abwärtsperturbation ignoriert wurde.

Insgesamt kompensierten 90 Prozent aller Probanden für die Aufwärtsperturbation, während nur etwa 31 Prozent aller Probanden für die Abwärtsperturbation kompensierten. Diese kompensatorische Asymmetrie scheint mit der phonemischen Asymmetrie der russischen geschlossenen Vokale übereinzustimmen. Es ist nämlich so, dass /i/ im Russischen nur palatalisierten Konsonanten folgen kann, während sowohl /ɨ/ als auch /u/ ausschließlich nach nicht-palatalisierten Vokalen erscheinen (siehe Bolla, 1981, S. 108-110). Da das dominante Perzeptionsmerkmal der Palatalisierung die Höhe der F2-Frequenz zu Beginn eines Vokals ist, erscheint es sinnvoll anzunehmen, dass die meisten Probanden auf die Aufwärtsperturbation reagierten, da sie das perturbierte /ɨ/ mit erhöhter F2-Frequenz als phonemischen Fehler der Palatalisierung klassifizierten. Auf der anderen Seite, reagierten weniger Sprecher auf die Abwärtsperturbation, da dies zur keiner Veränderung des phonemischen Status des wahrgenommenen Vokals führte.

Eine alternative Hypothese, die die kompensatorische Asymmetrie erklären könnte, besteht in der Idee, dass während es möglich war, für die Aufwärtsperturbation ausschließlich durch das Absenken der F2-Frequenz zu kompensieren, eine Kompensation der Abwärtsperturbation von den Sprechern erforderte, dass sie neben der F2-Frequenz auch die F3-Frequenz anheben, da beide Frequenzen sich relativ nahe zueinander im Ziellaut /ɨ/ befinden. Obwohl es für die Sprecher möglich sein sollte, F2 und F3 gleichzeitig durch das Verändern eines einzigen Artikulationsparameters wie der horizontalen Zungenposition anzuheben, ist die Anpassung der F3-Frequenz möglicherweise einfacher mit zusätzlichen artikulatorischen Veränderungen wie dem Grad der Lippenspreizung zu bewerkstelligen. Aus der Literatur zu russischen Vokalen ist es nämlich bekannt, dass /ɨ/ normalerweise mit einer etwas breiteren Lippenspreizung als /i/ produziert wird, das viel höhere F3-Werte aufweist (siehe Bolla, 1981, S. 109-110). Wenn in diesem Szenario die Sprecher zusätzlich zur Vorwärtsbewegung der Zunge ihre Lippen verengten, könnte dies zu einer Erhöhung der F3-Werte und zugleich zu einer stärkeren F2-Kompensation führen. Gewisse Evidenz für diese Hypothese lieferten die Ergebnisse einer Korrelationsanalyse der F2- und F3-Veränderungen für die beiden entgegengesetzten Perturbationsrichtungen, die zeigten, dass diese Korrelation auf der Gruppenebene signifikant und während der experimentellen Trials mit abwärts gerichteter Perturbation hoch positiv war, jedoch nicht während der Trials mit aufwärts gerichteter Perturbation.

Das Ziel des zweiten Expiriments war es, die Relevanz vom auditiven Feedback während der Produktion von Frikativen zu untersuchen und dabei methodologische Mängel von

früheren Studien zu berücksichtigen. Insbesondere haben wir die Perturbation nur auf den finalen Frikativ von CVC-Wörtern beschränkt und die Zielsegmente in Echtzeit perturbiert. Wir führten eine bidirektionale Perturbationsstudie des russischen Frikativs /sʲ/ durch, bei der das Spektrum des untersuchten Lauts in Abhängigkeit von dem experimentellen Stimulus ([lesʲ] oder [vesʲ]) in entgegengesetzte Richtungen perturbiert wurde (was zu einem höheren bzw. niedrigeren COG führte). Wir entschieden uns für das Russische, da das russische Konsonanteninventar die Reihe stimmloser Frikative /s/, /sʲ/ und /ʃʲ/ enthält, die durch qualitativ ähnliche Frequenzspektren charakterisiert sind. Diese akustische Nähe zwischen den drei Lauten ermöglichte es uns, auditiven Perturbationen des Ziellauts /sʲ/ durchzuführen, die es akustische entweder dem /s/ oder dem /ʃʲ/ ähnlicher machten. Dreiundzwanzig russische MuttersprachlerInnen (16 Frauen, 7 Männer) nahmen an der Studie teil.

Um das kompensatorischen Verhalten der Probanden umfassend zu analysieren, untersuchten wir verschiedene akustische Maße ihrer Sprachproduktion einschließlich der ersten drei spektralen Momente sowie zusätzlicher Parameter, die aus verschiedenen Frequenzbändern des Frikativspektrums extrahiert wurden (siehe Koenig, Shadle, Preston, & Mooshammer, 2013). Während der Datenanalyse untersuchten wir mithilfe eines überwachten Klassifizierungsalgorithmus (Random Forest (RF); Breiman, 2001) die Frage, ob bestimmte akustische Parameter identifiziert werden können, die sich während des Anpassungsprozesses systematisch verändern. Außerdem analysierten wir die zeitliche Dimension des Anpassungsprozesses, indem wir die Vorhersagen des Algorithmus über mehrere Zeitintervalle des Experiments anhand der akustischen Parameter modellierten, die für die Klassifizierung des durchschnittlichen Anpassungsverhaltens der Sprecher als relevant erachtet wurden. Um die individuellen Kompensationsstrategien nachzuvollziehen, modellierten wir die Veränderungen einzelner Parameter mittels verallgemeinerter additiver gemischter Modelle (GAMM).

Im Gegensatz zu Pertubationsstudien von Vokalen, bei denen sich normalerweise nur Formanten während des Anpassungsprozesses verändern, beobachteten wir ein kompensatorisches Verhalten, das durch umfangreichere spektrale Anpassungen bezüglich einer Reihe akustischer Parameter gekennzeichnet war. Mithilfe von Berechnungen der Wichtigkeit einzelner akustischen Parameter und anschließender RF-Modellierung konnten wir zeigen, dass COG zwar hilfreich für die Unterscheidung zwischen den unperturbierten Frikativen /s/, /sʲ/ und /ʃʲ/ war, aber keine Aussagekraft bei der Beschreibung des Anpassungsverhaltens der Probanden als Reaktion auf die Perturbation von /sʲ/ hatte. So zeigte ein entsprechendes GAMM-Modell, dass COG-Veränderungen für viele Sprecher unabhängig von der Perturba-

tionsrichtung auftraten. Mit anderen Worten, obwohl sich COG im Verlauf des Experiments im Durchschnitt signifikant veränderte, war es kein geeigneter Marker für das kompensatorische Verhalten und schien eher infolge von anderen Anpassungen aufzutreten.

Anschließend haben wir für alle untersuchten Parameter der /s^j/-Produktionen Wichtigkeitswerte separat für einzelne experimentelle Phasen berechnet. Diese Analyse ergab, dass zusätzlich definierte spektrale Maße wichtig waren, um vorherzusagen, unter welcher Perturbationsrichtung ein /s^j/-Token produziert wurde. Die durchschnittlichen Werte der Vorhersagegenauigkeit von RF-Modelle, die für einzelne Phasen berechnet wurden, stiegen während aufeinanderfolgenden Perturbationsphasen signifikant an, was darauf hindeutet, dass die Probanden ihre Kompensationsstrategien im Verlauf der experimentellen Sitzung verbesserten.

Um konkrete Kompensationsstrategien nachzuvollziehen, die von jedem Sprecher im Verlauf des Experiments angewendet wurden, berechneten wir zunächst individuelle Wichtigkeitswerte für alle Parameter basierend auf den /s^j/-Produktionen der Sprecher gegen Ende des Experiments. Anhand dieser Werte extrahierten wir die Funktionskurven einzelner Probanden aus den zuvor berechneten GAMM-Modellen. Dieses Verfahren ergab, dass etwa 42 Prozent der Sprecher die Amplitude des niedrigen Frequenzbands (600-5500 Hz) als Reaktion auf die Perturbation anpassten. Im Einzelnen hieß das, dass diese Sprecher die Amplitude des niedrigen Frequenzbands verringerten, wenn das gesamte Spektrum bei Abwärtsperturbation in Richtung niedrigerer Frequenzen verschoben wurde.

Unsere Beobachtung, dass einzelne Sprecher unterschiedliche akustische Parameter des Frikativspektrums als Reaktion auf die angewendeten auditiven Perturbationen modifizierten, steht im Einklang mit Ergebnissen früherer oral-artikulatorischen Perturbationsstudien (Hamlet & Stone, 1978; Flege, Fletcher, & Homiedan, 1988; Honda et al., 2002; Brunner et al., 2011). Wir glauben, dass die beobachtete Anpassungsvariabilität, verglichen zum Beispiel mit der Adaption bei Vokalen, auf den höheren Grad an akustisch-artikulatorischer Komplexität der Adaption bei Frikativen zurückzuführen ist.

Im vierten Kapitel hatten wir das Ziel, die Diskrepanz zwischen dem Grad der Anpassung zwischen Frikativen und Vokalen formal zu untersuchen. Zu diesem Zweck formulierten wir die Hypothese von Artikulationskomplexität einer Anpassungsaufgabe. Was mit diesem Konzept gemeint ist, wollen wir im Folgenden anhand von heterogenen Ergebnissen früherer Perturbationsstudien verdeutlichen. Als beispielsweise Fowler and Turvey (1980) die Produktion von Vokalen untersuchten, die unter Einsatz eines Beißblocks gesprochen wurden, stellten die Autoren fest, dass die Sprecher sich innerhalb von wenigen experimentellen Trials an diese Perturbationen anpassen konnten. Andererseits konnte in der Studie von Savariaux et al. (1995), bei der die Lippen der Sprecher während der Produktion von /u/ mit einem Plastikrohr blockiert wurden, nur die Hälfte der Sprecher für die labiale Perturbation teilweise kompensieren und nur ein einzelner Sprecher kompensierte vollständig durch Verändern der Verschlussstelle von der velo-palatalen zur velo-pharyngealen Region. In einer Folgestudie von Savariaux, Boë, and Perrier (1997) konnten zwei Sprecher nach einer artikulatorischen Trainingseinheit, bei der sie den Ziellaut /u/ nach /o/ produzierten, was eine stärkere Retraktion der Zunge bewirkte, vollständige Kompensation erzielen.

Für die beiden beschriebenen Perturbationsszenarien ist es plausibel anzunehmen, dass die Anpassung an die Beißblock-Perturbation eine artikulatorische Veränderung erfordert, die der unperturbierten Sprachproduktion ähnlicher ist als die, die bei der Plastikrohr-Perturbation erforderlich ist. Während der ersten Aufgabe müssen die Probanden lediglich ihre Zunge stärker als gewöhnlich anheben, da ihre öffnende Kieferbewegung blockiert ist. Während der Perturbation mit dem Plastikrohr müssen die Probanden die blockierte Lippenrundung durch das Zurückziehen der Zunge ausgleichen. Diese artikulatorische Anpassung ist weniger offensichtlich, da der Artikulator, der für die Kompensation eingesetzt wird, sowie seine Bewegungsrichtung weniger mit der unperturbierten Artikulationskonfiguration assoziiert werden. Infolgedessen sind weniger Sprecher in der Lage, die geeigneten artikulatorischen Anpassungen zu identifizieren, um für die Perturbation zu kompensieren. Diese Hypothese scheint geeignet zu sein, um den Unterschied zwischen den unterschiedlich starken Anpassungsgraden während der Perturbation von Frikativen und Vokalen zu erklären. Folglich haben wir den Anpassungsgrad der Sprecher an F2- und Frikativ-Perturbation mittels RF-Modellierung überprüft. Dies ermöglichte uns, Vorhersagegenauigkeiten für die Klassifikation von Abwärts- und Aufwärtsperturbation für Vokale zu berechnen und diese mit Vorhersagegenauigkeiten für Frikative zu vergleichen. Achtzehn Sprecher (14 Frauen, 4 Männer), die am ersten Experiment teilnahmen, schlossen auch das zweite Experiment ab. Dementsprechend konnten wir für die beiden Experimente berechneten Vorhersagegenauigkeiten für jeden Sprecher miteinander korrelieren.

Die Ergebnisse des Vokalexperiments stimmen mit unseren Ergebnissen aus dem ersten Kapitel überein. Die Werte der Vorhersagegenauigkeiten der berechneten RF-Modelle lassen darauf schließen, dass die Sprecher sich an die F2-Perturbation anpassen konnten, wobei alle Probanden einen Genauigkeitswert von mindestens 80 Prozent erreichten. Die für die untersuchten akustischen Parameter berechneten Wichtigkeitswerte legen nahe, dass die Sprecher im Durchschnitt eine Kompensationsstrategie entwickelten, die sich von Beginn der ersten Perturbationsphase an und bis zum Ende des Experiments auf die F2-Frequenz fokussierte. Für das Frikativexperiment beobachteten wir insgesamt geringere Genauigkeitswerte im Vergleich zum Vokalexperiment. Dies bedeutet, dass der Anpassungsgrad der Sprecher während der F2-Perturbation ihr Kompensationsverhalten bei Frikativperturbation nicht vorhersagen konnte. Dies wurde durch eine Korrelationsanalyse der für beide Experimente berechneten individuellen Genauigkeitswerte bestätigt. Im Gegensatz zum Vokalexperiment verbesserte sich die Vorhersagegenauigkeit nur langsam im Verlauf des Frikativexperiments.

Da wir den Einfluss von sprecherspezifischen Merkmalen auf den beobachteten Kompensationsgrad ausschließen können, sind wir der Ansicht, dass diese Diskrepanz auf die unterschiedliche Anzahl und Transparenz der artikulatorischen Parameter zurückzuführen ist, die die Sprecher anpassen mussten, um für die Perturbationen zu kompensieren, die beim ersten und zweiten Experiment angewendet wurden. Mit anderen Worten, die höhere Artikulationskomplexität der Anpassungsaufgabe im zweiten Experiment führte zu weniger erfolgreichen Anpassungsergebnissen. Zusammenfassend lässt sich sagen, dass die akustisch-artikulatorische Relation in Frikativen im Vergleich zu Vokalen weniger transparent ist und somit die Hypothese unterstützt, dass das Erreichen von Produktionszielen sich durch lautspezifische akustisch-artikulatorische Relationen auszeichnet (Perkell, 2012).

Im fünften Kapitel verfolgten wir das Ziel, den Grad der motorischen Äquivalenz zu untersuchen, der bei kompensatorischen Anpassungen als Reaktion auf auditive Perturbationen vorhanden ist. Zu diesem Zweck führten wir eine auditive Perturbationsstudie des Frikativs /s/ durch, in der wir zusätzlich Artikulationsbewegungen der Probanden mithilfe von EMA aufnahmen. Das Spektrum des untersuchten Lauts wurde in einem Wort abwärts perturbiert und in einem Kontrollwort unverändert belassen, wobei beide Wörter in einen einzelnen Stimulussatz eingebettet waren ([lasə (ʔ)ɛɐ̯hiːlt aɪ̯nɪ̯ tasə]; *Lasse erhielt eine Tasse*). Anhand der akustischen und artikulatorischen Daten der Probanden konnten wir untersuchen, wie sie ihre akustischen Kompensationsstrategien artikulatorisch umsetzen. An der Studie nahmen 19 weibliche, deutsche Muttersprachlerinnen teil.

Während der Datenanalyse untersuchten wir mithilfe von RF-Modellen die zeitliche Dimension des Anpassungsprozesses in der akustischen und artikulatorischen Dimension. Dies umfasste eine genauere Untersuchung von akustischen und artikulatorischen Parametern, die die Vorhersagegenauigkeiten der berechneten RF-Modelle beeinflussten. Anschließend identifizierten wir für einzelne Sprecher diejenigen akustischen Parameter, die für die Vorhersage des Perturbationsbedingung als am wichtigsten erachtet wurden. Anschließend konnten wir entsprechende Artikulationsparameter identifizieren, die für die Implementierung der akustischen Anpassungen als wichtig erachtet wurden.

Insgesamt stimmten die Ergebnisse in der akustischen Dimension mit den Beobachtungen aus dem dritten Kapitel in Bezug auf die spektrale Perturbation von /s^j/ überein. Die beobachteten Veränderungen waren durch relativ hohe Variabilität bezüglich der als am wichtigsten für die Kompensation erachteten akustischen Parameter gekennzeichnet. Für ungefähr 50 Prozent der Sprecher führten die akustischen Anpassungen zu Verschiebungen von spektraler Energie zwischen dem niedrigen (600-5500 Hz) und dem mittleren (5500-11000 Hz) Frequenzband. Für weitere 20 Prozent der Sprecher verschob sich die höchste Frequenz des mittleren Frequenzbandes in ihrer Produktion. Die Anzahl der modifizierten akustischen Parameter sowie die Vorhersagegenauigkeit akustischer RF-Modelle nahmen im Verlauf des Experiments stetig zu. Dies lässt darauf schließen, dass die Probanden unterschiedliche Produktionsstrategien für perturbierte und unperturbierte /s/-Tokens entwickelten.

Als wir die Artikulationsbewegungen der Sprecher untersuchten, die während der Perturbation auftraten, stellten wir fest, dass die Sprecher sofort auf die auditive Perturbation reagierten. Bei 40 Prozent der Sprecher beobachteten wir artikulatorische Anpassungen, die am relevantesten für die Unterscheidung zwischen perturbierten und unperturbierten /s/-Token waren, entweder als anterior-posteriore Bewegungen der Zungenspitze oder vertikale Verschiebungen der Unterlippe. Bei weiteren 20 Prozent der Sprecher waren vertikale Kieferbewegungen für die Kompensation am relevantesten. Insgesamt erhöhte sich die Vorhersagegenauigkeit von artikulatorischen RF-Modellen im Verlauf des Experiments, obwohl das Klassifizierungsmuster darauf hindeutet, dass Veränderungen in der artikulatorischen Dimension im Gegensatz zu der akustischen unregelmäßiger auftraten.

Schließlich zeigte ein Vergleich der über die akustische und artikulatorische Dimension hinweg wichtigen Parameter, dass eine Vielzahl von artikulatorischen Anpassungen zu vergleichbaren Veränderungen im akustischen Output führte. Zum Beispiel konnten Sprecher, die die spektrale Amplitude im niedrigen und mittleren Frequenzband balancierten, dies erreichen, indem sie entweder ihre Zungenspitze, die Unterlippe oder die Kieferposition veränderten. Zusammengenommen legen die Ergebnisse nahe, dass die Sprecher als Reaktion auf die Perturbation ihren Artikulationsraum erforschten, um das akustische Resultat ihrer Sprachproduktion auf ein bestimmtes Ziel hin anzupassen. Insgesamt stimmen diese Ergebnisse mit der Vorstellung überein, dass Sprachlaute perzeptuell-motorische Einheiten sind,

denen Artikulationsbewegungen zu Grunde liegen, die durch perzeptuelle Eigenschaften beeinflusst und geformt werden (Schwartz, Basirat, Ménard, & Sato, 2012).

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Goals of speech production

In their daily communication, speakers produce speech by pushing a controlled air stream past their vocal folds and through a vocal tract configuration formed by a set of articulators (tongue, lips, jaw, upper incisors etc.) which ultimately results in a certain acoustic output. In this sense, speech and, specifically, speech sounds can be understood as a relation between articulatory and acoustic dimensions. This idea is supported by more recent neuroimaging results which suggest that sensory representations of speech sounds are stored across auditory and somatosensory cortices and are characterized by neural auditory-somatosensory mappings (Hickok et al., 2011; Tourville & Guenther, 2011). The overall aim of the current dissertation is to improve our understanding of the functional nature of this relation.

While earlier theorizing was for some time dominated by the view that goals of speech production are articulatory by their nature (e.g., Browman & Goldstein, 1989; Saltzman & Munhall, 1989), most recent theories all in all agree that speech sound representations are defined in a multidimensional auditory-somatosensory space. However, despite their similarities many theories still differ in how they portrait the relation between the articulatory and acoustic dimensions. Subtle controversy arises around the issues whether there is a hierarchical relation between auditory and somatosensory signals and how they are integrated during speech production. For instance, several authors assume that targets of speech are rather auditory than articulatory while an auditory-motor-somatosensory mapping is learned by a speaker during speech acquisition (e.g., Guenther & Hickok, 2015). This position is similar to the idea that speech sounds are perceptuo-motor units comprising of articulatory movements which are shaped by perceptual properties and selected for their functional value for communication (Schwartz et al., 2012). Other authors more strongly emphasize the role of the somatosensory signal which is assumed to be employed by speakers to fine-tune their articulatory movements (Hickok, 2012) or to be in a constant trade-off relation with the auditory signal (Houde & Nagarajan, 2011).

Apart from recent neuroimaging research, modern theories of speech production rely on a bulk of behavioral evidence which shows that speakers employ somatosensory and auditory feedback signals to adjust for errors in their own speech production. To investigate these questions, it proved empirically fruitful to conduct oral-articulatory and auditory perturbation studies. During such experiments, speakers have to produce speech under aggravated conditions, e.g., under blockage of their articulatory movements or under altered auditory feedback. In order to retain intelligibility of their speech, speakers need to coordinate errors transmitted by their sensory feedback channels with appropriate corrective articulatory movements. In other words, speakers have to adapt their speech production to the aggravated conditions.

The earliest perturbation studies examined speakers' adaptive behavior to static mechanical perturbations of the articulatory apparatus. A number of studies, for instance, examined speakers' production of vowels when a bite-block was inserted between their teeth (e.g., Fowler & Turvey, 1980; Gay, Lindblom, & Lubker, 1981). These authors have demonstrated that speakers are able to adapt to this kind of static perturbations with very little practice and produce acoustic output equivalent to their unperturbed speech by reorganizing their articulatory strategies. On the one hand, these results suggest that speakers rely on articulatory and proprioceptive information to compensate for the perturbation as they would need to know the position of the bite-block in relation to the intended constriction degree and the tongue position to successfully produce the intended vowel. On the other hand, same results provide evidence for the idea that speakers do not map specific articulatory configurations to a certain acoustic output since they are able to produce the intended speech sound with different articulatory configurations when required. The later observation was repeatedly made across a range of oral-articulatory perturbation studies with artificial palates (e.g., Hamlet & Stone, 1976; McFarland & Baum, 1995), lip tubes (e.g., Savariaux et al., 1995; Aubin & Ménard, 2006), teeth prostheses (Jones & Munhall, 2003), and load perturbations of the jaw (e.g., Folkins & Abbs, 1975; Abbs & Gracco, 1984; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984).

The phenomenon that speakers can produce an intended acoustic goal by employing different articulatory strategies is more generally known as motor equivalence and was also observed outside of perturbation experiments. For instance, during the production of the English /r/ its characteristic acoustic properties can be produced either by raising the tongue tip and lowering the tongue dorsum or by lowering the tongue tip and raising the tongue dorsum (Westbury, Hashi, & Lindstrom, 1998; Zhou et al., 2008). Further examples of motor equivalent relations in speech can be found in Hughes and Abbs (1976) and Perkell, Matthies, Svirsky, and Jordan (1993). For an overview of motor equivalence in speech production see Perrier and Fuchs (2015).

Due to technical innovation, in more recent years it has become possible to study articulatory-acoustic relations in the context of specific acoustic parameters by altering, in near real-time, such parameters as fundamental frequency (f0; Jones & Munhall, 2000), vowel formants (F1, F2; Houde & Jordan, 1998), and frication noise (Shiller et al., 2009) in speakers' auditory feedback. During an auditory perturbation experiment speakers are usually asked to repeatedly produce a short word or syllable while they are hearing themselves over ear- or headphones. When the perturbation is applied, such that, for instance, F1 is increased in the auditory feedback, speakers begin to perceive the experimental stimulus differently from how they actually produce it. For instance, if the vowel /ɛ/ is perturbed in this way, it starts to resemble /æ/. In response, speakers typically decrease F1 in their produced speech in order to restore their percept of the intended word. That means that the produced F1 may become comparable to that of a /ɪ/. This compensatory response is generalizable across most investigated acoustic parameters and shows that under auditory perturbation speakers try to maintain their auditory target by adjusting their articulatory movements. For vowel formants, compensation was previously demonstrated with native speakers of English (e.g., Mitsuya et al., 2015), French (Mitsuya, Samson, Ménard, & Munhall, 2013), and Mandarin Chinese (Cai, Ghosh, Guenther, & Perkell, 2010).

The magnitude of the compensatory response is known to be influenced by some perceptual processes. For instance, Villacorta, Perkell, and Guenther (2007) demonstrated that speakers' individual auditory acuity scores with regard to F1 frequency significantly correlated with the magnitude of their compensatory response during F1 perturbation. Furthermore, Niziolek and Guenther (2013) findings suggest that the magnitude of the compensatory response does not depend purely on acoustic distance between produced and perturbed targets, but can become much larger when the perturbation results in a phonemic category change of the perturbed vowel compared to only sub-phonemic changes. This is consistent with findings by Reilly and Dougherty (2013) who showed that speakers react less strongly to F1 perturbations if F1 constitutes a less important perceptual cue for the identification and discrimination of the perturbed vowel.

While all – oral-articulatory as well as auditory – perturbation studies reviewed so far suggest that the acoustic dimension is essential for producing an intended speech sound, a few experiments were also able to demonstrate the independent contribution of the somatosensory feedback during speech production. Particularly, the experiments by Tremblay et al. (2003) and Nasir and Ostry (2006) demonstrated that speakers compensate for jaw movement perturbations delivered by a robotic arm even if these do not have any measurable effect on the acoustic outcome of speakers' articulation. In these studies, approximately 50 percent of speakers compensated for the applied jaw perturbation which suggests that speakers perceived the somatosensory errors and tried to correct for these. Furthermore, the authors did not find the same compensatory effects on trials where speakers produced opening non-speech jaw movements.

1.2 Relation between auditory and somatosensory feedback

Since the role of auditory and somatosensory feedback has been investigated in most studies separately, it is not completely clear how both feedback signals are incorporated during speech production. More recently, Lametti et al. (2012) hypothesized that speakers might exhibit individual preferences regarding the sensory feedback channel they predominantly employ to monitor their own speech production. In their study the authors investigated participants' responses to a simultaneous somatosensory jaw and auditory F1 perturbation. Lametti et al.'s results revealed a minor negative correlation across participants between the amount of observed somatosensory and auditory compensation. This means that speakers who changed their jaw position, compensating for the somatosensory perturbation, did not significantly change their F1 during auditory perturbation and vice versa.

Unlike Lametti et al.'s study, in which F1 was perturbed upwards causing an auditory error compatible with the simultaneously applied jaw opening perturbation, perturbations applied to either auditory or somatosensory feedback signal may induce incompatible information in the speech production system. Particularly, while auditory feedback might signal an error, somatosensory feedback might indicate that the appropriate target was achieved. Some authors suggested that such incongruence between feedback signals could be a potential reason for partial compensations observed during formant perturbation (MacDonald, Goldberg, & Munhall, 2010; Katseff et al., 2012). That is, these authors hypothesize that when acoustic parameters of speakers' speech are diverted from the target, speakers will compensate for the acoustic error as long as the discrepancy between the auditory and somatosensory feedback signals does not become too large.

The importance of congruency between specific auditory and somatosensory targets was, however, questioned by some empirical findings. For instance, Rochet-Capellan and Ostry (2011) demonstrated that speakers can simultaneously use multiple articulatory configurations to produce one vowel. To show this, the authors let their speakers repeatedly produce the words 'head', 'bed', and 'ted' while F1 in the vowel /ɛ/ was perturbed in opposing directions in two stimuli and remained unchanged in a control stimulus. On average, speakers were able to consistently compensate for the opposing F1 perturbations as well as to keep their F1 unchanged in the control stimulus. In other words, in that particular case speakers employed three different articulatory configurations to produce the vowel /ɛ/ as long as their auditory feedback suggested that they achieved the F1 value corresponding to their usual acoustic target of this vowel.

Consistent with Rochet-Capellan and Ostry's findings, Feng et al. (2011) demonstrated that speakers compensated for the closing perturbation of the jaw during the production of vowels /ɛ/ and /æ/ only when it resulted in a measurable F1 decrease. When, at the same time, F1 was increased in participants' auditory feedback to match their intended acoustic output, they no longer compensated for the jaw perturbation. The authors took this finding as a strong evidence for the hypothesis that the relation between articulatory and acoustic feedback signals is characterized by a hierarchy where the acoustic dimension takes precedence over the articulatory dimension.

In summary, the findings of cross-modal perturbation studies by Feng et al. (2011) and Lametti et al. (2012) suggest that as long as the somatosensory and auditory feedback are perturbed independently or are changed simultaneously in a consistent manner, speakers mostly compensate for errors in both signals, possibly with individual preferences for one or the other feedback signal. However, when somatosensory and auditory signals provide inconsistent feedback, speakers seem to prefer the auditory signal for error monitoring. The last observation is independently supported by Rochet-Capellan and Ostry (2011) results.

The discussed findings on the interaction between auditory and somatosensory feedback are, however, mostly limited to F1 perturbation in open-mid and open vowels /ɛ/ and /æ/. Therefore, any generalizations based on these results may be difficult as more recent auditory perturbation research suggests that the contribution of the somatosensory feedback to the production of vowels might differ across different phonemes. In particular, the compensatory magnitude to auditory perturbations is expected to be weaker for close vowels such as /i/ compared to non-close vowels since the former are characterized by a larger physical contact between active (tongue) and passive (hard palate) articulators (see Mitsuya et al., 2015). In other words, the incongruence between the auditory and somatosensory feedback signals might play a more important role for close vowels compared to non-close vowels.

Furthermore, in the case of downward F1 perturbation speakers' compensatory movements are more strongly restricted by physical boundaries imposed by the palate or the upper incisors. That means that, even when adjusted, speakers' articulatory movements may well remain within the limits of their unperturbed speech production such that the discrepancy between the auditory and somatosensory feedback signals remains too small to have an observable effect on speakers' compensatory behavior. Thus, the first goal of the current investigation is to re-examine the role of congruency between auditory and somatosensory targets under more restricting articulatory conditions. Among other things, we aim to examine the influence of coarticulation on the compensatory magnitude.

1.3 Role of auditory feedback for consonants

The hypothesis that, in some instances, the somatosensory dimension may prevail over the auditory dimension during production of a speech sound can be traced back to the distinction between vowels and consonants that is often made in the speech production research (e.g., Guenther et al., 1998; Perkell, 2012). At the core of this distinction lies the idea that consonants are rather defined in the articulatory target space while vowels are defined in the auditory space. Proceeding from this idea, it follows that auditory feedback should play a minor role during the production of consonants. However, despite the steady progress which has been made in the study of auditory feedback in the past 20 to 25 years, its role for the production of consonants is still debated. One of the reasons for this situation is the fact that auditory perturbation has been almost exclusively applied to alter acoustic characteristics of vowels but not those of consonants (for an overview see Caudrelier & Rochet-Capellan, 2019). As it turns out, there are practical and technical limitations associated with the auditory perturbation of consonants.

In contrast to vowels, which are continuous sound streams whose acoustic spectra contain prominent features (i.e., formants), many consonants are short bursts of auditory noise. Thus, while speakers can be asked to prolong their vowels to provide sufficient time for the perturbation algorithm to identify and target a specific frequency, this approach is not feasible for consonants. We are aware only of two attempts to cope with these limitations, namely the ones by Shiller et al. (2009) and by Casserly (2011) who investigated auditory perturbation of frication noise with English speakers. Similarly to vowels, fricatives are relatively long sounds which make it possible to expose speakers to longer periods of auditorily perturbed signal. However, due to the random nature of the fricative spectra it is not possible to target a specific frequency for perturbation but rather the whole spectrum can be shifted which effectively leads to higher energy concentration in higher or lower frequency bands.

In both mentioned studies, participants had to produce sibilant fricatives /s/ or /ʃ/ while their acoustic spectrum was alternated in near real-time such that the spectral center of gravity (COG) was decreased or increased in speakers' auditory feedback. While Shiller et al. (2009) found that their participants raised the COG compensating for the applied shift, Casserly (2011) observed three different behavioral patterns among her participants (no reaction, raising of the COG, and lowering of the COG). Although these results generally suggest that perturbation of auditory feedback affects speakers' consonant production, in contrast to the mostly consistent compensatory responses observed during auditory perturbation of vowels, perturbation of fricatives appears to cause a more variable response behavior. In fact, when we turn our attention to studies that investigated speech production under oral-articulatory perturbation, we observe analogous differences of compensatory variability between fricatives and vowels.

One of the earliest works investigating speakers' reaction to oral-articulatory perturbation of alveolar consonants (including alveolar fricatives) was the electropalatographic (EPG) study by Hamlet and Stone (1978). In this study, participants were required to produce the target segments with inserted artificial palates of 4 mm thickness. Immediately after the insertion of the palate, the authors found that there were more palatal contacts during the production of all tested consonants suggesting significant tongue overshoot. In the case of the fricatives, the overshoot led to a production of stops since the thick palate also reduced the groove size during fricative production. After a two week period of wearing and speaking with the artificial palate, participants were able to decrease the number of palatal contacts by changing the constriction location (either retracting or advancing their tongue). However, even after two weeks of adaptation the articulatory variability across participants remained high.

Similar results concerning the variability during the perturbed production of /s/ sounds are reported by Flege et al. (1988) who investigated speech of English and Arabic speakers perturbed by a bite-block. Contrary to Hamlet and Stone (1978) who investigated speakers' adaptation over a lengthy period of time, Flege et al. (1988) examined short-time adaptation within the same experimental session. Speakers were recorded immediately after the insertion of a bite-block during two experimental blocks that were interrupted by a 10-minute conversation with the bite-block in place. As Hamlet and Stone (1978), Flege et al. (1988) demonstrated that among their participants some compensated for the applied perturbation by changing the constriction location. Furthermore, the results suggested that the direction of the change was not consistent across participants. Additionally, a few participants varied the groove size between the unperturbed and perturbed conditions.

Finally, Honda et al. (2002) and Brunner et al. (2011), who investigated perturbed production of alveolar fricatives in Japanese and German speakers by means of artificial palates and electromagnetic articulography (EMA), demonstrated that along with compensatory adjustments of the constriction location and groove size some speakers also changed the jaw height to compensate for the perturbations. Furthermore, Brunner et al. (2011) found a complementary relation between the jaw height and the groove size of the tongue: speakers who lowered the jaw adjusted the tongue to a convex shape, while speakers who retained a high jaw position exhibited a more concave tongue shape. Since both adjustments led to more high-frequency energy in the feedback signal, this finding suggests that participants may employ different articulatory strategies to compensate for a perturbation.

Taken together, the reviewed findings suggest that the successful compensation in fricatives is articulatorily highly demanding as it requires fine control of a series of articulatory parameters (constriction location, tongue grooving, and jaw height) to accurately produce the target sound. In this respect, fricatives appear to differ significantly from vowels. McFarland and Baum (1995), for instance, found that French speaking participants were able to almost completely retain the spectral properties of vowels but not of fricatives after 15 minutes of speaking with a bite-block. The authors explicitly attribute this difference to higher articulatory demands of accurate fricative production.

In a follow-up study, McFarland, Baum, and Chabot (1996) came to the same conclusion after using artificial palates of 3 mm and 6 mm thickness to perturb the alveolar fricatives /s, ʃ/ and the vowels /i, a, u/. The authors found that immediately after the insertion of the palates, the fricative spectra were severely affected by the perturbation and the subsequently assessed perceptual goodness ratings for the perturbed fricatives were rather low. On the other hand, vowel production was only slightly affected by the perturbation overall. This finding corroborated earlier results by Hamlet and Stone (1976) who investigated vowel production of English participants perturbed with artificial palates of three different thicknesses. In McFarland et al. (1996) study, participants were able to only slightly improve their fricative acoustics after a 15-minute conversation with the artificial palate.

The above results demonstrate that compensating for perturbation in fricatives is generally more demanding compared to vowels, probably as it requires accurate control of a larger set of articulatory parameters. Although this hypothesis can potentially explain the compensatory variability observed across auditory perturbation studies by Shiller et al. (2009) and Casserly (2011), it also implies that to successfully assess the magnitude of compensation in fricatives it might be crucial to investigate different or additional spectral parameters aside from COG.

This idea is further supported by the results of the study by Jones and Munhall (2003) who evaluated compensation during the production of /s/ in English speakers whose speech was perturbed by extending their upper incisors by 5 to 6 mm without affecting their bite. During the experimental session, participants had to complete two adaptation phases each consisting of 15 blocks where in each block they had to produce the syllable /tas/ 10 times.

After the insertion of the teeth prosthesis, the spectral COG dropped on average by 1689 Hz. Initial COG analyses of the subsequent adaptation blocks did not show any significant compensatory improvements. However, the authors observed that the slope difference between the spectral regression lines for the regions between 0.5–2.5 kHz and 2.5–8 kHz, that are associated, among others, with the jaw height, and which initially increased when the teeth prosthesis was inserted, approached the unperturbed condition over the course of the adaptation phases. These findings demonstrate that, while COG previously proved to be an appropriate measure to differentiate between different voiceless fricative sounds (e.g., Forrest, Weismer, Milenkovic, & Dougall, 1988), it fails to provide an adequate measure for speakers' compensatory adjustments.

At this point, it is far from clear whether the observed compensatory variability during perturbation of fricatives is mostly attributable to a diminished role of auditory feedback for consonant production or rather to the complexity of articulatory control required for a successful adaptation. Thus, the second goal of the current investigation is to gain a clearer understanding of the role of auditory feedback for consonant production by extending on the studies by Shiller et al. (2009) and Casserly (2011), and developing a formal analysis which allows us to assess the amount of auditory adaptation in fricatives in a more objective way.

1.4 Multidimensional (acoustic and articulatory) analysis of adaptation

In order to avoid spurious results, a central feature of an objective adaptation analysis has to be its independence of specific acoustic parameters which are chosen to measure compensatory adjustments (i.e., COG, spectral shape, regions of spectral energy concentration). In other words, our goal is to estimate the overall magnitude of compensatory adjustments independently from their specific acoustic manifestations. Ultimately, this method should make it possible to numerically compare compensatory magnitudes across fricatives and vowels, which are characterized by not directly comparable spectral parameters (e.g., COG vs. F1/F2). Thus, the third goal of the current investigation is to provide a meaningful comparison between compensatory responses to auditory perturbation of fricatives and vowels to further deepen our insights into articulatory-acoustic relations across different speech sounds. With this comparison, we also intend to examine the influence of articulatory complexity of an adaptation task on its successful outcome.

Having an analytical instrument at hand which allows us to assess the magnitude of adaptation across different units of measurement, we can explore compensatory effects across acoustic and articulatory dimensions by assessing adaptation magnitude in spectral and spatial parameters recorded in parallel by an EMA system during auditory perturbation. Although it is possible to infer the overall degree of adaptation solely on grounds of acoustics, as it is regularly done in perturbation research, we believe that in order to fully reconstruct how speakers implement their compensatory adjustments it is necessary to examine articulatory movements directly. By doing so, we intend to advance our comprehension of the interaction between articulatory and acoustic dimensions during production of speech sounds.

A conclusive answer to the question how speakers implement compensatory adjustments in reaction to auditory perturbation is far from trivial since the degree of adaptation effects might differ across acoustic and articulatory dimensions due to motor equivalence. Specifically, it is imaginable that speakers adjust a certain acoustic parameter in reaction to the applied perturbation, but achieve the intended acoustic goal by means of different articulatory strategies. As far as we know, there are only a handful of studies addressing this issue, which, however, report conflicting evidence and consequently draw different conclusions.

For instance, in the previously discussed study by Feng et al. (2011), the authors were able to identify articulatory displacements in the vertical dimension of the jaw and tongue tip sensors consistently across all eight investigated English speakers. The observed displacements were characterized, depending on the stimulus and sensor, by a magnitude of 0.5–2 mm and were observed in parallel to compensatory changes that occurred to F1 in the acoustic dimension. Similarly, Trudeau-Fisette, Tiede, and Ménard (2017), who investigated articulatory displacements as reaction to F2 perturbation in 10 French speakers, observed lip and tongue sensors displacements that occurred in parallel to acoustic compensation. The articulatory displacements had the magnitude of 0.5–4 mm depending on the sensor; stronger changes were observed in the tongue tip and tongue back sensors compared to lip sensors.

On the other hand, Max, Wallace, and Vincent (2003), who investigated articulatory movements of three English speakers by means of EMA during F1/F2 perturbation, observed a consistent and distinct compensatory response in the acoustic dimension, but were not able to find consistent compensatory effects which occurred in parallel in any of the investigated articulatory parameters (jaw and tongue sensor positions). Max et al. (2003, p. 1053) interpreted their findings such that the observed articulator "displacements indicated motor-equivalent adaptation in the overall gestures".

Similar results are reported by Neufeld (2013) who perturbed F1/F2 in the production of 16 English speakers. Although the author observed typical compensatory responses in participants' acoustic signal on the group level, corresponding articulatory changes were characterized by a high degree of inter-speaker variability with regard to the identity as well as the number of affected articulators (lips, tongue, and jaw). Thus, the fourth goal of our investigation is to provide a systematic comparison of the observed compensatory changes across acoustic and articulatory dimensions and to investigate more closely their interaction.

1.5 Research aims

The current dissertation attempts to formulate a general research framework to investigate articula-tory-acoustic relations of speech sounds. To do so, in Chapter 2 we first re-examine the question whether the congruency between acoustic and articulatory error signals influences the magnitude of compensatory responses. In Chapter 3, we take a closer look at the role of auditory feedback for the production of fricatives and develop a formal analysis which allows us to assess the magnitude of adaptation independently from specific units of measurement. We evaluate this method in Chapter 4 by comparing compensatory behavior across auditory perturbation of fricatives and vowels. In Chapter 5, we examine the degree of adaptation across acoustic and articulatory dimensions by means of auditory perturbation and EMA. Finally, Chapter 6 will summarize and discuss the overall results with respect to the initially formulated research aims of the dissertation.

Chapter 2

The influence of coarticulatory and phonemic relations on compensatory responses

2.1 Introduction

The goal of the current chapter is to investigate the magnitude of compensatory effects associated with auditory perturbation applied to a close vowel since it is expected to be weaker compared to non-close vowels due to more prominent somatosensory feedback (Mitsuya et al., 2015).[1] In contrast to previous studies, we decided to perturb F2 frequency which, roughly speaking, is an indicator of the horizontal tongue displacement. In combination with the perturbation of a close vowel, this change allowed us to evaluate our hypothesis under more restricting somatosensory feedback conditions since, in order to compensate for the applied perturbation, our speakers were required to always retain the linguopalatal contact in the target vowel while moving its constriction location along the anterior-posterior axis.

As basis for our experimental design, we chose the Russian phoneme space since it includes the close central unrounded vowel /ɨ/ which is enclosed by the two close vowels /i/ and /u/. During the study, while participants repeatedly produced the target vowel /ɨ/ as part of CV syllables, its F2 was perturbed in opposing directions depending on the preceding consonant (/d/ or /g/). The bidirectional perturbation of F2 was intended to encourage participants

[1] The results of this chapter were previously published in Klein, Brunner, and Hoole (2019a).

to use two different compensatory strategies to produce the perturbed vowel, and the two different consonantal contexts (alveolar vs. velar) were chosen in such a way that the required compensatory directions were either compatible or incompatible with the usual coarticulatory relation between /dɨ/ and /gɨ/. The interaction between the place of articulation (alveolar vs. velar) and the F2 perturbation direction (upward vs. downward) was evenly counterbalanced between all participants. This resulted in two different coarticulatory configurations that were tested across two experimental groups.

Due to coarticulatory effects, speakers' F2 was expected to be higher in /dɨ/ compared to /gɨ/ before any compensatory response occurred. In one group, F2 was decreased in /dɨ/ and increased in /gɨ/ such that the initial F2 values of the two syllables were expected to drift apart over the duration of the experiment but to remain in their initial coarticulatory relation (/dɨ/ > /gɨ/). In the other group, the perturbation direction was swapped for both syllables, putatively preventing effective compensation as it was counteracted by coarticulatory effects, and the initial F2 values had to intersect for the two syllables over the duration of the experiment. This means that participants in the second group had to produce the two experimental syllables with an unusual coarticulatory pattern where F2 for /gɨ/ would be higher compared to /dɨ/.

One phonemic idiosyncrasy of Russian close vowels, which was expected to have a further influence on the magnitude of speakers' compensatory responses, is the fact that while /i/ appears only after palatalized consonants, both /ɨ/ and /u/ follow only non-palatalized ones (see Bolla, 1981, pp. 108-110). The acoustic feature which is most strongly associated with palatalized consonants is the height of the F2 frequency at the beginning of the following vowel which has highest values for /i/ compared to /ɨ/ and /u/. The height of F2 is such dominant for Russian speakers as perceptual cue to palatalization that even cross-spliced syllables containing non-palatalized consonants and vowels with high initial F2 frequency are mostly perceived as palatalized (Bondarko, 2005). Since the difference in F2 values between /i/ and /ɨ/ is on average substantially smaller (about 300 Hz) compared to /ɨ/ and /u/ (about 1000 Hz), it seems reasonable that, under equivalent amount of upward and downward F2 perturbation, participants should more readily classify instances of /ɨ/ perturbed towards /i/ as phonemic errors of palatalization compared to the perturbation of /ɨ/ towards /u/ which does not induce a change of the phonemic status of the perceived vowel. Consequently, considering the findings by Niziolek and Guenther (2013), speakers' compensatory responses should be on average stronger for the upward perturbation compared to the downward perturbation. However, since statistical analyses of the compensatory behavior provided in Niziolek and

Guenther (2013) were performed exclusively on the group level, it remains unclear whether a phonemic category change of the target vowel influences the magnitude of compensatory response equally across all speakers.

Our experiment was designed to provide information about how different articulatory and auditory factors, known to have a major influence on the general compensatory response to auditory perturbation on its own, may impact speakers' individual compensatory performance. Thus, we might observe distinct compensatory patterns within and across speakers associated with specific perturbation conditions. With the current experiment, we pursued two goals. First, we wanted to know whether speakers would use two distinct compensatory strategies to produce one vowel which is characterized by a high degree of linguopalatal contact. More specifically with regard to this question, it is unknown whether speakers predominantly rely on auditory feedback even if it requires them to adopt compensatory strategies which considerably deviate from their usual coarticulatory pattern. Secondly, we were interested in the question of whether we would be able to observe any systematic individual differences dependent on the phonemic contrast between the produced and the perturbed acoustic signal.

Furthermore, to understand the spatial and temporal evolution of the adaptation process, we analyzed the formant data with generalized additive mixed models (GAMMs) which allowed us to observe non-linear changes in participants' responses to perturbation. By doing this, we seek to overcome the shortcomings of previous perturbation studies which concentrate on the comparison of speakers' performance between the beginning and the end of the perturbation task, i.e., during the first and the last 15-20 trials of an experiment. Unfortunately, this aggregation approach allows only for pairwise time-uncorrelated comparisons (e.g., Feng et al., 2011) while the evolution of the adaptation process is often presented only in the form of exploratory scatterplots (e.g., Rochet-Capellan & Ostry, 2011; Mitsuya et al., 2015).

2.2 Methods

2.2.1 Participants

Thirty-two native speakers of Russian (25 females, 7 males) without reported speech, language, or hearing disorders participated in the study. The participants were recruited from the pool of Russian exchange students and young professionals living in Berlin. The mean age of the group was 25.3 years and participants have spent on average 2.9 years in Germany at the time of the recordings. The study was approved by the local ethics committee and all speakers gave their written consent to participate in the study.

2.2.2 Equipment

Each experimental session was recorded in a sound attenuated booth. Participants were comfortably seated in a chair in front of a 19 inch LCD flat screen computer monitor which served to display the stimuli and experimental instructions. All texts were presented in Russian using Cyrillic font. Participants' speech signal was recorded with a Beyerdynamic Opus-54 neck-worn microphone, perturbed in real-time, and fed back via foam-tipped E-A-RTONE 3A insert earphones, which attenuated the air-conducted sound by approximately 25-30 dB while the microphone gain was set in such a manner that it resulted in an approximate feedback level of 75 dB SPL. This volume was chosen impressionistically during pilot recordings to strike a balance between listening comfort and masking effect of the bone conduction (see overview on bone conduction in Rahman & Shimamura, 2013). Throughout the recordings, the microphone gain was fixed across all participants. In order to shift the F2 frequency produced by the participants, tracking and real-time formant perturbation was accomplished with AUDAPTER, which is a C++ real-time signal processing application compiled to a MEX-file executable within a MATLAB environment (see for technical details Cai et al., 2010).

The correctness of the formant perturbation delivered by AUDAPTER was investigated with the help of an independent MATLAB script which calculated the formant values of the original and perturbed signals by means of LPC analysis of the signals' cepstra. Subsequently, the results of these calculations were visually inspected at random for all participants (Figure 2.1). Based on this procedure, we concluded that despite the applied F2 perturbation all formants (F1, F2, and F3) remained present in the modified signal as distinct peaks. Furthermore, we assured ourselves that the F2 peak was not interchanged with the F3 peak as result of the upward F2 perturbation.

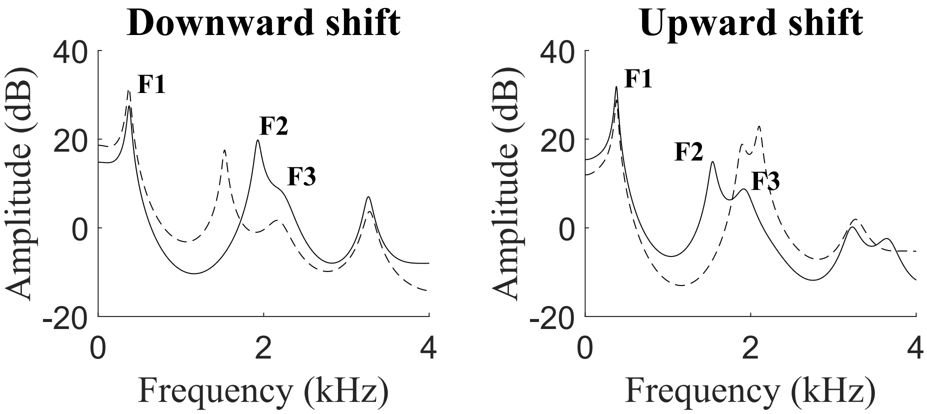

Figure 2.1: Example LPC-spectra of the original (solid lines) and perturbed (dashed lines) vowel /ɨ/ during the last shift phase of the experiment.

Direct measurements were performed to determine the delay of the feedback loop of the perturbation system by comparing the onsets of the acoustic response on its input and the output channels. The average delay amounted to 24 ms (*SD* = 4 ms). The original and perturbed signals were digitized and saved with a sampling rate of 16 kHz. Along with audio recordings, AUDAPTER stored data files containing the formant values (F1, F2, and F3) tracked during each stimulus production.

2.2.3 Experimental procedure and speech stimuli

For our study we chose Russian, since its vowel inventory includes the close central vowel /ɨ/ which is enclosed within the F2 space on each side by the two vowels /i/ and /u/. This constellation allowed us to investigate a two-sided compensation in /ɨ/ with bidirectional perturbation of the F2 frequency. The vowel /ɨ/ has a special status in the Russian vowel system since it never appears after palatalized consonants (Bolla, 1981, p. 109). This detail will become important during the discussion of participants' compensatory behavior in section 2.3 on page 23.

Each recording session lasted for approximately 20-25 minutes and consisted of four experimental phases: baseline, shift 1, shift 2, and shift 3 phases. Before the start of the first experimental phase, participants completed a few practice trials with unrelated speech material to ensure they understood the task and were able to perform it accurately.

During the 60 baseline trials, no auditory perturbation was applied and on each trial, which had an approximate duration of 2 seconds, participants were visually prompted to produce one of the four CV syllables /di/, /dɨ/, /gɨ/, and /gu/. This was done to assess participants' initial formant space whose structure was expected to have an influence on their compensatory behavior. The interstimulus interval between trials was approximately 1.5 seconds long. The visual presentation of the stimuli was controlled by a customized MATLAB software package developed at the Institute of Phonetics and Speech Processing, LMU Munich.

During the three following shift phases, of which each lasted for 50 trials, on each trial participants were prompted to produce the close central unrounded vowel /ɨ/ either in alveolar or velar consonantal context (i.e., /dɨ/ or /gɨ/). The F2 was shifted during the production of the vowel either upwards or downwards depending on the context (Table 2.1). Within each shift phase all stimuli were presented in pseudorandom order. That means that a participant could experience one perturbation direction on one trial and the other direction on the immediately following one. On the other hand, the same perturbation direction was never applied on more than two consecutive trials.

Table 2.1: Summary of the experimental conditions.

	Group A (compatible)		**Group B** (incompatible)	
Constriction	alveolar (/dɨ/)	velar (/gɨ/)	alveolar (/dɨ/)	velar (/gɨ/)
F2 perturbation	downward	upward	upward	downward

The magnitude of the applied F2 perturbation amounted to 220 Hz in the first shift phase and increased incrementally for each shift phase by 150 Hz reaching 520 Hz in the last shift phase. The amount of perturbation did not change within each shift phase. This perturbation scale was chosen on grounds of previous piloting as striking an optimal balance between moderate initial F2 perturbation and the experimental goal to let participants learn two strongly distinct compensatory strategies for the vowel /ɨ/.

Participants were instructed to produce all syllables with prolonged vowels. The prolongation of the vowel segments made for one thing the formant tracking more reliable and for the other maximized the amount of time during which participants were exposed to perturbed vowels. To keep the prolongation duration somewhat consistent across participants, they were assisted by a visual go-and-stop signal during their production. The go-and-stop signal had the form of a frame. Between the trials, while the frame stayed red, the stimulus for the upcoming trial appeared on the display and stayed within the frame. When a trial started, the frame color turned green which gave participants the signal to initiate their response. The resulting average duration of the produced vowels was 952 ms (*SD* = 270 ms).

Following the experimental session, all participants were asked if they noticed anything unusual in their auditory feedback during the experiment. A few of the participants reported that their pronunciation was different from what they were used to or that they perceived an acoustic difference between the syllables /dɨ/ and /gɨ/. Most participants attributed these pronunciation differences to the effect of listening to own speech on audio recordings, so when asked if and how these differences affected their production, participants reported to have ignored these. From previous research, however, it is known that participants are not able to voluntarily control their reaction to auditory perturbation even if they are told to ignore it (Munhall, MacDonald, Byrne, & Johnsrude, 2009). Furthermore, participants were asked whether they became aware of any systematic position changes of their articulators (specifically, the tongue position) in the course of the experiment. None of the participants reported to have noticed anything unusual.

2.2.4 Interaction between perturbation and coarticulatory effects

The interaction between the place of articulation (alveolar vs. velar) and the perturbation direction (upward vs. downward) was evenly counterbalanced between the 32 participants which resulted in two different coarticulatory configurations represented by experimental group A (14 females, 2 males) and group B (11 females, 5 males). Due to coarticulation, baseline F2 was expected to be higher in /dɨ/ compared to /gɨ/. In group A, F2 was decreased in /dɨ/ and increased in /gɨ/, such that compensatory movements were expected to act in the same direction as coarticulatory effects (compatible condition; Figure 2.2A). In that case, F2 values produced for /dɨ/ and /gɨ/ during the baseline phase were expected to drift apart during the shift phases of the experiment due to compensation but remain in the same relation (/dɨ/ > /gɨ/). In group B, the perturbation direction was swapped for both syllables and the baseline F2 values were expected to intersect for the two syllables during the shift phases of the experiment (incompatible condition; Figure 2.2B).

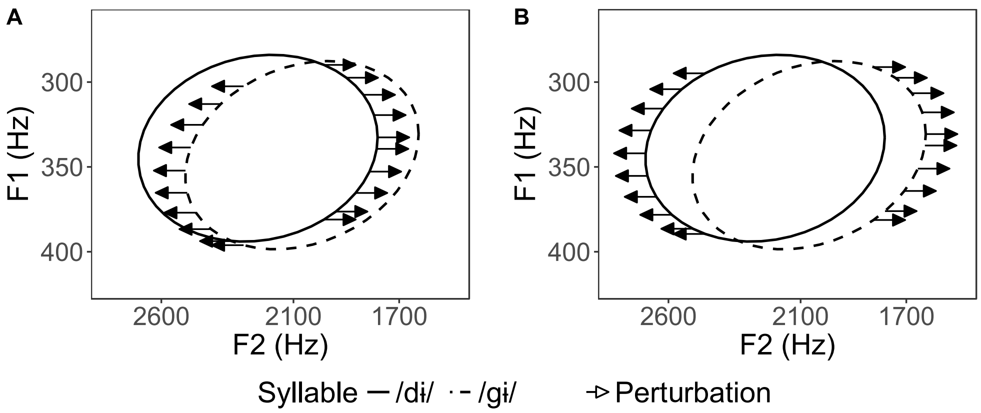

Figure 2.2: Distinct perturbation configurations applied for the experimental groups A and B. The ellipses for the two syllables /dɨ/ (solid line) and /gɨ/ (dashed line) are plotted based on the F1-F2 data of 480 baseline repetitions each. Note that the directions of the figure's axes are reversed. (A) For the two syllables to drift apart, F2 was increased in /gɨ/ and decreased in /dɨ/. (B) For the two syllables to intersect, F2 was increased in /dɨ/ and decreased in /gɨ/.

2.2.5 Data pre-processing and statistical analyses

All recordings of 32 participants amounted to 6720 trials. The onset and offset of the vowel segment produced on each trial were labeled manually based on spectrograms using MATLAB'S graphic input facilities. Subsequently, the corresponding formant vectors were extracted from AUDAPTER's data files based on the labeled onset and offset boundaries. The middle 50 percent portion of each formant vector was used to compute the formant means produced on each trial.

All statistical analyses were performed in R (version 3.4.1; R Core Team, 2017). To understand whether speakers' initial formant space could potentially provide an explanation for the occurrence of certain compensatory patterns, we performed an analysis on non-perturbed trials to derive the mean formant frequencies (F1, F2, and F3) for the vowels contained in the syllables /di/, /dɨ/, /gɨ/, and /gu/. By means of pairwise t-tests, we examined the relation between different formant frequencies of the perturbed vowel /ɨ/ as well as its relations to the neighboring sounds /i/ and /u/. To control for the use of multiple t-tests comparisons, the p-value was adjusted applying the Bonferroni correction. Since for each dependent variable (F1-F3), three comparisons were made (/di/ vs. /dɨ/, /dɨ/ vs. /gɨ/, and /gɨ/ vs. /gu/), the alpha level of .05 was set to .05/3 = 0.0167.

An additional analysis was performed comparing mean formant frequencies (F1, F2, and

F3) of the two syllables /dɨ/ and /gɨ/ produced by speakers on non-perturbed trials with formant values produced during the last shift phase of the experiment. By means of pairwise t-tests, we examined how both experimental groups A and B adjusted to the applied perturbation in the target vowel /ɨ/. To control for the use of multiple t-tests comparisons, the p-value was adjusted applying the Bonferroni correction. Since for each dependent variable (F1-F3), two comparisons were made (/dɨ/ vs. adapted /dɨ/ and /gɨ/ vs. adapted /gɨ/ for each group A and B), the alpha level of .05 was set to .05/2 = 0.025.

To identify and classify individual compensatory patterns, we recomputed individual compensation magnitudes reached by participants during the last shift phase of the experiment as percentage scores. This allowed us to subdivide participants into different groups with respect to their compensatory performance for both perturbation directions (upward vs. downward).

To examine average formant changes in participants' production of the two syllables /dɨ/ and /gɨ/ across the four experimental phases, we fitted a generalized additive model (GAM; Hastie & Tibshirani, 1987) which is a significant extension of a generalized linear regression model as it allows the modeling of non-linear relationships between the dependent and independent variables (Wood, 2017a). Therefore, GAMs are much more flexible compared to a linear regression model. The non-linear relationships are modeled via complex functions (smooths) which are constructed from ten basis functions (e.g., linear, quadratic, and cubic functions) with an adjustable number of basis dimensions. The number of basis dimensions indicates the upper limit of how complex the constructed function can be and is estimated directly from the data during the modeling process. That means that the usage of GAMs does not require a predefined specification of a certain (non-linear) function as it is derived directly from the data. To prevent overfitting of the data, i.e., modeling of functions which are too complex and therefore might obscure any generalizable patterns in the data, GAMs are estimated using penalized likelihood estimation and cross-validation (see for details Wood, 2006). One further advantage of GAMs is the possibility to include random effects into the model structure to account for individual response variability across but also within speakers. To denote the inclusion of random effects in the fitted model, it is dubbed generalized additive *mixed* model (GAMM). For a hands-on introduction to GAMMs with a focus on dynamic speech analysis see Wieling (2018).

The GAMM offers three main advantages for analyzing the data from auditory perturbation experiments. First, it is possible to analyze the data as a function of time which allows the investigation of the whole adaptation process rather than just its outcome. Secondly, the nonlinearity of parameter smooths does not make any assumptions regarding the temporal or spatial characteristics of the adaptation process. Finally, the parameter smooths can be estimated including random effects which allows one to capture the individual variability of the adaptation process which is important considering the variability which is repeatedly observed in perturbation studies.

Prior to building the GAMM model, participants' raw formant frequencies were normalized by subtracting each participant's mean formant frequency produced during the baseline phase for the respective syllable (/dɨ/ or /gɨ/). This was done to exclude participant-specific differences regarding their absolute formants magnitude (e.g., due to gender differences). By means of this normalization, the average F1, F2, and F3 values for /dɨ/ and /gɨ/ were set at zero for the baseline phase.

Subsequently, using the *mgcv* package (version 1.8-19; Wood, 2017b) we fitted one GAMM model for each formant (F1, F2, and F3) with normalized frequency averaged across all participants and all experimental trials as dependent variable. The data of the unperturbed syllables /di/ and /gu/, which were uttered by participants only during the baseline phase, were not included in the resulting GAMMs. All GAMM models were evaluated, interpreted, and visualized by means of the *itsadug* package (version 2.3) by (van Rij, Wieling, Baayen, & van Rijn, 2017).

In the model structure, we included random factor smooths with an intercept split for the perturbation direction (upward vs. downward) in order to assess (potentially non-linear) individual compensation magnitude differences over the course of the experiment. The model also included a fixed effect which assessed the 'constant' effect of the perturbation direction independently from the individual and temporal variation. The interaction between the perturbation direction (upward vs. downward) and the experimental group (A vs. B) did not significantly improve the model fit, as revealed by the goodness of fit assessed by the Akaike Information Criterion (AIC). Therefore, the data of both experimental groups (A and B) was pooled together for the GAMM analysis.

2.3 Results

We will start our review of the results by summarizing the initial formant frequencies produced by all participants during the baseline phase of the experiment when no perturbation was applied. After that, we will compare the adapted formant space of experimental groups A and B produced during the last shift phase to the baseline formants. Concluding this comparison, we will discuss different compensatory patterns observed among participants. Finally, we will turn to the presentation of speakers' average compensatory behavior. In particular, we will discuss changes in F1, F2, and F3 frequencies produced by participants over the course of the whole experiment.

2.3.1 Initial formant space

The mean F1, F2, and F3 frequencies produced by all participants during the baseline phase are summarized in Figure 2.3. Overall, the formants observed in this study for the vowels /i/, /ɨ/, and /u/ were comparable with previous descriptive studies of the Russian vowel space (e.g., Lobanov, 1971). The average within-speaker differences between F1 values of the investigated vowels were partially statistically significant, but exhibited rather minor effect sizes, and amounted to 11.84 Hz between /di/ and /dɨ/ (95% CI [–15.89 –7.80], t = 5.748, p < .001), to –3.80 Hz between /dɨ/ and /gɨ/ (95% CI [–8.06 0.46], t = –1.750, p = .08), and to 4.75 Hz between /gɨ/ and /gu/ (95% CI [–8.77 –0.72], t = –2.317, p = .02). The F2 values, on the other hand, revealed statistically significant within-speaker differences between the investigated vowels which also exhibited prominent effect sizes. The average F2 difference between /di/ and /dɨ/ was 241.22 Hz (95% CI [215.67 266.76], t = 18.532, p < .001), 166.92 Hz between /dɨ/ and /gɨ/ (95% CI [140.78 193.07], t = 12.532, p < .001), and 1300.04 Hz between /gɨ/ and /gu/ (95% CI [1278.62 1321.46], t = 119.140, p < .001). As expected, F2 was higher for /dɨ/ compared to /gɨ/ likely due to coarticulation. Regarding the F3 values, the three syllables /dɨ/, /gɨ/ and /gu/ were very similar in contrast to /i/ where F3 was distinctly higher. Specifically, the F3 differences amounted to 556.36 Hz between /di/ and /dɨ/ (95% CI [525.90 586.82], t = 35.847, p < .001), 112.11 Hz between /dɨ/ and /gɨ/ (95% CI [85.27 138.95], t = 8.196, p < .001), and 77.95 Hz between /gɨ/ and /gu/ (95% CI [48.68 107.23], t = 5.226, p < .001).

From Figure 2.3, it is apparent that the vowel /ɨ/ is enclosed by its neighboring sounds

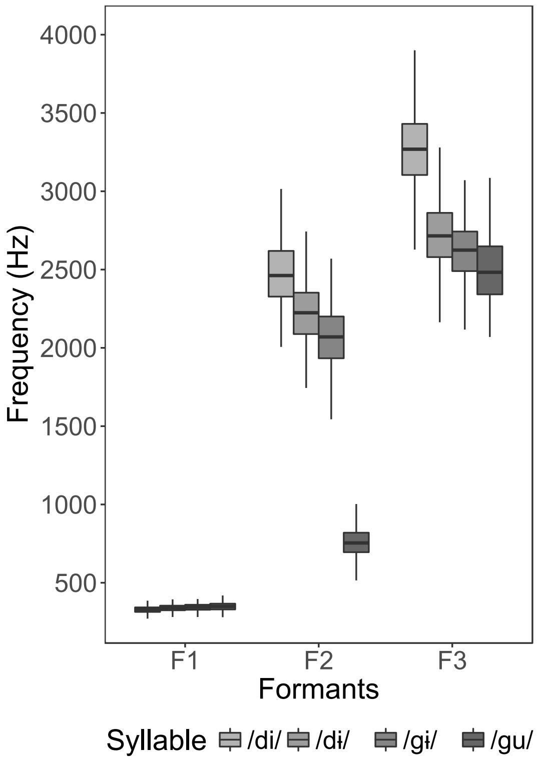

Figure 2.3: F1, F2, and F3 frequencies produced by each participant during the baseline phase (no perturbation) for the four syllables /di/, /dɨ/, /gɨ/, and /gu/.

/i/ and /u/ within the F2 dimension. However, there is also an asymmetry with respect to the F2 distance between the perturbed vowel /ɨ/ and the upper /i/ on the one hand, and between /ɨ/ and the lower /u/ on the other. Specifically, the F2 distance between /i/ and /ɨ/ is lower compared to the F2 distance between /ɨ/ and /u/. Furthermore, while the distance between F2 and F3 frequencies is quite high for both /di/ (–789.53 Hz, 95% CI [–818.84 –760.21], *t* = 52.86, *p* < .001) and /gu/ (–1751.29 Hz, 95% CI [–1776.46 –1726.13], *t* = –136.63, *p* < .001), the two frequencies are very close together in the perturbed syllables /dɨ/ (–474.31 Hz, 95% CI [–501.24 –447.54], *t* = 34.671, *p* < .001) and /gɨ/ (–529.20 Hz, 95% CI [–555.34 –503.07], *t* = –39.742, *p* < .001); the implications of these properties of the investigated vowel space on the current findings regarding compensation performance are addressed in the discussion section.

2.3.2 Adapted formant space

The mean F1, F2, and F3 frequencies produced by the experimental groups A and B for the vowel /ɨ/ during the baseline and the last shift phase are summarized in Figure 2.4. The average within-speaker changes of F1 values were statistically significant for group A and amounted to 11.23 Hz between non-adapted and adapted /dɨ/ (95% CI [14.80 7.67], *t* = 6.191, *p* < .001) and 11.22 Hz between non-adapted and adapted /gɨ/ (95% CI [15.19 7.25], *t* = 5.55, *p* < .001). For group B, the F1 changes were not statistically significant and amounted to 1.59 Hz between non-adapted and adapted /dɨ/ (95% CI [8.70 –5.53], *t* = 0.438, *p* = .66) and 1.31 Hz between non-adapted and adapted /gɨ/ (95% CI [5.75 –8.36], *t* = 0.364, *p* = .77).

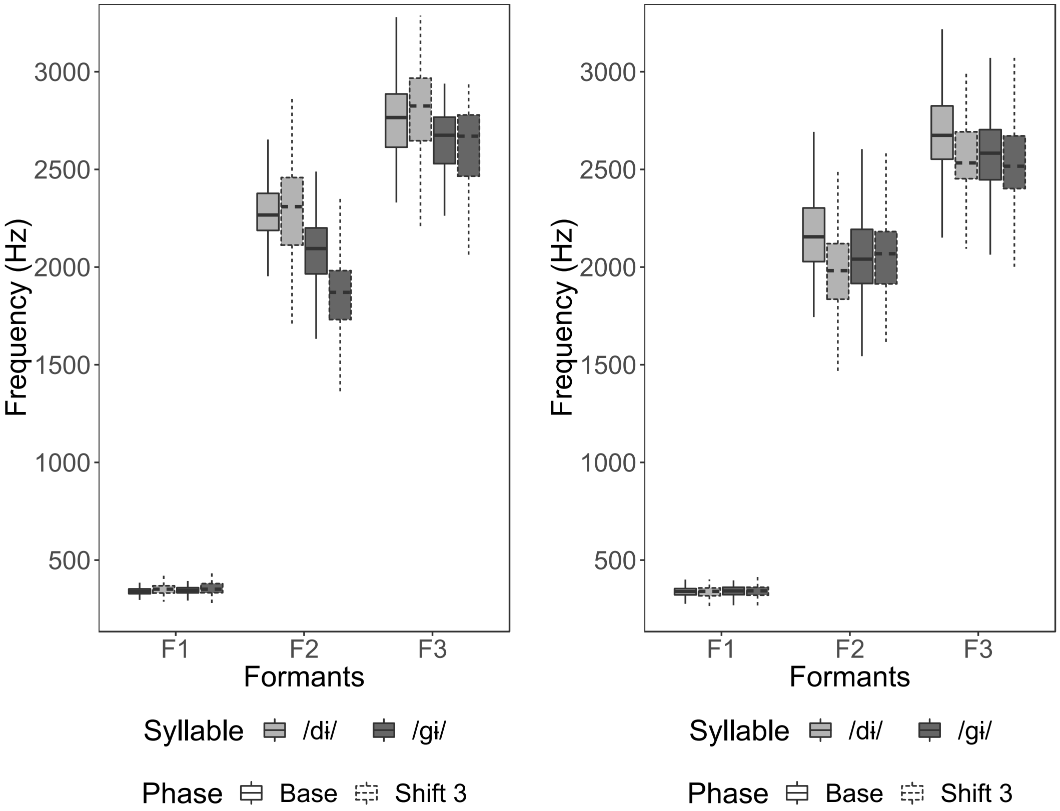

Figure 2.4: F1, F2, and F3 frequencies produced by participants of groups A (left panel) and B (right panel) during the baseline (no perturbation) and the shift 3 phase (520 Hz perturbation) for the syllables /dɨ/ and /gɨ/.

For group A, the average F2 change between non-adapted and adapted /dɨ/ amounted to 17.44 Hz (95% CI [50.52 14.83], *t* = 1.073, *p* = .28) and –216.01 Hz between non-adapted and adapted /gɨ/ (95% CI [–182.93 –249.09], *t* = –12.824, *p* < .001). For group B, the change between non-adapted and adapted /dɨ/ amounted to –193.56 Hz (95% CI [–158.97 –228.15], *t* = –10.992, *p* < .001) and 20.31 Hz between non-adapted and adapted /gɨ/ (95% CI [57.88 17.26], *t* = 1.062, *p* = .28). For group A, the F3 change between non-adapted and adapted /dɨ/ amounted to 43.62 Hz (95% CI [76.09 11.16], *t* = 2.639, *p* = .008) and –30.05 Hz between non-adapted and adapted /gɨ/ (95% CI [–0.46 –59.64], *t* = –1.995, *p* = .04). Finally, for group B, the F3 change between non-adapted and adapted /dɨ/ amounted to –133.74 Hz (95% CI [–95.57 –171.90], *t* = –6.885, *p* < .001) and –23.09 Hz between non-adapted and adapted /gɨ/ (95% CI [14.98 –61.17], *t* = –1.191, *p* = .23).

As predicted, for group A, the initial F2 values for the syllables /dɨ/ and /gɨ/ drifted further apart by the last shift phase of the experiment such that the F2 distance increased on average by 233.45 Hz. On the other hand, for group B, the initial F2 values for the syllables /dɨ/ and /gɨ/ moved towards each other on average by 173.25 Hz by the end of the experiment. Considering the average F2 distance of 124.53 Hz between /dɨ/ and /gɨ/ produced by speakers of group B during the baseline phase, we can conclude that, on average, the F2 values for both syllables intersected by the end of the experiment by 48.72 Hz.

2.3.3 Individual compensatory differences

To get a better overview of individual differences regarding the compensatory magnitude, we recalculated it as individual percentage scores that were reached by participants on average during the last shift phase (Table 2.2). Based on the direction of F2 changes as well as the magnitude of the F2 difference between both perturbation directions (upward vs. downward) observed for each participant during the last shift phase, we were able to group all participants. This initial grouping was then confirmed visually by examining the slopes of the compensatory changes (Figure 2.5).

With regard to F2 frequency, 10 participants exhibited what we dubbed a ‘symmetrical’ compensatory pattern since these participants adjusted their F2 values in the opposite direction to the upward and downward shifts by about the same compensatory magnitude (Figure 2.5A). Specifically, symmetrical adapters produced the target vowel /ɨ/ with an average compensatory magnitude of 30 percent (*SD* = 21) and 28 percent (*SD* = 12) on trials with upward and downward perturbation, respectively. The symmetrical pattern was observed among participants from the experimental group A (five speakers) as well as group B (five speakers).

Table 2.2: Average F2 and F3 changes calculated for each participant as percentage scores for the last shift phase of the experiment (520 Hz ≙ 100%). The table includes information regarding participants' coarticulatory configuration (gr.) and the applied perturbation direction (up vs. down). Based on their compensatory behavior, speakers were assigned into different groups: the symmetrical (sym.), the asymmetrical (asym.), and the negative (neg.) compensatory pattern (pat.). Three speakers did not display any specific compensatory pattern (-).

ID	gr.	**F2 (%)**		**F3 (%)**		pat.	ID	gr.	**F2 (%)**		**F3 (%)**		pat.
		up	*down*	*up*	*down*				*up*	*down*	*up*	*down*	
f1	A	–65	–16	11	–1	asym.	f15	B	–58	6	15	16	asym.
f2	A	5	–11	–14	0	-	f16	B	–3	23	–15	12	-
f3	A	–22	44	–2	27	sym.	f17	B	–40	16	–25	7	sym.
f4	A	–68	26	–37	20	sym.	f18	B	–63	2	–59	–7	asym.
f5	A	–66	9	5	14	asym.	f19	B	–11	24	3	6	sym.
f6	A	–36	4	0	40	asym.	f20	B	–66	–1	13	13	asym.
f7	A	–58	12	–3	3	sym.	f21	B	–54	–9	–76	–28	asym.
f8	A	–90	–67	–7	–17	neg.	f22	B	–37	24	–15	34	sym.
f9	A	–103	6	–8	1	asym.	f23	B	–4	40	–17	–12	sym.
f10	A	–32	–36	–42	–47	neg.	f24	B	–42	–12	–38	–24	neg.
f11	A	44	20	9	25	-	f25	B	–15	16	–3	–1	sym.
f12	A	–54	–25	–9	–23	neg.	m3	B	–31	–1	–24	–17	asym.
f13	A	–45	8	8	13	asym.	m4	B	–72	–21	–36	–13	asym.
f14	A	–19	39	–4	37	sym.	m5	B	–28	–10	–7	4	neg.
m1	A	–24	41	17	43	sym.	m6	B	–40	–28	–83	–39	neg.
m2	A	–30	0	–23	0	asym.	m7	B	–31	4	–47	–15	asym.

On the other hand, another 13 participants exhibited an 'asymmetrical' compensatory pattern where they on average compensated by 55 percent (SD = 21) for the upward shifts, but by 0 percent (SD = 9) for the downward shifts. That means that asymmetrical adapters produced /ɨ/ under downward perturbation with approximately the same absolute F2 values as in their baseline phase (Figure 2.5B). The asymmetrical pattern was observed among participants from the experimental group A (six speakers) as well as group B (seven speakers).

In contrast to asymmetrical adapters, another six participants considerably lowered their F2 frequency for both of the applied perturbation directions (Figure 2.5C). For that reason, we dubbed their compensatory pattern as the 'negative' one. These speakers adjusted their F2 frequency by –48 percent (SD = 23) during the upward and by –30 percent (SD = 21) during the downward perturbation. The negative pattern was observed among participants from the experimental group A (three speakers) as well as group B (three speakers).

Finally, there were three non-adapters – two speakers from group A and one speaker from group B – who did not exhibit a distinct compensation behavior which would match any of the patterns described above. Furthermore, their reaction to the applied perturbation appeared to be rather unsystematic, so we do not discuss their data any further.

With regard to F3 frequency, symmetrical adapters of F2 changed it on average by –8 percent (*SD* = 15) on trials with upward F2 perturbation and by +16 percent (*SD* = 18) on trials with downward F2 perturbation. Asymmetrical adapters adjusted their F3 values by –17 percent (*SD* = 30) on trials with upward F2 perturbation and by +1 percent (*SD* = 18) on trials with downward F2 perturbation. Finally, negative adapters changed their F3 frequency by –31 percent (*SD* = 30) during the upward shifts and –24 percent (*SD* = 18) during the downward shifts.

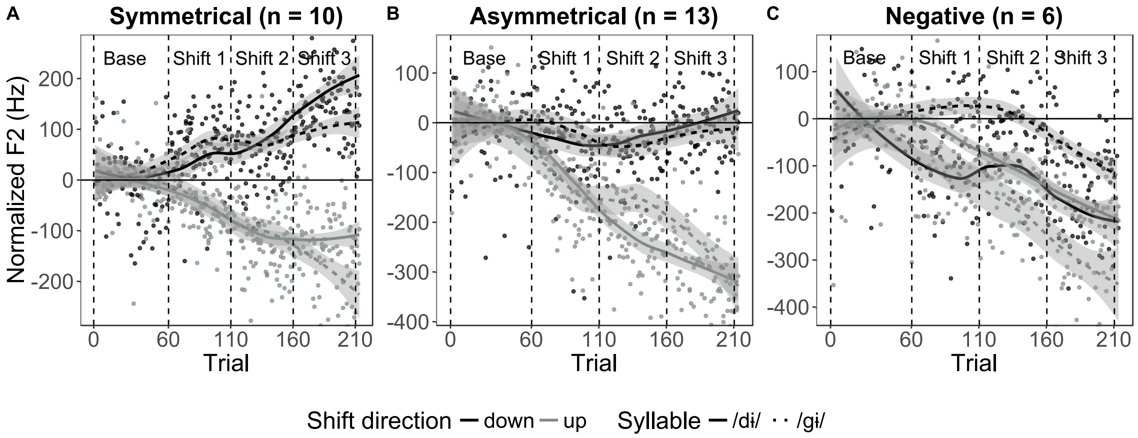

Figure 2.5: Average compensatory effects in F2 for downward and upward perturbation over the course of the three shift phases. Participants' data is divided into three subplots based on their compensatory pattern: (A) symmetrical adapters, (B) asymmetrical adapters, (C) negative adapters. The data is additionally split by the produced syllable. The plot does not contain the data of the three non-adapters. Individual y-axis scales were applied due too big differences across compensatory patterns.

2.3.4 Relation between F2 compensation and F3 changes

To understand the relation between the F2 compensatory magnitude and the corresponding adjustments of F3, we correlated F2 and F3 changes that occurred during the last shift phase of the experiment (Figure 2.6). For trials with upward F2 perturbation, Pearson's correlation coefficients revealed a weak, non-significant positive correlation between the compensation magnitudes observed for F2 and F3 (r = 0.12, p = .51). In contrast, for trials with downward F2 perturbation, we observed a strong, significant positive correlation (r = 0.7, p < .001). These findings mean that most participants compensated consistently for the upward F2 perturbation by decreasing their F2 beyond the corresponding baseline (vertical dashed line in Figure 2.6). In this case, no systematic changes occurred to the corresponding F3 values which appeared to freely fluctuate around their baseline (horizontal dashed line in Figure 2.6). On the other hand, when participants successfully compensated for F2 shifts on trials with downward perturbation by increasing their F2 values beyond the baseline (vertical dashed line in Figure 2.6), it was mostly accompanied by higher F3 values.

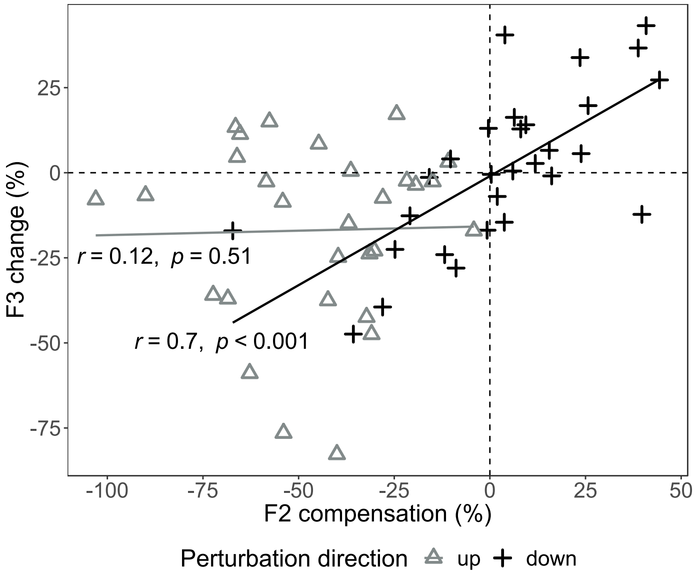

Figure 2.6: Correlation between percentage scores of the corresponding F2 and F3 changes achieved by each speaker during the last shift phase. Correlation is calculated separately for the upward and downward perturbation directions.

In Table 2.3, we summarize the Pearson's correlation coefficients for the relation between F2 and F3 changes independently for the symmetrical, asymmetrical, and negative compensatory patterns. As seen from the table, the three compensatory patterns described in the last section differed systematically with regard to the degree of F2-F3 correlation observed among them. While symmetrical and asymmetrical patterns were characterized by stronger F2-F3 correlation on trials with downward perturbation, the participants exhibiting the negative pattern displayed a weaker relation between F2 and F3 in this case. On the other hand, while asymmetrical and negative patterns were characterized by negative F2-F3 correlation for trials with upward perturbation, the same F2-F3 relation held for the symmetrical adapters as on trials with downward perturbation.

Table 2.3: Pearson's correlation coefficients for the F2-F3 relation calculated separately for each compensatory pattern and applied perturbation direction. None of the correlations was statistically significant likely due to the small number of participants per compensatory pattern and the resulting lack of statistical power.

Compensatory pattern	**F2-F3 correlation**	
	up	*down*
symmetrical	$r = 0.51$	$r = 0.46$
asymmetrical	$r = –0.17$	$r = 0.49$
negative	$r = –0.39$	$r = 0.22$

2.3.5 Average compensatory behavior

As previously mentioned in section 2.2.5 on page 20, the interaction between the perturbation direction (upward vs. downward) and the experimental group (A vs. B) did not significantly improve the fits of the investigated GAMM models. Indeed, the AIC scores were consistently lower for all models which did not contain the interaction between the perturbation direction and the experimental group (–1.2 for the F1 model, –4.74 for the F2 model, and –1.16 for the F3 model). These findings are consistent with the independent analysis of participants' adapted formants presented in section 2.3.2 on page 25 and demonstrates further that the information regarding the coarticulatory configuration speakers were assigned to was irrelevant to model their average compensatory behavior. This also likely means that the compensatory behavior did not significantly differ across the experimental groups A and B. To verify this finding, we visually investigated the model fits which included the interaction (figure not given). Since we did not observe any apparent differences in the compensatory behavior of the two experimental groups, below we present the GAMM models which were fitted without the interaction to focus our later discussion on significant findings.

The GAMM estimated for F1 suggested that the applied perturbation did not have a significant fixed effect on the produced F1 values since they did not significantly differ from the baseline either on trials with upward F2 perturbation (0.63 Hz, *t* = 0.207, *p* = .83) or on trials with downward F2 perturbation (3.63 Hz, *t* = 1.797, *p* = .07). Taking the temporal variation over the course of the experiment into account, the model did not reveal a F1 difference from the baseline for either of the two perturbation directions (Figure 2.7A). Random non-linear smooths of the F1 model suggest that there were unsystematic participant-specific F1 changes which are most likely not related to the applied perturbation (Figure 2.7B). Furthermore, a direct comparison between trials with applied upward and downward perturbation revealed no significant difference in their F1 curves (Figure 2.7C). The average F1 difference amounted to 0.61 Hz (95% CI [–6.52 5.31]) by the end of the first shift phase, –0.73 Hz (95% CI [–7.65 6.18]) by the end of the second shift phase, and –0.86 Hz (95% CI [–10.08 8.37]) by the end of the experiment.

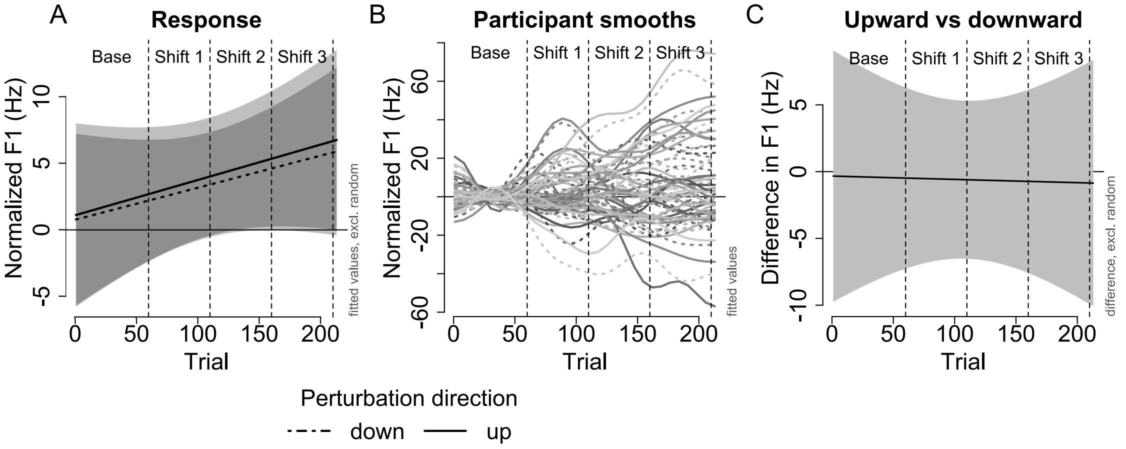

Figure 2.7: (A) Average compensatory effects (excluding random participant effects) in F1 for downward and upward perturbation during the three shift phases. (B) Random model smooths for each participant. (C) The average difference in F1 between the opposing compensatory effects.

The GAMM estimated for F2 suggested that the applied perturbation had a significant fixed effect on the produced F2 values on trials with upward (128.06 Hz, $t = 5.616$, $p < .001$) but not on trials with downward perturbation (12.12 Hz, $t = 0.961$, $p = .33$). However, the direction of the fixed effect was opposed to the direction of the applied perturbation during upward and downward perturbation. Examining the effect of the perturbation over time, the model revealed that the compensation effect increased for both perturbation directions over the course of the experiment (Figure 2.8A). Judging from the figure, the effect appears to be on average stronger for the upward perturbation compared to the downward perturbation. The random F2 smooths fitted individually for each participant indicate that above and beyond the general tendency to counteract the applied perturbation, participants' compensatory adjustments exhibited high variability in both investigated dimensions (formant frequency and time; Figure 2.8B). As indicated by the confidence interval in Figure 2.8C, the F2 difference between trials produced under opposite perturbation directions became significant almost immediately at the beginning of the first shift phase and increased, as expected, over the three perturbation phases. The average F2 difference amounted to 114.82 Hz (95% CI [62.54 167.10]) by the end of the first shift phase and to 182.64 Hz (95% CI [127.78 237.50]) by the end of the second shift phase. By the end of the experiment, the average F2 difference reached 255.74 Hz (95% CI [189.69 321.79]).

The GAMM estimated for F3 suggested that in the course of the experiment participants systematically changed their F3 values even though only F2 perturbation was applied during the experiment. The average fixed effect on the produced F3 values on trials with upward F2 perturbation was –57.61 Hz ($t = -3.459$, $p < .001$) and 7.64 Hz ($t = 0.682$, $p > .49$) on trials with downward F2 perturbation. The direction of the fixed effect was the opposite of the direction of the applied perturbation during upward and downward perturbation. Examining the effect of the perturbation over time, the model revealed that this effect increased for both perturbation directions over the course of the experiment (Figure 2.9A). The random participant smooths suggest that the magnitude of F3 changes varied significantly between participants (Figure 2.9B). Similarly to F2 compensation, it appears that participants' F3 ad-

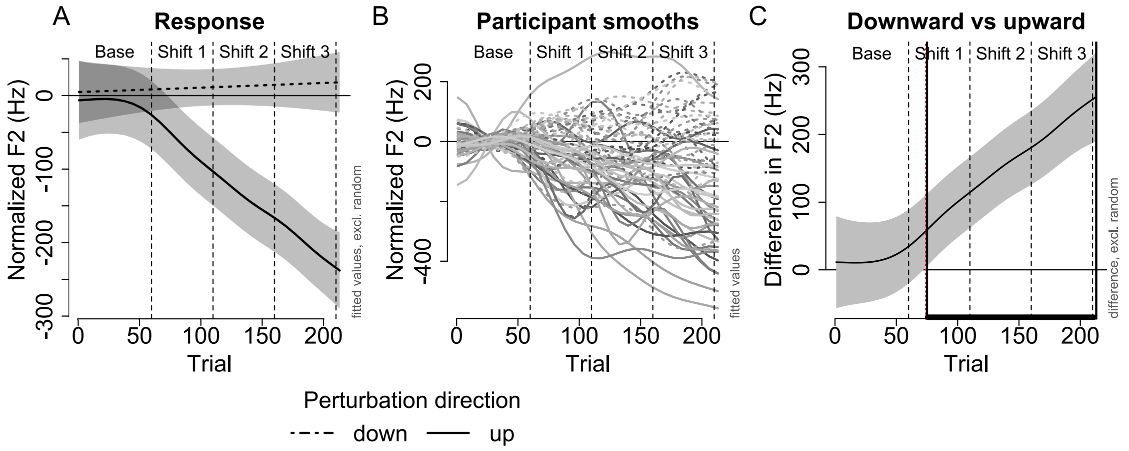

Figure 2.8: (A) Average compensatory effects (excluding random participant effects) in F2 for downward and upward perturbation during the three shift phases. (B) Random model smooths for each participant. (C) The average difference in F2 between the opposing compensatory effects. Solid vertical lines denote the region of significant difference.

justments were on average stronger on trials with upward perturbation. As indicated by the confidence interval in Figure 2.9C, the F3 difference between trials produced under opposite perturbation directions became significant after the first half of the first shift phase and increased over the three perturbation phases. The average F3 difference amounted to 52.56 Hz (95% CI [20.09 82.03]) by the end of the first shift phase and to 80.08 Hz (95% CI [42.01 118.14]) by the end of the second shift phase. By the end of the experiment, the average F3 difference reached 107.66 Hz (95% CI [56.59 158.73]).

2.4 Discussion

From previous perturbation studies it is known that speakers can simultaneously use multiple compensatory strategies to achieve the intended acoustic target (Rochet-Capellan & Ostry, 2011; Feng et al., 2011). However, these results are limited to low vowels /ɛ/ and /æ/ and do not generalize since later research established that the magnitude of compensatory responses to auditory perturbation appears to differ across different vowels due to different degrees of physical contact between the tongue and the hard palate involved during the articulation

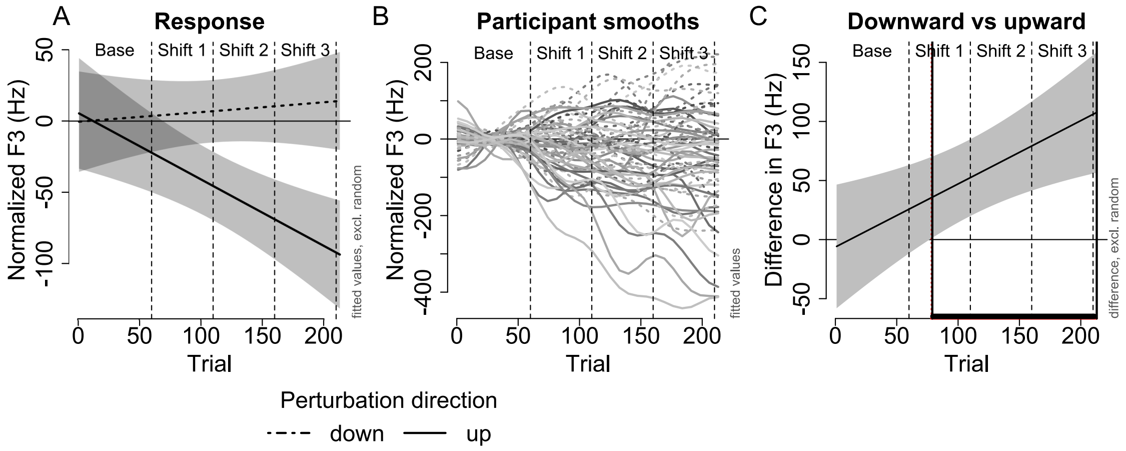

Figure 2.9: (A) Average compensatory effects (excluding random participant effects) in F3 for downward and upward perturbation during the three shift phases. (B) Random model smooths for each participant. (C) The average difference in F3 between the opposing compensatory effects. Solid vertical lines denote the region of significant difference.

of a particular vowel (Mitsuya et al., 2015). The current study investigated the potential influence of large linguopalatal contact on speakers' ability to simultaneously use multiple compensatory strategies.

During three shift phases of the experiment, participants produced the close central unrounded vowel /ɨ/ while its F2 frequency was perturbed with increasing magnitude of 220, 370, and 520 Hz in opposing directions (upward or downward) depending on the preceding consonant (/d/ or /g/). The bidirectional shift was intended to encourage participants to employ two distinct compensatory strategies to produce the vowel in /dɨ/ and /gɨ/. We investigated two experimental groups, where for the first group F2 was decreased in /dɨ/ and increased in /gɨ/, while for the second group, the perturbation directions were swapped for both syllables. This was done to additionally examine the potential influence of the coarticulatory pattern characterizing the relation between syllables /dɨ/ and /gɨ/ on speakers' compensatory production.

Examining participants' average compensatory behavior, we found out that they employed two distinct compensatory strategies depending on the direction of the applied perturbation. This result is consistent with findings made by Rochet-Capellan and Ostry (2011) who demonstrated that participants can employ multiple compensatory strategies for low vowels in the context of F1 perturbation. Adding to this result, our data show, first, that speakers are able to develop multiple compensatory strategies for a vowel with high degree of linguopalatal contact, and second, that speakers adopt multiple compensatory strategies even if these deviate from the coarticulatory relations of their unperturbed speech. These findings demonstrate that auditory feedback serves an important role for online error correction during speech production (see Houde & Jordan, 1998; Purcell & Munhall, 2006; Villacorta et al., 2007), and seem to further indicate that speakers disregard somatosensory errors as long as auditory errors are corrected for (Feng et al., 2011). However, a more detailed analysis of individual compensatory strategies among our participants revealed a fine-grained picture hinting at the influence of additional factors on speakers' individual compensatory responses.

In particular, only about 72 percent of the investigated speakers (23 out of 32) were able to develop two distinct production strategies for the target vowel while the remaining participants failed to do so. Roughly half of the speakers employing two compensatory strategies (10 out of 23) exhibited a symmetrical compensatory pattern, compensating in equal amounts for both applied perturbation directions, and the other half (13 out of 23) exhibited an asymmetrical compensatory pattern, compensating more strongly for the upward perturbation and mostly ignoring the downward perturbation.

At first glance, a hypothesis which could explain the emergence of the asymmetric compensatory pattern might entail the idea that in this case speakers' compensatory movements were bound by different physical restrictions associated with the two experimental syllables /dɨ/ and /gɨ/. In the case of /dɨ/, for instance, the magnitude of the forward movement of the tongue body, which was required to compensate for downward shifts, was most likely restricted by the alveolar ridge and the upper incisors. On the other hand, the compensation movement in the case of the upshifted /gɨ/ was directed towards the pharynx allowing the tongue to travel a farther distance along the palate. However, as straightforward this hypothesis appears to be, it can account only for a part of the current data since the asymmetrical compensatory pattern also occurred when /gɨ/ was perturbed downwards although there should be no physical restrictions which were comparable to the case of downward perturbation of /dɨ/. In other words, the symmetrical compensatory pattern was observed among speakers independently of whether they were assigned to the experimental group A or B (compatible or incompatible coarticulatory configuration). This observation suggests that the physical restrictions associated with a specific place of articulation of the two syllables /dɨ/ and /gɨ/ (alveolar vs. velar) probably did not have a crucial effect on the emergence of symmetrical and asymmetrical compensatory patterns.

Among the 28 percent of speakers (9 out of 32) who failed to develop two distinct production strategies for the target vowel /ɨ/, six participants exhibited a negative compensatory pattern significantly decreasing their F2 frequency irrespective of the perturbation direction, and the remaining three participants failed to compensate consistently for any of the two perturbation directions. However, their production of the vowel /ɨ/ also underwent significant changes in the course of the experiment.

There remains a possibility that changes present in speakers with inconsistent compensatory responses were caused by a formant drift, for instance, due to fatigue. Presence of such formant drift, however, presupposes that a speaker does not actively adjust her/his formants and therefore the observed changes are expected to have rather low effect sizes across all formants (F1-F3), both produced syllables (/dɨ/ and /gɨ/), and remain independent of the perturbation direction (upward and downward). Contrary to these assumptions, formant changes observed for the three inconsistent adapters were rather heterogeneous with respect to the effect size and the affected formant frequency.

In particular, while there were no significant F1 changes in their speech, all three speakers adjusted their F2 for at least one perturbation direction by approximately 150 to 250 Hz, sometimes following the direction of the perturbation. Consequently, we incline to believe that these speakers perceived the induced auditory errors but were either not able to classify the directions of the applied F2 shifts, and identify the articulatory parameters they needed to adjust in order to restore the intended acoustic goal, or have failed to successfully coordinate the required articulatory adjustments in a manner required for each perturbation direction.

As with symmetrical and asymmetrical compensatory patterns, the negative pattern and inconsistent compensatory responses occurred equally among participants assigned to both experimental groups (compatible and incompatible coarticulatory configuration). This further supports the assumption that physical restrictions associated with a particular syllable in combination with a particular perturbation direction appear to have played but a minor role in the emergence of individual compensatory patterns. This suggests that other factors were responsible for shaping of speakers' individual compensatory responses.

Looking at the individual data ignoring different compensatory patterns, we see that while 90 percent of participants (29 out of 32) compensated for the upward perturbation, only about 31 percent of participants (10 out of 32) compensated for the downward perturbation. This compensatory asymmetry seems to be congruent with the asymmetry present in the phonemic space of Russian high vowels.

As described in the introductory section, in Russian, /i/ appears only after palatalized consonants and both /ɨ/ and /u/ follow only non-palatalized ones (see Bolla, 1981, pp. 108-110). Since the dominant perceptual cue of palatalization for Russian speakers is the height of F2 frequency at the beginning of the following vowel, it seems reasonable that most participants classified instances of /ɨ/ with increased F2 as phonemic errors of palatalization while, on the other hand, less speakers reacted to decreasing F2 as it did not induce a change of the phonemic status of the perceived vowel. Additionally, this perceptual effect may have been strengthened by the fact that the F2 frequency peak overlapped spatially with the F3 peak during upward perturbation making the shifts auditorily more salient compared to downward shifts of F2.

Averaging the individual data across all speakers, we arrive at results which appear to be consistent with findings made by Niziolek and Guenther (2013) who showed that speakers react on average much more strongly to perturbations which result in changes of the phonemic category of the perturbed vowel compared to perturbations which result only in sub-phonemic changes. Speaking in specific terms, in our case the compensatory magnitude amounted on average to 45 percent on trials with upward and to 3 percent on trials with downward perturbation during the last shift phase of the experiment. However, this is a simplification of the actual compensatory behavior since we know that some speakers (symmetrical adapters) compensated equally for upward and downward perturbation, irrespective of whether it alternated the phonemic category of the perturbed vowel, and other speakers (asymmetrical adapters) reacted essentially only to upward perturbation which alternated the phonemic category of the perturbed vowel.

Considering these detailed observations, it appears that while for symmetrical adapters the perceptual processes involved during compensation related more to vowel discrimination, asymmetrical adapters relied more strongly on vowel identification (see discussion in Reilly & Dougherty, 2013). In agreement with findings made by Villacorta et al. (2007) about the influence of auditory acuity on the compensatory magnitude, we think that symmetrical adapters were presumably more sensitive to F2 changes independent of the phonemic status of the perceived vowel.

Another hypothesis which could potentially explain the compensatory asymmetry is that although it was feasible to compensate for the upward F2 perturbation in /ɨ/ by lowering exclusively the F2 frequency, a compensation for the downward F2 perturbation required from speakers to raise their F2 along with F3 since both frequencies lie quite close in the target vowel /ɨ/. Although speakers should be able to raise F2 and F3 simultaneously by changing one articulatory parameter, such as the horizontal tongue body position, the adjustment of the F3 frequency might be easier to achieve involving additional articulatory changes such as the degree of lip opening or spreading. From previous literature on Russian vowels, it is known that /ɨ/ is normally produced with a slightly wider lip opening compared to /i/ which has much higher F3 values (see Bolla, 1981, pp. 109-110). In this scenario, if speakers narrowed their lips additionally to the forward movement of the tongue, that could raise their F3 values and contribute to a stronger F2 compensation. Some support for this hypothesis is provided by the results of a correlational analysis of F2 and F3 changes for the two opposite perturbation directions which demonstrated that on the group level this correlation was significant and highly positive on trials with downward perturbation but not on trials with upward perturbation.

An additional correlational analysis performed separately for speakers exhibiting different compensatory patterns revealed furthermore that the positive F2-F3 correlation held for symmetrical adapters not only on trials with downward, but also on trials with upward perturbation. This suggests that these speakers developed a compensatory strategy which encompassed F2 and F3 changes for both perturbation directions. On the other hand, asymmetrical adapters developed a compensatory strategy encompassing only the F2 frequency, which was sufficient to compensate for the upward perturbation but not for the downward perturbation.

The hypothesis that speakers employ individual compensatory strategies involving different degrees of articulatory complexity may also provide some explanation for the occurrence of the negative compensatory pattern. Judging from the correlational analysis, negative adapters developed a compensatory strategy for the upward perturbation which was similar to that of the asymmetrical adapters, but used it for both perturbation directions. Although the exact mechanisms behind this remain to be seen, it is possible that negative adapters tried to correct for the perceptually more salient auditory errors, which occurred during the upward perturbation and, at the same time, adhered to some kind of articulatory economy by employing the same compensatory strategy also during the downward perturbation.

The two explanations for the emergence of different compensatory patterns provided from perspectives of speech perception and articulation are not mutually exclusive. On the contrary, following the idea that representations of speech sounds are defined in a multidimensional auditory-articulatory space, it appears plausible that both dimensions could and should have an influence on speakers' ability to compensate for perturbations. Indeed, the presence of different compensatory patterns in our data hints at the idea that speakers might differently weight the information provided by different feedback channels (see discussion in Lametti et al., 2012).

In summary, our analyses suggest that although speakers are able to use multiple compensatory strategies even for vowels with strong linguopalatal contact, there is an array of auditory and, probably, articulatory factors that have an additional influence on speakers' individual compensatory performance. While some of the previous work has pointed out the importance of several factors like phonemic relations and non-redundant perceptual cues of the perturbed vowel, the current study provides evidence for the advantage of analyzing individual participants' performances since this may provide deeper insights into the mechanisms of feedback control and the nature of speech targets.

Chapter 3

The relevance of auditory feedback for production of fricatives

3.1 Introduction

The goal of the current chapter is to systematically evaluate the relevance of auditory feedback during fricative production taking into account methodological shortcomings of the previous auditory perturbation studies of fricatives by Shiller et al. (2009) and Casserly (2011).[1] To do so, let us first review both studies in more detail.

In Shiller et al.'s study, participants produced CV and CVC words starting with /s/ sounds. Their responses were pitch shifted in real-time around –3 semitones and fed back via headphones to the participants. The pitch shifting reduced the spectral center of gravity (COG) of the fricative on average by 1430 Hz. The authors found that the participants raised their COG on average by 529 Hz in reaction to the shift. However, in this case it is not possible to attribute observed compensation effects solely to the fricative perturbation since the whole response including the following vowel and consonant segments was affected by the pitch shift. That means that the observed compensatory effect might be a consequence of a general f0 compensation, an effect that was frequently demonstrated in the literature before (e.g., Jones & Munhall, 2000).

By perturbing specifically the fricative segments, Casserly (2011) improved on Shiller et al.'s holistic perturbation approach. During her experiment, participants produced VCV pseu-

[1] The results of this chapter were previously published in Klein, Brunner, and Hoole (2019b).

dowords containing the sound /ʃ/ while they received auditory feedback over headphones. Simultaneously to participants' production of the fricative, synthetic turbulent noise was mixed into participants' feedback raising the COG by about 700 Hz. Contrary to Shiller et al., Casserly identified not one, but three different behavioral patterns among her participants: i) no reaction to the perturbation, ii) raising of the COG ('imitating the perturbation'), and iii) lowering of the COG ('counteracting the perturbation'). Despite methodological improvements, Casserly's perturbation solution has its own potential problems such as asynchronous starts of the produced fricative and the synthetic noise, and duration and amplitude mismatches between the two. These problems can result in unnatural sounding stimuli and undesired acoustic artifacts which might have an influence on participants' behavior. Furthermore, it is difficult to evaluate Casserly's interpretation regarding imitating and counteracting responses since due to the unidirectional perturbation design even unrelated drifts in speakers' COG, e.g., due to general fatigue, remain indistinguishable from systematic behavior.

With respect to the perturbation method, the current investigation expands on the studies by Shiller et al. (2009) and Casserly (2011) addressing some of their key methodological issues. In contrast to Shiller et al. (2009), we apply perturbation only to the final fricative segment of CVC words leaving the two preceding segments unaffected by the manipulation. By doing this, we seek to avoid potential influence of overall f0 changes on participants' compensatory behavior. In contrast to Casserly (2011) study, the target segments are perturbed in real-time eliminating such potential acoustic artefacts as amplitude and duration mismatches between participants' speech and the feedback signal.

Given the facts that the compensation in fricatives involves a diverse set of articulatory parameters and that COG might be insufficient to reveal speakers' compensatory adjustments (see discussion in Chapter 1), it appears to us necessary to investigate a broader range of acoustic variables associated with fricative production in order to successfully evaluate the outcome of our experiment. Consequently, we chose to examine a series of measures drawing upon previous work on acoustic correlates of sibilant fricatives in general (see overview in Koenig et al., 2013) and in Russian – the target language of our investigation – specifically (Kochetov, 2017).

We conducted a bidirectional auditory perturbation study of the sibilant fricative /s^j/ in which the spectrum of the investigated sound was perturbed in opposite directions depending on the experimental stimulus it was embedded in such that the spectral balance was shifted towards higher or lower frequencies on each experimental trial (leading to higher or lower

COG, respectively). In order to compensate for the bidirectional shifts, participants had to coordinate their compensatory movements in two different ways to produce the target sound /s^j/. By employing this design, we made sure that acoustic changes in speakers' production that occurred irrespectively of the applied perturbation direction would not be incorrectly interpreted as compensation or imitation. Furthermore, we included a noise-masked speaking condition to investigate whether participants would retain their compensatory adjustments in production of /s^j/ when they no longer could rely on the auditory feedback, and whether they would immediately return to compensating as soon as they were again provided with the perturbed feedback.

During the data analysis, we sought to answer two related questions by applying a supervised classification algorithm (Random Forest (RF); Breiman, 2001). First, we aimed to identify those acoustic parameters which were systematically affected during the adaptation process, thus trying to eliminate parameters changing throughout the experimental session independently of the applied perturbation direction. Secondly, we investigated the evolution of the adaptation process by modeling the algorithm's predictions over several time intervals of the experiment based on the values of acoustic parameters deemed relevant to classify average speakers' adapting behavior. Finally, to break down the adaptation process into individual compensatory strategies, we traced specific acoustic changes that occurred over the course of the experiment by means of GAMMs.

3.2 Methods

3.2.1 Participants

Twenty-three native speakers of Russian (16 females, 7 males) without reported speech, language, or hearing disorders participated in the study. The participants were recruited from the pool of Russian exchange students and young professionals living in Berlin. The mean age of the group was 24.6 years. Participants had spent on average 2.7 years in Germany prior to the recordings. The study was approved by the local ethics committee and all speakers gave their written consent to participate in the study.

3.2.2 Equipment

The overall experimental set-up was identical to the one described in section 2.2.2 on page 16. During the experiment, segment tracking and real-time perturbation was accomplished with AUDAPTER. In order to perturb fricative spectra, we employed AUDAPTER's pitch shifting facilities. When applied to voiceless fricative spectra, this manipulation results in an overall spectral shift such that the COG of the fricative is increased or decreased depending on the direction of the pitch shift. To identify the onset and the offset of the fricative in real-time, AUDAPTER performed an analysis of the speech signal's short-time root-mean-square (RMS). While RMS amplitude is a general intensity measure, we defined an RMS ratio as an indicator of high frequency intensity present in the signal by dividing the high-pass filtered RMS curve by a smoothed RMS curve (Figure 3.1A). When the RMS ratio curve crossed a threshold of 0.03 and kept on rising for the following 15 ms, suggesting the onset of the fricative, pitch shifting was activated. Subsequently, AUDAPTER deactivated pitch shifting when the RMS ratio curve fell below a threshold of 0.06 and kept on falling for the following 15 ms (Figure 3.1B). The chosen thresholds were defined during pilot recordings of a native speaker's /s^j/-productions. To maximize the tracking success rate, the experimental stimuli were constructed in such a way that on each trial there was a unique time interval characterized by high spectral energy (see next section).

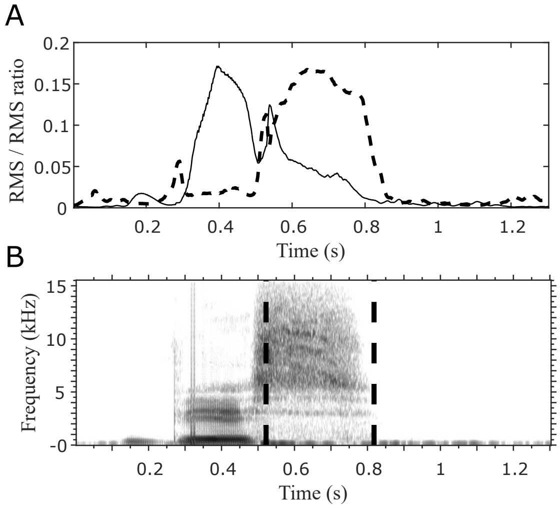

Figure 3.1: Example of a single experimental trial: (A) RMS (solid line) and RMS ratio curve (dashed line) of the speech signal. (B) Fricative onset and offset (dashed lines) tracked by AUDAPTER overlaid on a spectrogram of the speech signal.

Direct measurements were made to determine the delay of the auditory feedback loop by comparing the onsets of the incoming acoustic signals from the input and the output channel of the perturbation system. The delay was about 15 ms when no perturbation was applied and increased up to 22 ms on trials when pitch shifting was activated. The original and perturbed signals were digitized and saved with a sampling rate of 32 kHz.[2]

3.2.3 Speech stimuli and experimental manipulation

For our study we chose Russian since its consonant inventory includes a series of voiceless sibilants /s/, /sʲ/, and /ʃʲ/ which are produced with quite small articulatory differences resulting in frequency spectra of qualitatively similar shapes.[3] Both /s/ and /sʲ/ are alveolar consonants which are produced by a narrowing which is formed in the center, front part of the oral cavity between the tongue tip and the region of the upper incisors and the alveolar ridge. In contrast to /s/, /sʲ/ is palatalized which means that the tongue body moves further forward during this sound. Finally, /ʃʲ/ is a palatalized post-alveolar consonant which is produced by a narrowing of the tongue dorsum which forms a gap along the medial line of the palate. For a more thorough overview of the articulatory characteristics of Russian voiceless sibilants see works by Bolla (1981, pp. 87-88, 90-92) and Skalozub (1963, pp. 28-35).

All three investigated sounds are acoustically characterized by wide and continuous frequency spectra which are mostly differentiated by prominent noise peaks which occur around 5000-6000 Hz for /s/, 4750-6000 Hz for /sʲ/, and 3300-3700 Hz for /ʃʲ/ (Bolla, 1981); similar average values are also reported in Padgett and Żygis (2007) (see also Figure 3.2). The acoustic proximity between the sounds /s/, /sʲ/, and /ʃʲ/ allowed us to perform delicate auditory perturbations of the target sound /sʲ/ which made it sound either more similar to /s/ or to /ʃʲ/.

During a baseline phase, before any spectral perturbation was applied, participants had to produce six CVC words each containing a final voiceless fricative to assess their usual speech production of these sounds (Table 3.1).[4]

During three shift phases, participants produced words containing the palatalized fricative

[2] For this we had to modify the original AUDAPTER software to operate at a higher sampling frequency.

[3] Russian also includes the voiceless sibilant /ʃ/ which is not investigated in the current study.

[4] The relative word frequencies for the six stimuli words as suggested by the Russian National Corpus (http://www.ruscorpora.ru/en/index.html) are 6.52e–05 for [les], 2.37e–06 for [lesʲ], 5.55e–07 for [leʃʲ], 2.31e–05 for [ves], 3.08e–04 for [vesʲ], and 5.14e–05 for [veʃʲ].

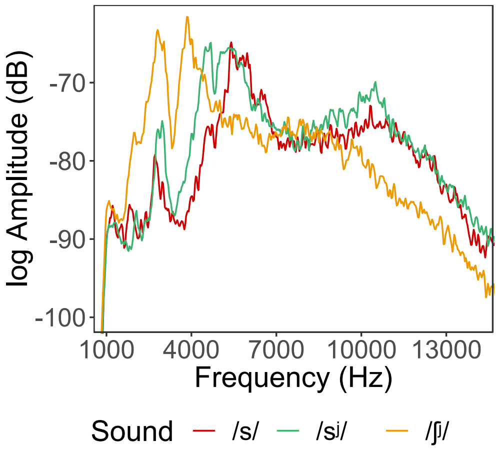

Figure 3.2: Example power spectra of the investigated sounds /s/, /sʲ/, and /ʃʲ/.

Table 3.1: The experimental stimuli.

[les]	’wood’	[lesʲ]	’climb’	[leʃʲ]	’bream’
[ves]	’weight’	[vesʲ]	’whole’	[veʃʲ]	’thing’

/sʲ/ embedded in the words [lesʲ] and [vesʲ]. During the production of the target sound /sʲ/, pitch was decreased by five semitones in the word [vesʲ] and increased by five semitones in the word [lesʲ]. Note that the pitch shift was applied exclusively to the fricative portion of each stimulus word and did not affect the first two segments. As can be seen in Figure 3.3, across both perturbation directions, the applied pitch shift altered the overall spectral balance of the fricative /sʲ/ affecting the amplitudes of the lower (2.5-5.5 kHz) and higher (10-12 kHz) frequency bands in a complementary fashion.

Specifically, when pitch was shifted downwards, the amplitude of the lower frequency band increased while the amplitude of the higher frequency band decreased such that the frication noise of the target sound /sʲ/ resembled that of a more low-frequency fricative. The upward pitch shift caused analogous acoustic changes but in the opposing spectral direction such that the target sound /sʲ/ resembled a high-frequency fricative. Although the amplitude in the middle frequency band (5.5-10 kHz) was also prone to some minor changes, we think that these modifications were perceptually less salient compared to complementary perturbations that occurred in lower and higher bands (see Table 3.2).

The shift effects were also observable by means of the first (COG) and second (standard deviation; SD) spectral moment. As expected, the downward pitch shift caused a significant

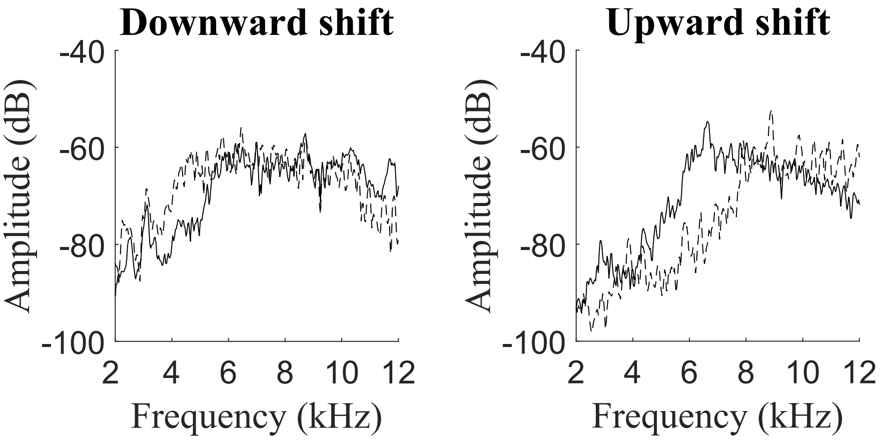

Figure 3.3: Example power spectra of the original (spoken by a participant; solid lines) and perturbed (heard by a participant; dashed lines) fricative segments during the shift phases.

decrease in COG and SD, while the opposite was true for the upward shift. The magnitude of the applied perturbation remained constant throughout the three shift phases and resulted in average shift effects summarized in Table 3.2 based on the data of the first shift phase. Generalized additive mixed modeling of the speech signals perceived by participants throughout the experiment confirmed that the magnitude of shift effects was consistent and equivalent across all speakers (see Figure 3.4).

3.2.4 Experimental procedure

Each recording session lasted for approximately 25-30 minutes and consisted of seven experimental phases (Table 3.3). Before the experiment began, participants completed few practice trials with unrelated speech material to adjust the microphone gain enabling reliable fricative tracking.

During the baseline phase, no auditory perturbation was applied and participants were able to familiarize themselves with the experimental situation of receiving auditory feedback over earphones. On each trial, which had an approximate duration of 2 seconds, participants were visually prompted to produce one of the six CVC words [les], [ves], [les^j], [ves^j], [leʃ^j], or [veʃ^j]. The interstimulus interval between the trials was approximately 1.5 seconds long.

Participants were asked to produce each CVC word with a neutral (flat) intonation, in a normal speech tempo to improve online tracking of the fricative. To keep the speech amplitude equal across all experimental phases, participants were provided with a real-time

Table 3.2: Mean differences between produced and perceived frication noise with respect to the amplitude of the low (Level$_{\mathrm{Low}}$), mid (Level$_{\mathrm{Mid}}$), and high (Level$_{\mathrm{High}}$) frequency band as well as the first two spectral moments (COG and SD). Standard deviation is given in parentheses. The data are split by gender (f = female; m = male). For more details on the role of the reported acoustic parameters for fricative production see section 3.2.6.

Parameter	**downward shift** ([ves^j])		**upward shift** ([ves^j])	
	f (n = 14)	*m (n = 5)*	*f (n = 14)*	*m (n = 5)*
Level$_{\mathrm{Low}}$ (dB)	6.86 (3.94)	1.65 (1.2)	–5.57 (3.17)	–5.27 (0.52)
Level$_{\mathrm{Mid}}$ (dB)	–2.56 (1.82)	–1.73 (1.48)	–0.44 (1.25)	0.82 (2.35)
Level$_{\mathrm{High}}$ (dB)	–17.6 (3.51)	–17.84 (1.15)	8.28 (1.85)	8.35 (3.32)
COG (Hz)	–1863 (208)	–1750 (334)	2431 (233)	2161 (273)
SD (Hz)	–492 (89)	–579 (82)	452 (120)	575 (34)

Table 3.3: The experimental sequence.

	Experimental phase							
	Baseline	Shift 1	Noise 1	Shift 2	Noise 2	Shift 3	Noise 3	Total
Trials	48	30	24	30	24	30	24	210

graphical display of the microphone gain and asked to keep a steady intensity range. They were also assisted by a visual go-and-stop signal during their production.

After the baseline phase, participants completed three cycles of alternating shift and noise phases. During the shift phases, participants were prompted to produce the voiceless palatalized alveolar fricative /s^j/ embedded in one of the CVC words [les^j] or [ves^j]. The whole spectrum was shifted during the production of the fricative by five semitones either downwards for the word [ves^j] or upwards for the word [les^j]. Within each shift phase the perturbed words were presented in pseudorandomized order. The first shift phase began with a ramp phase which lasted for 10 trials and served for gradually increasing the magnitude of the applied perturbation. The measurements collected during the ramp phase (10 trials for each participant) were not included in the analysis. The second and the third shift phases did not contain a ramp phase.

During noise phases, participants were prompted to produce all six CVC words [les], [ves], [les^j], [ves^j], [leʃ^j], and [veʃ^j] while perceiving pink noise (range 0-20 kHz) which was played over earphones at approximately 75 dB SPL. Considering that the air-conducted sound was attenuated by 25-30 dB by the earphones and that participants were required to main-

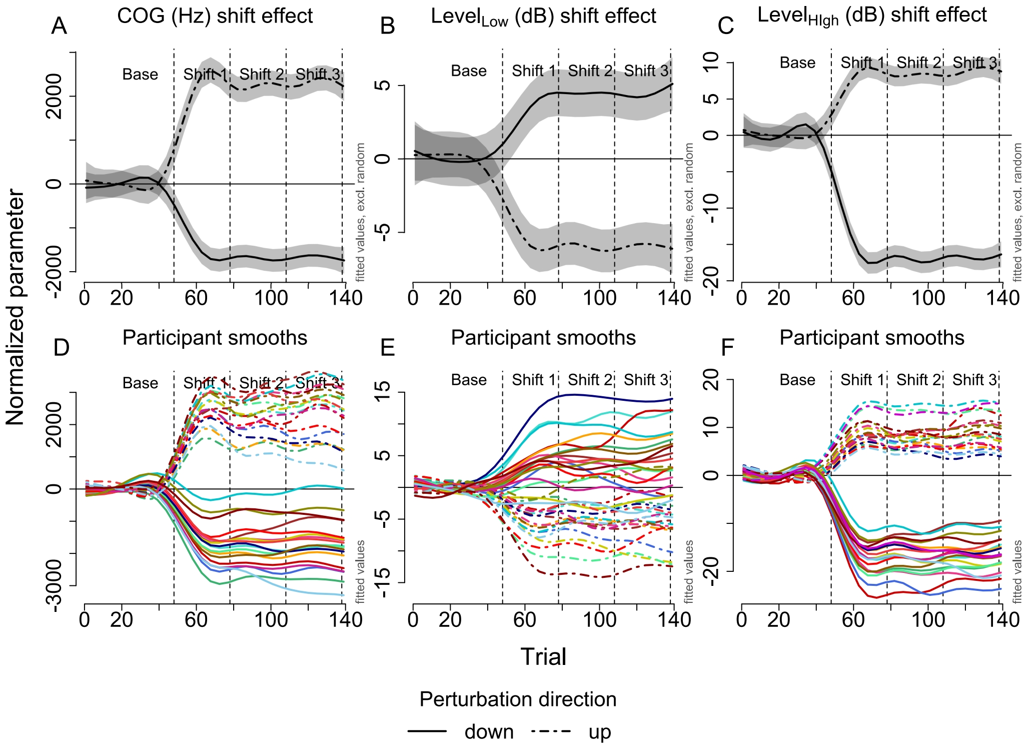

Figure 3.4: Visual summary of the GAMM models fitted for perceived COG as well as $\text{Level}_{\text{Low}}$ and $\text{Level}_{\text{High}}$ amplitudes. (A-C) Average shift magnitude for downward and upward perturbation plotted across baseline and shift phases of the experiment. Grey bands represent 95% confidence intervals. (D-F) Random smooths of the corresponding models color-coded for single participants.

tain their usual speaking intensity, the employed noise level effectively masked speakers' responses as confirmed by self-experimentation. The noise phases were included to investigate whether participants would retain their adapted pronunciation under masked feedback.

Following the experimental session, all participants were asked if they noticed anything unusual in their auditory feedback during the experiment. A few of the participants stated that their pronunciation of the fricative was different in some words from what they are used to. Most participants attributed these pronunciation differences to technical reasons so when asked if and how these differences affected their production, participants reported to have mostly ignored them.

3.2.5 Data pre-processing

All recordings of 23 participants amounted to 4830 trials. The onsets and offsets of the final fricative segments and the vowels preceding them were labeled for each trial manually using MATLAB's graphical input facilities. Subsequently, the acoustic measures were extracted from the vowel and fricative portions of each response based on the labeled onset and offset boundaries.

For each vowel, average formant values (F1 and F2) were obtained over the final 25 percent of the vowel duration. To this end, the recorded signal was first resampled to 10 kHz and pre-emphasized by approximately 6 dB per octave. Then, a 14-pole or a 10-pole LPC-analysis was performed to extract formants for male or female speakers, respectively. For each fricative, the speech signal was first high pass filtered at 600 Hz to exclude any potential influence of intrusive voicing. Then, power spectral densities (PSD) were computed with MATLAB's pwelch() function with a default window size over the middle 50 percent of the fricative. Finally, the spectral measures were computed for the time averaged power spectra.

To determine whether AUDAPTER was able to track the fricative, and therefore to successfully deliver perturbation, each trial was visually inspected and all trials where the interval for pitch shifting did not correspond with the speaker's /s^j/-production were discarded from the analysis. Hence, the data of four participants were excluded completely from further analysis since the application of perturbation failed in more than 30 percent of their shift trials. This left the data of 19 speakers for the analysis. Then, from the remaining data, a further 5.9 percent of all shift trials were discarded due to erroneous fricative tracking.

3.2.6 Experimental measures

In order to get a comprehensive analysis of participants' compensatory behavior, we investigated several acoustic measures of their responses produced under perturbed auditory feedback. To be able to compare our findings to the results of previous auditory perturbation experiments investigating sibilant fricatives, we computed the first three spectral moments including the COG along with the spectral SD and skewness (Forrest et al., 1988). Previous studies on spectral moments reported that these measures can be used to effectively discriminate different sibilants. Specifically, higher COG and more negative skewness scores (more energy in high frequencies) are reported for /s/ compared to /ʃ/. Furthermore, higher

SD scores are reported for sibilants with broader noise spectra (e.g., Jongman, Wayland, & Wong, 2000).

Following the insights summarized in Koenig et al. (2013) who evaluated a range of spectral parameters for the investigation of sibilants, we extracted additional measures from different bands of the fricative spectrum. To this end, we divided the spectrum into three bands: 600-5500 Hz (low band), 5500-11000 Hz (mid band), and 11000-16000 Hz (high band) (Figure 3.5). The specific boundaries for each frequency band were defined based on explorative analysis of the /s^j/ tokens produced by all speakers during current experiment.

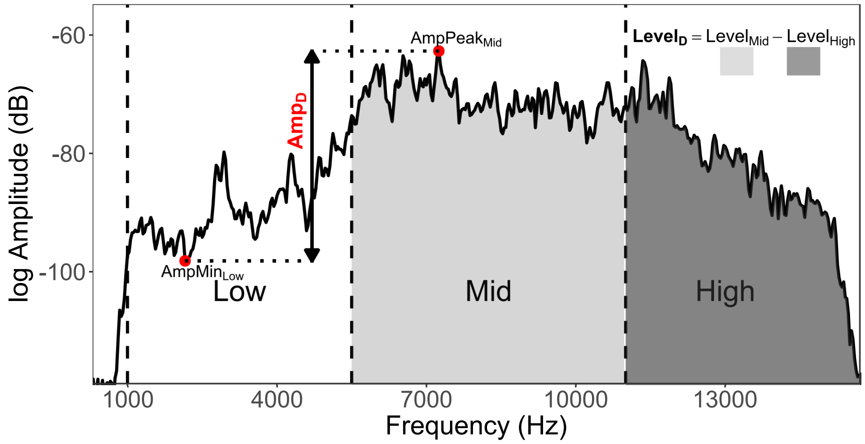

Figure 3.5: Example of a spectrum of the fricative /s^j/ divided into three spectral bands (low, mid, and high) to sequentially compute the investigated spectral parameters. Figure adapted from Koenig et al. (2013).

From the mid band, we extracted the corresponding frequency peak (Freq_{Mid}) which is expected to be lower when the front cavity is longer and/or the lip opening becomes smaller during the production of the sibilant fricative (Iskarous, Shadle, & Proctor, 2011; Jongman et al., 2000). Furthermore, we computed the difference between the amplitudes at the frequency peak over the mid band and at the minimum frequency over the low band ($\text{AmpD}_{\text{Mid-MinLow}} = \text{AmpPeak}_{\text{Mid}} - \text{AmpMin}_{\text{Low}}$). This difference quantifies a balance of acoustic energy in the low and the mid frequency bands and serves as a measure of the 'goodness' of the produced sibilant noise (Koenig et al., 2013). The difference $\text{AmpD}_{\text{Mid-MinLow}}$ is systematically lower for fricatives in which the frication noise is produced by the air stream hitting an obstacle (i.e., upper incisors in the case of /s/) downstream from the constriction location (Shadle, 1990). A similar logic is true for the difference between the energy levels of the high band and the mid band ($\text{LevelD}_{\text{High-Mid}} = \text{Level}_{\text{High}}-\text{Level}_{\text{Mid}}$) which was computed as a second relative measure. Both relations, $\text{AmpD}_{\text{Mid-MinLow}}$ and $\text{LevelD}_{\text{High-Mid}}$, are conceptually related to the spectral slope measures computed in Evers, Reetz, and Lahiri (1998), Jesus and Shadle (2002), and Jones and Munhall (2003) for different regions of the fricative spectrum and provide essentially the same information.

In addition to the spectral moments and the three measures described in the above paragraph, we extracted several acoustic parameters including amplitude peaks, RMS levels, and their relations computed over different frequency bands (Table 3.4). The latter parameters were extracted to provide additional acoustic variables with potential predictive power for the classification algorithm which was used to identify the acoustic parameters relevant for the adaptation process.

Table 3.4: Definition of the fricative-internal acoustic parameters.

Parameter	**Definition**
COG	Center of gravity of the whole spectrum
SD	Standard deviation of the whole spectrum
Skewness	Skewness of the whole spectrum
$\text{AmpMin}_{\text{Low}}$*	Minimum amplitude over the low-frequency band (600-5500 Hz)
$\text{AmpPeak}_{\text{Low}}$*	Peak amplitude within the low-frequency band (600-5500 Hz)
$\text{AmpPeak}_{\text{Mid}}$*	Peak amplitude within the mid-frequency band (5500-11000 Hz)
Freq_{Mid}	Frequency at $\text{AmpPeak}_{\text{Mid}}$
$\text{AmpPeak}_{\text{High}}$*	Peak amplitude within the high-frequency band (11000-16000 Hz)
$\text{AmpD}_{\text{Mid-MinLow}}$	$\text{AmpPeak}_{\text{Mid}}-\text{AmpMin}_{\text{Low}}$
$\text{AmpD}_{\text{High-Mid}}$	$\text{AmpPeak}_{\text{High}}-\text{AmpPeak}_{\text{Mid}}$
$\text{Level}_{\text{Low}}$	Sound level (RMS) over the low-frequency band (600-5500 Hz)
$\text{Level}_{\text{Mid}}$	Sound level (RMS) over the mid-frequency band (5500-11000 Hz)
$\text{Level}_{\text{High}}$	Sound level (RMS) over the high-frequency band (11000-16000 Hz)
$\text{LevelD}_{\text{High-Mid}}$	$\text{Level}_{\text{High}}-\text{Level}_{\text{Mid}}$
$\text{LevelD}_{\text{Mid-Low}}$	$\text{Level}_{\text{Mid}}-\text{Level}_{\text{Low}}$

* These measures were not directly used to classify the fricative spectra but served as intermediate variables to compute relative measures.

Besides fricative-internal measures, we investigated the formants (F1 and F2) of the vowel /e/ preceding each fricative since it was previously shown for Russian that F1 and F2 indicate the degree of palatalization (amount of tongue advancement) of the following fricative (Kochetov, 2017). Specifically, palatalized sibilant fricatives, including /s^j/, are associated with considerably higher F2 and somewhat lower F1 values at the end of a preceding vowel. These measures were included in the analysis since speakers may adjust the formants of the preceding vowel to compensate for the fricative perturbation.

3.2.7 Statistical analyses

All analyses were performed in R (version 3.4.1; R Core Team, 2017). To identify acoustic parameters potentially involved in the adaptation process, we applied the RF algorithm to responses produced by participants under varying experimental conditions (i.e., unperturbed feedback, perturbed feedback, noise-masked feedback) using the implementation provided in the *randomForest* package (version 4.6-14) by Liaw and Wiener (2002).

RF is a supervised ensemble classification technique with the objective to predict the values of one particular response variable, in our case, the sound category (/s/ vs. /s^j/ vs. /ʃ^j/) or the perturbation direction (downward vs. upward), from a set of predictive variables, in our case the previously defined acoustic measures (Table 3.4). RF classification is performed by aggregating over the predictions of multiple binary decision trees (usually several hundreds or thousands), whereas each decision tree is a prediction model with a branching structure and, simply speaking, the goal to split the given data into subsets which are as homogeneous as possible with respect to the values of the response variable (for an introduction to decision trees see Breiman, Friedman, Olshen, & Stone, 1984). In an RF, decision trees are developed on random bootstrap samples of the original data and a random subset of the predictive variables is used at each data split to minimize the correlation between each tree and thus potential bias towards certain predictive variables. One of the questions which we asked ourselves in the context of our data was whether an RF model could be found which would classify participants' responses reasonably well into those produced under downward vs. upward perturbation based solely on the acoustic measures we provided it with.

In addition to classification, RF models provide an importance measure for each predictive variable defined as the loss of accuracy of classification which occurs when the values of the given variable are randomly permuted. First, the loss of accuracy for a particular predictive variable is computed separately for all trees which use this variable for classification, and then its mean and standard deviation are computed across the whole forest. However, since the RF computations involve stochastic processes the importance measure of each predictive variable fluctuates in different runs of the model. This circumstance may complicate the identification of predictive variables which are relevant for the optimal classification, and thus to the understanding of the underlying adaptation process.

To overcome this practical concern of RF usage, we applied a feature selection procedure as implemented in the *Boruta* package (version 5.3.0) by Kursa and Rudnicki (2010). The main idea of this procedure is to compare the importance of actual predictive variables with the importance of their corresponding "shadow" variables, which are their copies consisting of randomly permuted values. The random permutation applied to "shadow" variables destroys any possible relation with the response variable which means that any corresponding importance score bigger than zero is caused by noise. After computing an importance measure for every "shadow" variable, the maximal one of these values is selected as a reference value against which the importance measures of actual predictive variables are compared in a two-sided t test to decide if a particular variable is relevant in predicting the value of the response variable. To get valid results, the whole Boruta procedure is performed until an importance decision is made with respect to all predictive variables. The empirical power and the stability of the results provided by the Boruta algorithm were evaluated with several simulated and actual data sets by Degenhardt, Seifert, and Szymczak (2017).

When performing RF modeling throughout the current investigation, we adhered to a certain sequence of computational steps and a fixed set of options. First, we identified the predictive variables deemed relevant for a given classification task by means of the Boruta procedure. The decision whether a parameter was relevant for a given classification task was made at the significance level 0.01. Then, we computed an RF model with 5000 trees including only the predictive variables deemed previously relevant. Each RF model was set up to use all of the chosen predictive variables to perform a split on the data.

First, we explored participants' baseline production of the target sound /s^{j}/ and its acoustic differences to its contrasting sounds /s/ and /ʃ^{j}/ in order to understand which of the experimental measures could discriminate the investigated sibilants and thus were potentially expected to change during the adaptation process. To that end, we first collapsed the baseline data of the six CVC words into three sound categories according to the final segment ([les] and [ves] into /s/, [les^{j}] and [ves^{j}] into /s^{j}/, and [leʃ^{j}] and [veʃ^{j}] into /ʃ^{j}/). Then, we performed RF modeling in conjunction with the Boruta procedure to compute the importance scores for all investigated acoustic parameters.

We additionally examined speakers' baseline production of the most relevant acoustic parameters in more detail by fitting linear-mixed models using the *lme4* package (version

1.1-13; Bates, Mächler, Bolker, & Walker, 2015). One model was fitted for each of the relevant acoustic parameters. Each model included the produced syllable and the interaction between the syllable and gender as fixed effects and the respective acoustic parameter as the dependent variable. Furthermore, all models included random intercepts for each participant as well as random slopes for each sound category. P-values were obtained with the *lmerTest* package (version 2.0-33) by Kuznetsova, Brockhoff, and Christensen (2017).

Subsequently, we examined speakers' production of the target sound /sʲ/ over the course of the experimental session. To this end, we modeled the data of each experimental phase separately, which allowed us to identify the acoustic parameters which participants systematically adjusted in reaction to upward and downward perturbation, and to investigate whether these adjustments became more robust as the experimental session went on. Furthermore, since we ran RF models with responses produced under noise-masked feedback, we were able to evaluate whether participants maintained the acquired adjustments in their production when their auditory feedback was replaced with noise.

In contrast to computations performed on the baseline data, all RF modeling procedures aimed at the investigation of the adaptation process were applied not to raw but to normalized parameter values which were obtained by subtracting each participant's mean of each acoustic parameter produced during the baseline phase in the respective word ([lesʲ] or [vesʲ]). This was done to exclude participant-specific differences regarding their absolute frequency magnitudes (e.g., due to gender differences) and considerably improve the computation time of the classification algorithm. By means of this normalization, the average values of all investigated parameters for [lesʲ] and [vesʲ] were set at zero for the baseline phase productions.

To investigate individual aspects of the adaptation process, we first computed variable importance scores for the investigated acoustic parameters separately for each speaker for the last shift phase of the experiment. Then, we analyzed the magnitude and temporal evolution of speakers' compensatory responses focusing on individually relevant parameters.

To analyze the adaptation process over time we employed GAMMs. We fitted models with normalized parameter values averaged across all participants and all experimental trials as dependent variables using the *mgcv* package (version 1.8-19; Wood, 2017b). The data of the unperturbed words containing the sounds /s/ and /ʃʲ/ were not included in the resulting GAMMs. All GAMM models were evaluated, interpreted, and visualized by means of the *itsadug* package (version 2.3) by van Rij et al. (2017).

In the model structure, we included random factor smooths with an intercept split for the perturbation direction (downward vs. upward) in order to assess individual compensation magnitude differences over the course of the experiment. The model also included a fixed effect which assessed the ‘constant’ effect of the perturbation direction independently from the temporal variation. Visual model inspection revealed that the residuals of the fitted GAMMs followed a normal distribution for all investigated measurements.

By extracting the fitted parameter curves estimated individually for each participant by the GAMM models, we were able to classify participants’ compensatory behavior into different groups.

3.3 Results

We begin to present our results with findings concerning speakers’ baseline productions of the three investigated sounds /s/, /s^j/, and /ʃ^j/. Then, in order to assess adaptational changes that occurred over the course of the experiment, we first identify acoustic parameters most relevant to the discrimination of the three fricatives as well as the applied perturbation direction. This information allows us to investigate the robustness and the temporal evolution of the adaptation process. Finally, investigating individual parameter importance measures, we are able to pinpoint specific compensatory adjustments applied by each speaker over the course of the experiment.

3.3.1 Baseline production of /s/, /s^j/, and /ʃ^j/

In this section, we provide a summary of the baseline values of all acoustic parameters across all investigated fricatives. The mean baseline measurements are summarized for all 19 analyzed participants in Table 3.5 on the following page. As expected, the target fricative /s^j/ was acoustically ‘flanked’ by the sounds /s/ and /ʃ^j/ for female as well as for male participants with regard to most investigated acoustic parameters.

Although all 13 acoustic parameters were identified as important to classifying the three sounds /s/, /s^j/, and /ʃ^j/, the outcome of the variable importance procedure also implied that F1, F2 and COG were by far the most relevant parameters for the discrimination between the investigated sound categories. The latter three parameters were separated by a larger gap

Table 3.5: Means of the investigated acoustic parameters produced for each sound during the baseline phase (no perturbation). Standard deviation is given in parentheses. Data are split by gender (f = female; m = male).

Parameter	/s/		/s^j/		/ʃ^j/	
	f (n = 14)	*m (n = 5)*	*f (n = 14)*	*m (n = 5)*	*f (n = 14)*	*m (n = 5)*
COG (Hz)	8337 (927)	7416 (1298)	7521 (822)	6761 (1640)	4336 (480)	3600 (258)
SD (Hz)	1805 (373)	2059 (425)	1892 (322)	2029 (431)	1556 (248)	1119 (194)
Skewness	0.24 (1.02)	0.44 (1.00)	0.44 (0.66)	0.78 (1.01)	1.82 (0.59)	3.01 (0.42)
Freq$_{\text{Mid}}$(Hz)	8071 (1030)	7588 (1423)	7381 (892)	7383 (1229)	6451 (405)	6118 (227)
AmpD$_{\text{Mid-MinLow}}$ (dB)	45.14 (5.04)	41.46 (3.63)	43.30 (5.35)	38.39 (3.91)	32.45 (4.96)	30.55 (2.15)
AmpD$_{\text{High-Mid}}$ (dB)	–9.55 (4.32)	–10.18 (5.69)	–11.71 (3.09)	–10.53 (4.94)	–16.36 (4.39)	–16.66 (4.28)
Level$_{\text{Low}}$ (dB)	–16.60 (4.66)	–8.39 (4.28)	–11.51 (5.80)	–7.49 (3.16)	–0.49 (4.28)	2.05 (4.89)
Level$_{\text{Mid}}$ (dB)	–2.10 (4.16)	–4.82 (6.65)	–2.93 (3.74)	–7.58 (6.22)	–8.67 (4.51)	–10.39 (7.0)
Level$_{\text{High}}$ (dB)	–11.77 (5.01)	–15.19 (10.23)	–14.82 (4.22)	–18.41 (10.10)	–25.01 (4.98)	–26.77 (10.17)
LevelD$_{\text{High-Mid}}$ (dB)	–9.66 (3.83)	–10.38 (5.51)	–11.89 (2.79)	–10.83 (4.92)	–16.34 (4.18)	–16.38 (4.08)
LevelD$_{\text{Mid-Low}}$ (dB)	14.50 (5.64)	3.58 (4.88)	8.58 (6.57)	–0.09 (6.85)	–8.18 (2.57)	–12.43 (2.21)
F1 (Hz)	587 (42)	508 (24)	442 (36)	391 (15)	442 (37)	404 (42)
F2 (Hz)	1979 (108)	1832 (88)	2340 (145)	2127 (127)	2334 (138)	2109 (166)

with respect to their mean importance scores compared to the remaining 10 parameters which were grouped much closer among each other (Table 3.6 on the next page).

The subsequent RF model was fed the data consisting of 300 tokens of each sound category produced by participants during the baseline phase, and employed all 13 acoustic parameters to classify the three sounds /s/, /s^j/, and /ʃ^j/. This model performed with a mean accuracy of 97.23 percent. However, a simpler model, including only F1, F2, and COG as predictive variables, performed insignificantly worse with a mean accuracy of 96.78 percent. In other words, the COG of a given sibilant fricative as well as F1 and F2 of the preceding vowel were practically sufficient to predict its phonemic category. These results are in general consistent with previous studies on Russian sibilants (e.g., Kochetov, 2017).

Looking in more detail at the differences regarding the three most distinctive acoustic parameters for the sounds /s/, /s^j/, and /ʃ^j/, we found that for female participants the F1 difference was significant between /s/ and /s^j/ (–145 Hz, $t = –20.326$, $p < .05$), but not between /ʃ/ and /s^j/ (0 Hz, $t = 0.001$, $p > .05$). In contrast to female speakers, for male participants average F1 was lower by 79 Hz ($t = 3.926$, $p < .05$) for /s/, –52 Hz ($t = –3.030$, $p < .05$) for /s^j/, and –38 Hz ($t = –1.911$, $p > .05$) for /ʃ^j/. For female speakers, F2 was on average significantly

Table 3.6: Acoustic parameters deemed important to classify whether a produced sound was /s/, /s^j/, or /ʃ^j/. Results of the variable importance computation performed on the data from the baseline phase (no perturbation).

Parameter	Mean importance score	Relevance decision
F1	61.60	important
F2	47.79	important
COG	38.68	important
$\text{LevelD}_{\text{Mid-Low}}$	22.72	important
$\text{Level}_{\text{Low}}$	20.53	important
$\text{AmpD}_{\text{Mid-MinLow}}$	19.05	important
Freq_{Mid}	15.92	important
$\text{Level}_{\text{High}}$	14.25	important
Skewness	13.80	important
$\text{LevelD}_{\text{High-Mid}}$	12.90	important
$\text{Level}_{\text{Mid}}$	12.80	important
$\text{AmpD}_{\text{High-Mid}}$	11.85	important
SD	11.24	important
shadowMax*	2.40	decision boundary

alpha = 0.01

* shadowMax is a “dummy” parameter computed by the Boruta algorithm to determine the importance decision boundary. See section 3.2.7 on page 52 for more details.

higher for /s^j/ in comparison to /s/ (361 Hz, t = 12.918, p < .05). However, the difference between /ʃ^j/ and /s^j/ was not significant (–6 Hz, t = –0.436, p > .05). In contrast to female speakers, for male participants, average F2 was lower by 146 Hz (t = –2.707, p < .05) for /s/, –213 Hz (t = –2.871, p < .05) for /s^j/, and –224 Hz (t = 2.937, p < .05) for /ʃ^j/. The COG values for the three investigated sounds were considerably different with /s/ exhibiting the highest and /ʃ^j/ the lowest COG. For female participants, average COG difference between /s/ and /s^j/ amounted to –816 Hz (t = –6.493, p < .05) and to 3185 Hz (t = 10.853, p < .05) between /s^j/ and /ʃ^j/. Compared to female participants, the COG values were on average lower for all three fricatives for male participants with a significant difference only for /ʃ^j/ (–736 Hz, t = –3.222, p < .05). These results are consistent with previous observations that COG can be used as a central tendency measure to discriminate between different sibilant categories, and that F1/F2 values of a vowel are an indicator of the palatalization status of the following sibilant. Furthermore, the results demonstrate similar relations among the three investigated sounds /s/, /s^j/, and /ʃ^j/ for female and male speakers.

3.3.2 F1, F2, and COG adjustments over the course of the experiment

Given the fact that the parameters F1, F2, and COG were deemed the most relevant measures for discriminating between the investigated sibilant categories, it appears reasonable to assume that participants' adaptation to spectral perturbations of /s^j/ would be reflected in one or more of these variables. To evaluate this hypothesis, in this section we examine production changes that occurred in F1, F2, and COG over the course of the experiment (see Figure 3.6).

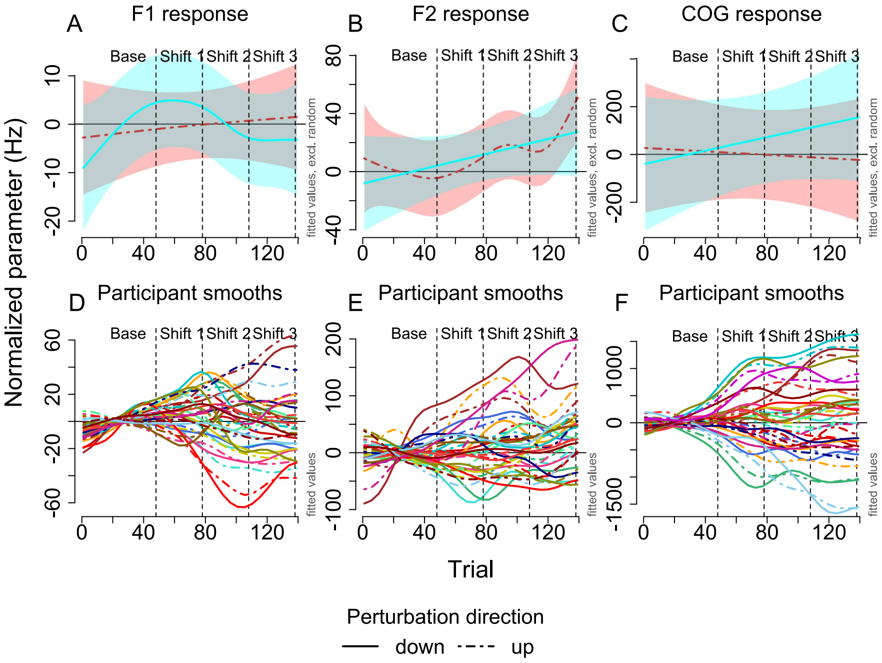

Figure 3.6: Visual summary of the GAMM models fitted for F1, F2, and COG. (A-C) Average compensatory effects (excluding random participant effects) for downward and upward perturbation plotted across baseline and shift phases of the experiment. Red and blue bands represent 95% confidence intervals. (D-F) Random smooths of the corresponding models color-coded for single participants.

Apart from the fact that all three parameters are characterized by changes and fluctuations that occurred over the course of the experiment, none of the variables takes on distinct average values depending on the applied perturbation direction, which would be a sign that participants developed different production strategies for the opposing shifts. Only for COG values do we observe a slight diverging tendency (Figure 3.6C). The random smooths suggest that the examined parameters are characterized by a high degree of adaptational variability across single participants.

We can think of several potential reasons for the observed null effects. The most obvious explanation would be that speakers were unable to learn to compensate for the applied perturbations. However, the random smooths presented in Figures 3.6D-F suggest that participants' production did indeed change after the baseline phase as the experimental session progressed further. Thus, an alternative explanation is that the acoustic parameters F1, F2, and COG, deemed previously important to discriminate between the sounds /s/, /s^j/, and /$ʃ^j$/ are not necessarily adequate measures for the evaluation of participants' adapting behavior. Additionally, considering the insight from previous oral-articulatory perturbation studies that compensation in fricatives involves a set of varying articulatory parameters (see discussion in Chapter 1), it might be completely inappropriate to try to assess the adaptation process in fricatives based solely on one specific or a fixed set of acoustic parameter(s). During the remaining analysis, we examine which acoustic parameters were in fact relevant to discriminating whether a given token of the fricative /s^j/ was produced under downward or upward perturbation.

3.3.3 /s^j/-production under unperturbed and perturbed feedback

In this section, we present the results of the variable selection procedure for all 13 investigated acoustic parameters separately for all experimental phases where participants spoke under unperturbed or perturbed auditory feedback (baseline, shift 1, shift 2, and shift 3 phases). All importance decisions made by the Boruta algorithm for the corresponding experimental phases are summarized in Table 3.7.

Although no actual perturbation was applied during the baseline phase, we were nonetheless technically able to calculate variable importance decisions for the classification of /s^j/ tokens for this phase since each perturbation direction (downward vs. upward) was associated with a certain experimental stimulus ([ves^j] vs. [les^j]). The variable importance decisions were calculated for the baseline phase for completeness' sake as well as to provide a sanity check of the applied variable selection procedure since none of the investigated acoustic parameters was expected to become relevant for the baseline data.

For each of the three shift phases, the Boruta procedure identified different numbers of relevant parameters. While no particular parameter was important for predicting the direction of the applied perturbation during the first shift phase, 6 and 11 parameters were chosen

Table 3.7: Overview of the importance decisions for all experimental measures regarding the classification task of /s^j/ tokens with regard to the applied perturbation direction. The decisions are given for the baseline, shift 1, shift 2, and shift 3 phases.

Parameter	**Baseline**	**Shift 1**	**Shift 2**	**Shift 3**
COG (Hz)	-	-	important	important
SD (Hz)	-	-	-	important
Skewness	-	-	-	important
Freq$_{\text{Mid}}$(Hz)	-	-	-	important
AmpD$_{\text{Mid-MinLow}}$ (dB)	-	-	-	important
AmpD$_{\text{High-Mid}}$ (dB)	-	-	-	important
Level$_{\text{Low}}$ (dB)	-	-	important	important
Level$_{\text{Mid}}$ (dB)	-	-	important	-
Level$_{\text{High}}$ (dB)	-	-	-	-
LevelD$_{\text{High-Mid}}$ (dB)	-	-	-	important
LevelD$_{\text{Mid-Low}}$ (dB)	-	-	important	important
F1 (Hz)	-	-	important	important
F2 (Hz)	-	-	important	important

for the second and the third shift phase, respectively. These results are consistent with the idea that speakers progressively adjusted their compensatory strategies for the two opposing perturbation directions in the course of the experiment.

For the third shift phase, the variable selection procedure deemed 11 out of 13 investigated parameters relevant (Table 3.8 on the next page). The five variables deemed most important were the second (SD) and the third (skewness) spectral moments, the RMS level of the lowest frequency band (600-5500 Hz), the amplitude difference between the highest (11000-16000 Hz) and the mid frequency band (5500-11000 Hz), as well as the first formant (F1) of the vowel preceding the perturbed sibilant. The COG was the 6th most important variable for the correct classification of the applied perturbation direction.

3.3.4 SD, skewness, and Level$_{\text{Low}}$ adjustments over the course of the experiment

The GAMM models estimated for the three acoustic parameters which achieved the highest variable importance scores during the third shift phase (SD, skewness, Level$_{\text{Low}}$) revealed

Table 3.8: Acoustic parameters deemed important to classify whether a token of /s^j/ was produced under downward or upward perturbation. Results of the variable importance computation performed on the data from the third shift phase.

Parameter	**Mean importance score**	**Relevance decision**
SD	9.59	important
Skewness	7.62	important
$\text{Level}_{\text{Low}}$	6.99	important
F1	6.61	important
$\text{AmpD}_{\text{High-Mid}}$	6.09	important
COG	5.44	important
$\text{LevelD}_{\text{High-Mid}}$	5.01	important
F2	4.01	important
$\text{AmpD}_{\text{Mid-MinLow}}$	3.82	important
Freq_{Mid}	3.05	important
$\text{LevelD}_{\text{Mid-Low}}$	2.92	important
shadowMax*	2.63	decision boundary
$\text{Level}_{\text{Mid}}$	-	unimportant
$\text{Level}_{\text{High}}$	-	unimportant

alpha = 0.01

* shadowMax is a “dummy” parameter computed by the Boruta algorithm to determine the importance decision boundary. See section 3.2.7 on page 52 for more details.

that the applied perturbation had a measurable effect on participants’ production (Figures 3.7A-C). Even though the most observed average effects were not statistically significant, their direction was opposite the direction of the applied perturbation for all three parameters.

On trials with downward perturbation, the observed average changes amounted to 76.10 Hz (t = 3.364, p < .05) for SD, –0.053 (t = –0.826, p > .05) for skewness, and –0.58 dB (t = –1.446, p > .05) for $\text{Level}_{\text{Low}}$. On trials with upward perturbation, the constant effect of perturbation resulted in an average change of –64.71 Hz (t = –1.865, p > .05) for SD, 0.057 (t = 0.629, p > .05) for skewness, and 0.82 dB (t = 1.489, p > .05) for $\text{Level}_{\text{Low}}$.

Examining the effect of perturbation over time, the fitted models revealed that, despite minor differences across the investigated parameters, the absolute magnitude of the average compensatory effects increased for both perturbation directions. As can be seen in Figures 3.7A-C, the average adaptation magnitude started to diverge after the baseline and kept increasing in the course of the remaining experiment.

Examining the mean difference between changes for both perturbation directions, the

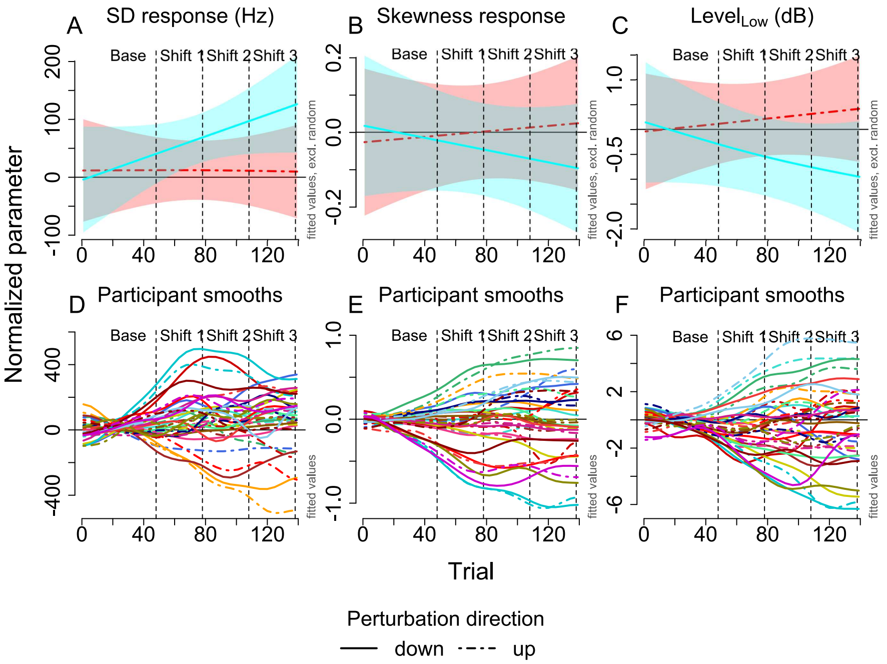

Figure 3.7: Visual summary of the GAMM models fitted for spectral SD, skewness, and Level$_{\mathrm{Low}}$. (A-C) Average compensatory effects (excluding random participant effects) for downward and upward perturbation plotted across baseline and shift phases of the experiment. Red and blue bands represent 95% confidence intervals. (D-F) Random smooths of the corresponding models color-coded for single participants.

model revealed that only spectral SD displayed a significant effect of compensation by the last shift phase of the experiment. The average SD difference between both perturbation directions (downward vs. upward) amounted to 57 Hz (95% CI [–9 124]) by the end of the first shift phase and to 87 Hz (95% CI [3 171]) by the end of the second shift phase. By the end of the experiment, the average SD difference reached 116 Hz (95% CI [0 232]). Although the parameters skewness and Level$_{\mathrm{Low}}$ displayed trends in the expected directions, the average compensatory effects in these parameters were not significantly different between the two opposing perturbation directions.

3.3.5 Responses produced under noise-masked feedback

In this section, we present the results of the variable selection procedure for the three experimental phases where participants spoke under noise-masked auditory feedback (noise 1, noise 2, and noise 3 phases). Although no perturbation was delivered during noise phases, we were technically able to calculate variable importance decisions for the classification of /s^j/ tokens since each perturbation direction (downward vs. upward) was associated with a certain experimental stimulus ([ves^j] vs. [les^j]). All importance decisions made by the Boruta algorithm for the corresponding experimental phases are summarized in Table 3.9.

Table 3.9: Overview of the importance decisions for all experimental measures regarding the classification task of /s^j/ tokens with regard to the perturbation direction. The decisions are given for the noise 1, noise 2, and noise 3 phases.

Parameter	**Noise 1**	**Noise 2**	**Noise 3**
COG (Hz)	-	-	-
SD (Hz)	-	important	-
Skewness	-	-	-
Freq_{Mid}(Hz)	important	-	-
$\text{AmpD}_{\text{Mid-MinLow}}$ (dB)	-	-	-
$\text{AmpD}_{\text{High-Mid}}$ (dB)	-	important	-
$\text{Level}_{\text{Low}}$ (dB)	-	-	important
$\text{Level}_{\text{Mid}}$ (dB)	-	-	-
$\text{Level}_{\text{High}}$ (dB)	-	-	-
$\text{LevelD}_{\text{High-Mid}}$ (dB)	-	-	-
$\text{LevelD}_{\text{Mid-Low}}$ (dB)	-	-	-
F1 (Hz)	-	-	-
F2 (Hz)	-	-	-

In comparison to the number of relevant parameters for phases completed by participants under perturbed auditory feedback (Table 3.7), the number of parameters deemed important during the noise phases is much smaller. Furthermore, there is no observable trend in this number across the three noise phases. These findings suggest that, in general, participants did not transfer any considerable production adjustments from the shift to the noise phases.

For the third noise phase, the variable selection procedure deemed only $\text{Level}_{\text{Low}}$ with an importance score of 9.79 and a decision boundary score of 3.19 important. That means that the average importance score for the remaining 12 parameters either could not be computed or was below the score of 3.19.

3.3.6 Prediction accuracy as a measure of adaptation

Based on the variable importance decisions computed in sections 3.3.3 and 3.3.5, we ran an RF model for each experimental phase which allowed us to investigate whether the applied perturbation direction could be predicted from the acoustic parameters of participants' /s^j/-productions, and whether this prediction would improve in the course of the experimental session hinting at a progression of the adaptation process. In order to establish an initial prediction accuracy score, we ran an RF model on the baseline data knowing that no perturbation was applied in this phase. For phases in which no particular acoustic parameter was deemed relevant for predicting the applied perturbation direction (i.e., baseline and shift 1 phases; see Table 3.7), all 13 investigated parameters were used to compute the corresponding RF model.

For the baseline phase, the data fed to the RF model consisted of 300 /s^j/ tokens potentially produced under downward or upward perturbation. For the three shift phases, the data included on average 250 /s^j/ tokens produced under downward perturbation and 250 /s^j/ tokens produced under upward perturbation; the exact number of tokens fluctuated somewhat across the shift phases since a few tokens were removed from the data set due to failed perturbation delivery (see section 3.2.5 on page 49 for details). Finally, for the three noise phases completed under masked feedback, the data included 150 /s^j/ tokens. The prediction accuracy scores for all these RF models are summarized in Figure 3.8.

As can be seen in Figure 3.8, the mean accuracy prediction for the perturbation direction does not exceed 50 percent (chance level) during the baseline phase. This result reflects the fact that no perturbation was delivered during this phase, and the examined /s^j/ tokens should not differ in this regard. The mean accuracy of 55.20 percent for the first shift phase suggest that participants might have begun to develop different adjustments for the two opposing perturbation directions. For the second and the third shift phase, the prediction accuracy scores increase to 59.46, and then again to 66.59 percent indicating that participants were able to improve their compensatory adjustments as the experiment progressed. During each noise phase that followed a shift phase, the prediction accuracy drops on average supporting the hypothesis that participants did not maintain all of their recently gained production adjustments when auditory feedback was no longer available.

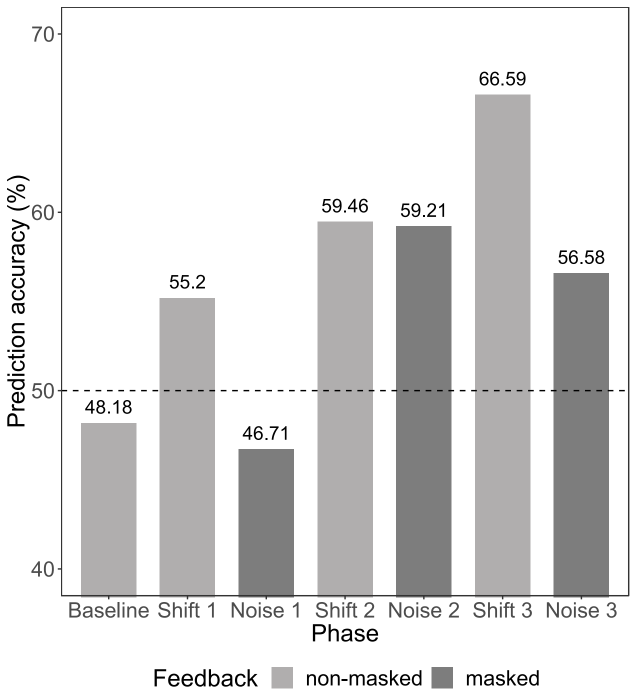

Figure 3.8: Summary of the prediction accuracy scores of the fitted RF models across all experimental phases with regard to the classification task of the applied perturbation direction. The dashed line at 50% denotes the chance level.

3.3.7 Individual adapting behavior

In this section, we discuss adapting behavior of individual speakers by focusing on specific acoustic parameters to further substantiate our overall findings. To this end, we first conducted the variable selection procedure for the third shift phase of the experiment separately for each of the 19 participants. These results are summarized in Table 3.10.

As can be seen in Table 3.10 on the following page, the number of acoustic parameters deemed important for the prediction of the applied perturbation direction varied considerably across participants ranging from one (f3) to eight (m5) variables. There was also one participant for whom no parameter was identified as important (m2). This observation is crucial as it seems that none of the investigated acoustic parameters was invariably employed by speakers to adapt to the applied perturbation in the course of the experiment. On the contrary, our participants appear to have modified different regions and properties of the frequency spectrum of the target fricative /s^j/.

For instance, a quite high number of participants (see importance decisions for f1, f5, f7,

Table 3.10: Overview of the importance decisions for all experimental measures regarding the classification task of /s^j/ tokens with regard to the applied perturbation direction. The variable importance computation was performed on speakers' individual data from the last shift phase of the experiment. Each parameter is marked either as being among the three most important (++) or important (+).

Parameter	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	m1	m2	m3	m4	m5	Total ++	Total +
COG (Hz)				++					++	++		++	++						++	6	-
SD (Hz)			++	+	++			+	++		++		+		++				++	6	3
Skewness	+			++			++	++	++				++				++		++	7	1
Freq$_{Mid}$(Hz)				++				+		+					+		++	+		2	4
AmpD$_{Mid\text{-}MinLow}$ (dB)				+	+							++			++				+	2	3
AmpD$_{High\text{-}Mid}$ (dB)	+	+		+	+	++	+		+						+				+	1	8
Level$_{Low}$ (dB)	++				++		++	++	+			++	+	++			++	++		8	2
Level$_{Mid}$ (dB)		++					+			++			+	++	++			++		5	2
Level$_{High}$ (dB)									+	+			++							1	2
LevelD$_{High\text{-}Mid}$ (dB)	+	++		+		++	+					+							+	2	5
LevelD$_{Mid\text{-}Low}$ (dB)	++	++			++			++	+		++	+	+						+	5	4
F1 (Hz)	++						++			++					+			++		4	1
F2 (Hz)						++				+	++			++					++	4	1
Total	6	4	1	7	5	3	6	5	7	6	3	5	7	3	4	-	3	4	8		

f8, f9, f12, f13, f14, m3, and m4) adjusted the values of $\text{Level}_{\text{Low}}$ increasing or decreasing the intensity of lower frequencies. On the other hand, fewer participants (f2, f7, f10, f13, f14, m1, m4) modified the values of $\text{Level}_{\text{Mid}}$, and even fewer speakers (f9, f10, f13) changed the values of $\text{Level}_{\text{High}}$. Only for two speakers did the intensity change across all three frequency bands (f10 and f13). For several speakers, the amplitude adjustments in single frequency bands were accompanied by changes in relative measures such as $\text{AmpD}_{\text{High-Mid}}$, $\text{Level}_{\text{High-Mid}}$, and $\text{Level}_{\text{Mid-Low}}$. However, there were only four speakers (f4, f6, f11, and m5) for whom the relative measures were deemed more important compared to amplitude intensities of single frequency bands.

Although the spectral moments COG, SD, and skewness were very often among the three individually highest ranking variables, there was only one participant whose compensatory adjustments could be solely described by SD changes (f3). For the remaining speakers, changes in one of the spectral moments rather accompanied adjustments in different frequency bands. This observation suggests that spectral moments may provide sufficient means to detect general changes in speakers' production, but are not sensitive enough to investigate speakers' compensatory adjustments in more detail.

Beginning with $\text{Level}_{\text{Low}}$, we assigned all participants to different groups depending on the parameter which in each case was deemed most important for predicting the direction of the applied perturbation. Each participant – except for m2 for whom none of the investigated acoustic parameters was deemed relevant for discriminating between the two perturbation directions – was assigned to a single group. In that way, we were able to define four groups. Then, for every group, we extracted participants' random smooths from the corresponding GAMM models (Figure 3.9). In Table 3.11 on page 69, we also summarized parameter values for each speaker as fitted by the GAMM models for the third shift phase.

In the first group, for which $\text{Level}_{\text{Low}}$ was determined as the most relevant parameter, seven out of eight speakers (f1, f5, f7, f8, f14, m3, and m4), produced low frequencies (600-5500 Hz) with higher amplitude, making them more salient, under upward perturbation compared to downward perturbation (Figure 3.9A). This difference was significant for five speakers (see f1, f7, f8, f14, and m4 in Table 3.11). All five speakers of the second group (f4, f9, f10, f13, and m5), produced the target sound /s^{j}/ with higher COG values on trials with downward perturbation compared to upward perturbation (Figure 3.9B). This difference was significant for three speakers (see f9, f13, and m5 in Table 3.11). Two out of three speakers of the third group produced the target fricative with significantly higher SD (see f3 and

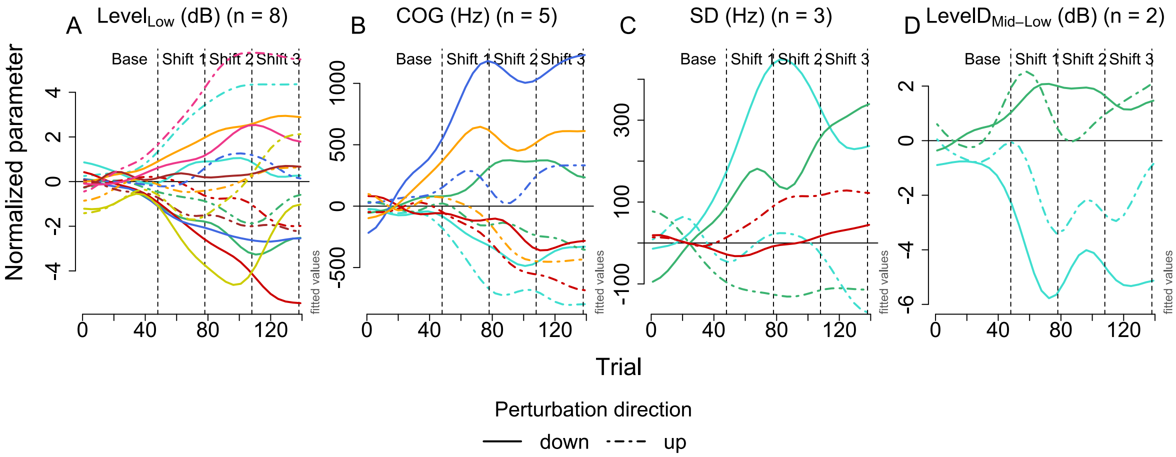

Figure 3.9: Random smooths of the GAMM models fitted for Level$_{\text{Low}}$, COG, spectral SD, and LevelD$_{\text{Mid-Low}}$ and plotted across baseline and shift phases of the experiment; lines are color-coded for single participants.

m1 in Table 3.11), i.e., broadening its spectrum when it was perturbed downwards (Figure 3.9C).

One speaker each from the first (f12) and the third (f11) group appears to have rather followed the applied perturbation since their reaction was swapped for both perturbation directions compared to the remaining speakers in their respective groups. However, the difference between the two perturbation directions was significant only for participant f12 (see Table 3.11).

While the results summarized above are quite straightforward, there are some speakers whose behavior is rather difficult to interpret. For instance, for speakers of the fourth group (f2 and f6), the parameter deemed most relevant for discriminating between both perturbation directions was LevelD$_{\text{Mid-Low}}$ (Figure 3.9D). While for f6 the difference between values produced on trials with upward and downward perturbation was not significant, f2 produced the /s^j/ spectrum with significantly lower LevelD$_{\text{Mid-Low}}$ values on trials with downward perturbation. Since for both participants Level$_{\text{Low}}$ was not considered a relevant parameter for predicting the perturbation direction, it appears that these speakers rather adjusted the amplitude of the mid frequencies (5500-11000 Hz). However, it is not clear how this adjustment influenced their perception of the perturbed tokens.

Although not a single parameter was deemed relevant for m2 to differentiate between tokens produced under opposite perturbation directions, this speaker significantly altered his

Table 3.11: Mean changes in the individually relevant parameters between the baseline and the last shift phase as fitted by the GAMM models. The parameter values are given in dB ($\text{Level}_{\text{Low}}$ and $\text{LevelD}_{\text{Mid-Low}}$) or Hz (SD and COG).

ID	**Parameter**	**Perturb.**		**Diff.**	**ID**	**Parameter**	**Perturb.**		**Diff.**
		up	*down*				*up*	*down*	
f1	$\text{Level}_{\text{Low}}$	4.36**	0.27	–4.10**	f11	SD	122	44	-78
f2	$\text{LevelD}_{\text{Mid-Low}}$	–0.85	-5.13**	–4.28**	f12	$\text{Level}_{\text{Low}}$	–2.24**	2.88**	2.90**
f3	SD	–171**	237**	408**	f13	COG	332	1231**	899**
f4	COG	–798**	–330	468	f14	$\text{Level}_{\text{Low}}$	2.13**	–5.01**	–3.15**
f5	$\text{Level}_{\text{Low}}$	–0.59	–2.52**	–1.93	m1	SD	–114	339**	453**
f6	$\text{LevelD}_{\text{Mid-Low}}$	2.13	1.47	-0.66	m2	-	-	-	-
f7	$\text{Level}_{\text{Low}}$	–1.99**	–5.45**	–3.46**	m3	$\text{Level}_{\text{Low}}$	0.64	2.88**	2.24
f8	$\text{Level}_{\text{Low}}$	0.14	-2.54**	-2.68**	m4	$\text{Level}_{\text{Low}}$	5.47**	1.79**	-3.68**
f9	COG	–355**	234	589**	m5	COG	–434	612**	1046**
f10	COG	–685**	–282	402					

**p < .05

production over the course of the experiment. Specifically, during the last shift phase he produced the target sound with significantly lower $\text{Level}_{\text{Low}}$ (–5.80/–6.31 dB) and significantly higher COG (1385/1632 Hz) values on trials with downward and upward perturbation. To exclude the possibility that the applied perturbation was less effective for m2, we analyzed his auditory feedback during the first shift phase. This analysis revealed no qualitative differences compared to the average values reported in Table 3.2 (see section 3.2.3 on page 44). On trials with downward perturbation, the amplitude of $\text{Level}_{\text{Low}}$ increased for m2 on average by 1.45 dB, while the amplitude for $\text{Level}_{\text{High}}$ decreased by 17.78 dB. On trials with upward perturbation, the amplitude of $\text{Level}_{\text{Low}}$ decreased for m2 by 5.56 dB, while the amplitude for $\text{Level}_{\text{High}}$ increased by 9.39 dB. As for the remaining speakers, the shift effects were observable as complementary changes in low and high frequency bands of m2's /s^{j}/-productions.

3.4 Discussion

In contrast to a growing body of auditory perturbation research detailing the role of auditory feedback for vowel production, equivalent empirical work dealing with consonants remains rare. Although a few studies have successfully demonstrated that speakers react to auditory perturbations of frication noise during /s/ and /ʃ/ production, the results of these studies are rather scattered and difficult to compare to each other due to methodological differences and limitations (Shiller et al., 2009; Casserly, 2011). In the current study, we sought to overcome these limitations by building upon past phonetic as well as auditory and oral-articulatory perturbation studies of fricatives.

During three shift phases of the current experiment, speakers had to produce the sibilant fricative /s^j/ while its spectrum was pitch shifted either downwards or upwards depending on the presented stimulus ([ves^j] vs. [les^j]). The alternating perturbation directions were there to ensure that changes observed in participants' speech were not due to unrelated factors, e.g., general fatigue, but were systematically dependent on the experimental manipulation. In contrast to perturbation studies of vowels, where typically only formants are affected during the adaptation process, we observed compensatory behavior which was characterized by broader spectral changes concerning a set of acoustic parameters. This finding is consistent with previous oral-articulatory perturbation studies (McFarland & Baum, 1995; McFarland et al., 1996).

The analysis of the adaptation process was impeded by the fact that the number of acoustic parameters modified in response to the perturbation varied across speakers and experimental phases. Consequently, it was not possible to apply more conservative modeling techniques without making any additional assumptions regarding the question of whether certain acoustic parameters serve as appropriate markers of speakers' compensatory behavior. For instance, past auditory perturbation studies of fricatives by Shiller et al. (2009) and Casserly (2011) made an implicit assumption that COG, which repeatedly has been shown to systematically differ across fricative phonemes (e.g., Forrest et al., 1988), serves as a sufficient compensatory measure.

By means of variable importance computations and subsequent RF modeling, we were able to show that COG was indeed helpful for discriminating between the sibilant fricatives /s/, /s^j/, and /ʃ^j/ produced without perturbation, but was on its own not meaningful in describing participants' adapting behavior in reaction to the perturbation of /s^j/. A GAMM model, which included COG as a dependent variable and was computed across baseline and shift phases, revealed that for many speakers COG changes were independent of the applied perturbation direction. In other words, although on average COG changed significantly over the course of the experiment, it was not an appropriate marker of compensatory behavior and appeared to rather reflect changes that were consequences of other compensatory adjustments.

To avoid additional, possibly unreasonable, assumptions regarding the status of certain acoustic parameters as compensation markers, we subjected all 13 investigated parameters of speakers' /sʲ/-productions to the analysis of variable importance scores for each of the experimental phases. Thus, we were able to identify acoustic parameters involved in the adaptation process in an objective and statistically valid fashion. This procedure was indispensable for further analysis of the data as it would be infeasible to impossible to select appropriate acoustic parameters manually, particularly since as the experimental session went on generally more and more parameters were prone to change.

The variable importance scores revealed that additionally defined spectral measures, such as $\text{Level}_{\text{Low}}$, were important for predicting under which perturbation direction a /sʲ/ token was produced. For a few participants, besides the adjustments of fricative-internal parameters, F1 and F2 values of the unperturbed vowel preceding the perturbed /sʲ/ were deemed important for predicting the applied perturbation direction. The fact that speakers employed F1 and F2 values to compensate for the fricative shifts is rather unsurprising considering that both parameters were previously deemed important for predicting the category of the investigated fricatives /s/, /sʲ/, and /ʃʲ/. This observation may be taken as strong evidence of adapting behavior since it suggests that speakers were anticipating the effect of the shifts and planned for their compensation already during the preceding segment.

The average prediction accuracy scores of RF models computed for each phase based on previously determined variable importance scores increased considerably over successive shift phases suggesting that on the group level participants improved their compensatory strategies as the experimental session progressed. This trend was consistent with the results of GAMM models computed across baseline and shift phases for acoustic parameters deemed the most important for predicting the applied perturbation direction on the group level (SD, skewness, and $\text{Level}_{\text{Low}}$). Although the acoustic difference between /sʲ/ tokens produced under downward vs. upward perturbation was significant only for spectral SD, overall RF modeling results suggest that speakers were able to develop two different compensatory strategies for the target sound /sʲ/.

Furthermore, by means of variable importance and prediction accuracy scores it was possible to identify differences between /sʲ/ tokens produced under perturbed and under noise-masked feedback. For the latter, only a few acoustic parameters were identified as important for predicting the applied perturbation direction such that the resulting RF models performed considerably worse in comparison to models computed for data produced under perturbed feedback. These findings are consistent with previous results by Honda et al. (2002) and Jones and Munhall (2003) who observed degraded compensation under noise-masked feedback. Nevertheless, we observed an overall small increase in prediction accuracy scores across the three noise phases which suggests that speakers were able to retain some of their compensatory adjustments after a certain number of experimental trials.

To trace specific compensatory strategies employed by each speaker over the course of the experiment, we computed individual variable importance scores based on speakers' /s^j/-productions near the end of the experiment. Guided by these individual scores, we extracted participants' individual smooths from GAMM models computed previously for the investigated acoustic parameters. Thus, it was possible to assign all participants to a few specific groups which were characterized by similar compensatory behavior.

This procedure revealed that about 42 percent of the speakers adjusted the amplitude of the low frequency band (600-5500 Hz) in reaction to the applied spectral shifts. Specifically, when the balance of the overall spectral energy was tilted into direction of lower frequencies on downward shifts, these speakers decreased the amplitude of the low frequency band. On the other hand, when the spectrum of the target sound was shifted upwards tilting the balance of the overall spectral energy into direction of higher frequencies, these speakers increased the amplitude of the low frequencies. Additionally, with this approach we were able to identify two speakers who appeared to follow the applied shifts and one speaker who was not able to develop two distinct compensatory strategies for the target sound /s^j/.

Our observation that individual speakers modified different acoustic parameters of the fricative spectrum in reaction to the applied auditory shifts is analogous to findings from previous oral-articulatory perturbation studies. These studies have demonstrated that speakers may adjust different articulatory parameters (e.g., constriction location, jaw height, tongue grooving) in reaction to articulatory perturbation (e.g., Hamlet & Stone, 1978; Flege et al., 1988; Honda et al., 2002; Brunner et al., 2011). This highly diverse nature of the adaptation process in fricatives suggests that previous accounts that tried to classify speakers' adapting behavior on the basis of single acoustic parameters as following or counteracting the perturbation were too simplistic. Instead, we suggest that increases or decreases in values of single parameters occurring independently of the applied perturbation direction are merely markers of exploratory behavior during which speakers try to identify appropriate compensatory movements to produce the intended acoustic output.

Alternatively, the adaptation variability observed across speakers might be caused by how different individual speakers may perceive the effects of the applied perturbation. Specifically in the case of spectral shifts, the resulting auditory changes in lower and higher frequency bands of the spectrum might be perceived as more or less salient, due, for instance, to a slight high-frequency hearing loss. Unfortunately, participants' hearing thresholds were not obtained during the experiment which would allow for a systematic assessment of this hypothesis. Although we cannot completely rule out this explanation, the individual participant smooths extracted from GAMM models computed across baseline and shift phases for several acoustic parameters of the target sound /s^j/ suggest that speakers were able to perceive the spectral shifts since all investigated parameters (F1, F2, COG, SD, skewness, and $\text{Level}_{\text{Low}}$) exhibited significant deviations from the baseline phase.

Ultimately, we believe that the adaptation variability observed in our study across individual speakers was rather introduced at the production stage. This hypothesis is further supported by an informal comparison between the current data and previous auditory perturbation studies of vowels (e.g., Purcell & Munhall, 2006). The central insight of this comparison is the fact that the number of speakers who were able to distinctively compensate for both perturbation directions in the current study is significantly lower (about 55-75 percent depending on how strict the decision criterion is applied) compared to the perturbation studies of vowels (about 90-95 percent). We suggest that this discrepancy is due to a higher degree of acoustic-articulatory complexity of the adaptation in fricatives compared to the adaptation in vowels. While the latter mostly requires either adjustments of the jaw height or vertical and/or horizontal displacements of the tongue to compensate for F1/F2 perturbations, compensatory adjustments in fricatives encompass a more complex coordination of such parameters as constriction location, tongue grooving, and jaw height. This hypothesis is consistent with previous oral-articulatory studies which arrived at a similar conclusion after the comparison of compensatory adjustments for vowels and fricatives (e.g., McFarland & Baum, 1995).

In summary, our analyses suggest that although speakers employ their auditory feedback to detect errors during fricative production, there are additional articulatory and auditory factors that have influence on their individual compensatory performance. While some of the previous phonetic work has demonstrated the importance of analyzing several acoustic parameters of fricative spectra, the current study provides support for the relevance of examining speakers' compensatory behavior with a multi-parameter approach that takes the complex acoustic-articulatory nature of fricatives into account.

Chapter 4

Articulatory complexity of the adaptation task

4.1 Introduction

In the previous chapter, we demonstrated that speakers systematically employ their auditory feedback to correct for perturbations during fricative production.[1] However, based on presented statistical analyses it is also apparent that the degree of adaptation for fricatives is smaller compared to what has been previously observed in studies examining formant perturbation. As outlined in the discussion section of the previous chapter, it appears to us that the main reason for this difference is a higher articulatory complexity of the adaptation task in the case of fricatives compared to vowels. Indeed, as discussed in Chapter 1 on page 8, similar observations were previously made in oral-articulatory perturbation studies cross-examining adaptation in vowels and consonants. A varying degree of adaptation, however, is not only observable when comparing vowels and consonants, but also occurs across studies that investigate similar sounds but employ rather different perturbation methods.

For instance, when Fowler and Turvey (1980) examined English speakers' productions of vowels /i, ɛ, a, ɔ, ʌ, u/ spoken with a 14 mm bite-block between their teeth, the authors found that speakers were able to adapt to these perturbations within the first couple of trials. Furthermore, the introduced perturbation did not affect participants' response times. These findings are overall similar to observations made in other studies with English and Swedish

[1] Some of the ideas presented in this chapter were previously published in Klein, Brunner, and Hoole (2018) and Klein, Brunner, and Hoole (2019c).

speakers employing analogous experimental paradigm (e.g., Lindblom, Lubker, & Gay, 1979; Gay et al., 1981; Kelso & Tuller, 1983).

On the other hand, in a study by Savariaux et al. (1995) when speakers' lips were blocked with a 25-mm tube during the production of the French /u/, only six out of 11 speakers were able to partially compensate for the labial perturbation and only one speaker compensated completely by changing the constriction location from a velo-palatal to a velo-pharyngeal region. In a follow-up study by Savariaux et al. (1997), which employed an identical experimental set-up, only one out of nine speakers was able to successfully compensate for the applied perturbation. Remarkably, two additional speakers were able to achieve complete compensation after a training session producing the target sound /u/ after an /o/ which induced a retraction of the tongue required to adapt for the lip-tube perturbation.

Savariaux et al. (1995) interpreted their initial results as support for the idea that the varying degree of compensation among participants was due to "speaker-specific internal representations of articulatory-to-acoustic relationships". This hypothesis appears to be obvious and was later reformulated by Lametti et al. (2012) in the context of somatosensory vs. auditory feedback preference although it somewhat clashes with observations made by Ghosh et al. (2010) who found that the degree to which individual speakers are sensitive to changes in articulatory and auditory feedback signals are positively correlated with each other.

Apart from inconsistencies with other findings, the attempt to explain adaptation variability by ascribing it to inter-speaker differences alone fails to offer a general explanation for variability that is observed across different experiments. Considering the observation that articulatory training might facilitate adaptation, Savariaux et al. (1997) entertain the idea that some speakers are initially not able to find the correct compensatory strategy due to an insufficient representation of an articulatory-to-acoustic relation. Put more generally, we believe that under certain perturbation conditions some speakers are not able to manage the articulatory complexity of the adaptation task at hand which has an impact on its outcome. Let us further expand on this idea with a comparison of compensatory strategies observed across bite-block and lip-tube perturbation experiments.

It is plausible to assume that the adaptation to bite-block perturbation during production of vowels requires an articulatory adjustment that is more similar to the unperturbed condition compared to the case of lip-tube perturbation during the production of /u/. During the first task, participants are merely required to lift their tongue more strongly than usual since their jaw, which normally assists at this task, is blocked. Indeed, this compensatory pattern was observed by Gay et al. (1981) for the production of a perturbed /i/ by means of x-ray imaging. Furthermore, the direction of the compensatory tongue movement does not change due to the perturbation.

During the lip-tube perturbation, on the other hand, participants have to compensate for blocked lip rounding by retracting their tongue. This articulatory adjustment is less obvious as the articulator used to compensate for the perturbation and its movement direction are less associated with the usual articulatory configuration used to produce the intended sound. As a consequence, the adaptation process may take longer and fewer speakers are able to identify the appropriate articulatory adjustments to compensate for the perturbation. See Savariaux et al. (1995) for corresponding x-ray images.

A difference regarding articulatory demands of two adaptation tasks, as in the case of bite-block vs. lip-tube perturbation, appears to equally offer a potential explanation for the smaller adaptation degree observed during fricative perturbation in the previous chapter compared to previous formant perturbation studies. We believe that while speakers are able to correct for F1 and/or F2 perturbations rather easily by adjusting their vertical and/or horizontal tongue positions, quite transparent articulatory movements required to produce different vowel sounds, a successful compensation in the case of fricatives involves arguably more, less obvious, articulatory adjustments such as constriction location, tongue grooving, and the jaw height (Brunner et al., 2011).

Thus, the goal of the current chapter is to investigate the influence of articulatory complexity of an adaptation task on the degree of adaptation by comparing speakers' compensatory patterns across both experiments described in two previous chapters. In order to do so, we re-examined speakers' adaptation degree to F2 perturbation by means of RF modeling. This allowed us to obtain prediction accuracy scores for downward and upward perturbation for vowels and to compare them to accuracy scores computed for fricatives. To exclude potential confounding effects of speaker-specific traits claimed to influence the adaptation outcome (Savariaux et al., 1995; Lametti et al., 2012), we let the same speakers complete both experiments. Consequently, we were able to correlate accuracy scores computed for both experiments within each speaker. Additionally, we examined variable importance scores of different acoustic parameters deemed important for the classification of speech tokens produced under downward or upward perturbation. Tracking the importance scores over the course of both experiments, we were able to uncover compensatory strategies employed by participants to adapt for applied perturbations.

If the articulatory complexity of an adaptation task does not have an influence on its outcome, individual speakers are expected to compensate in equal amounts during auditory perturbation of vowels and fricatives such that we should see a positive correlation between accuracy scores across both experiments. Furthermore, emergence of a single or dominant compensatory strategy for each experiment among participants would additionally undermine the influence of articulatory complexity on the outcome of an adaptation task. On the other hand, absence of a positive correlation of accuracy scores would provide some evidence against the hypotheses that compensation is characterized by speakers' individual articulatory-acoustic relations or feedback channel preferences.

4.2 Methods

4.2.1 Participants

Eighteen speakers (14 females, 4 males) who participated in the formant perturbation experiment described in Chapter 2 (Experiment 1), completed also the fricative experiment described in Chapter 3 (Experiment 2). The mean age of this group was 24.6 years. Both studies were approved by the local ethics committee and all speakers gave their written consent to participate.

4.2.2 Equipment, experimental procedure, and stimuli

The general experimental set-up was equal for both experiments and is described in great detail in section 2.2.2 on page 16. The speakers were seated in a chair in front of a computer monitor and completed a baseline as well as three shift phases for both experiments. The corresponding experimental procedures are described in more detail in sections 2.2.3 on page 17 and 3.2.4 on page 46, respectively. During each experimental session, participants' speech was recorded with a neck-worn microphone and fed back via foam-tipped insert earphones. In Experiment 1, participants were asked to produce the close central vowel /ɨ/ whose F2 frequency was perturbed in two opposite directions during the shift phases. In Experiment 2, participants produced the voiceless fricative /sʲ/ whose whole acoustic spectrum was perturbed in two opposite directions during the shift phases. Detailed descriptions of auditory perturbations applied in Experiment 1 and Experiment 2 can be found in sections 2.2.3 and 3.2.4, respectively.

4.2.3 Statistical analyses

All analyses were performed in R (version 3.4.1; R Core Team, 2017) based on respective subsets of the data presented in Chapters 2 and 3 (for details see sections 2.2.5 on page 20 and 3.2.5 on page 49, respectively). In order to compare adaptation magnitude across both experiments, we first applied the RF algorithm to responses produced by participants under downward and upward perturbation using the implementation provided in the *randomForest* package (version 4.6-14). All RF models were computed using normalized parameter values which were obtained by subtracting participants' individual means of each investigated acoustic parameter produced during the baseline phase in the respective experiment and stimulus.

For our investigation, we examined three formant frequencies (F1, F2, and F3) in the case of vowels, and 13 spectral parameters (COG, SD, skewness, Freq_{Mid}, $\text{AmpD}_{\text{Mid-MinLow}}$, $\text{AmpD}_{\text{High-Mid}}$, $\text{Level}_{\text{Low}}$, $\text{Level}_{\text{Mid}}$, $\text{Level}_{\text{High}}$, $\text{LevelD}_{\text{High-Mid}}$, $\text{LevelD}_{\text{Mid-Low}}$, F1, and F2 of the preceding vowel) in the case of fricatives in order to keep the current results consistent with the results reported in the previous chapter.

Subsequently, we were able to correlate the prediction accuracy scores observed for Experiment 1 and Experiment 2 within each participant. Furthermore, for both experiments we examined the evolution of variable importance scores computed for each experimental phase. By doing so, we were able to investigate which of the investigated acoustic parameters were driving the prediction accuracy scores of the RF models. This allowed us to quantify the degree to which speakers were able to develop and maintain common compensatory strategies throughout both experiments.

4.3 Results

We begin this section by summarizing and comparing the adaptation magnitude across the vowel and fricative perturbation experiments. Then, we take a closer look at individual acoustic parameters which were deemed most import for speakers' adaptation strategies in Experiment 1 and Experiment 2.

4.3.1 Adaptation magnitude across both experiments

In this section, we provide prediction accuracy scores computed for Experiment 1 and Experiment 2 for the baseline and all shift phases. To acquire the accuracy scores, we computed overall eight RF models, one for each experimental phase of the respective experiment.

This allowed us to quantify the degree to which the direction of the applied perturbation could be predicted from participants' /ɨ/- and /s^j/-productions. In order to establish an initial prediction accuracy score for each experiment, we ran RF models on the baseline data since no perturbation was delivered during this phase.

For the baseline phase of Experiment 1, the data fed to the RF model consisted of 530 /ɨ/ tokens potentially produced under downward or upward perturbation. For each of the three shift phases, the data included on average 440 /ɨ/ tokens produced under downward and 440 /ɨ/ tokens produced under upward perturbation. The accuracy scores computed for Experiment 1 are summarized in Figure 4.1.

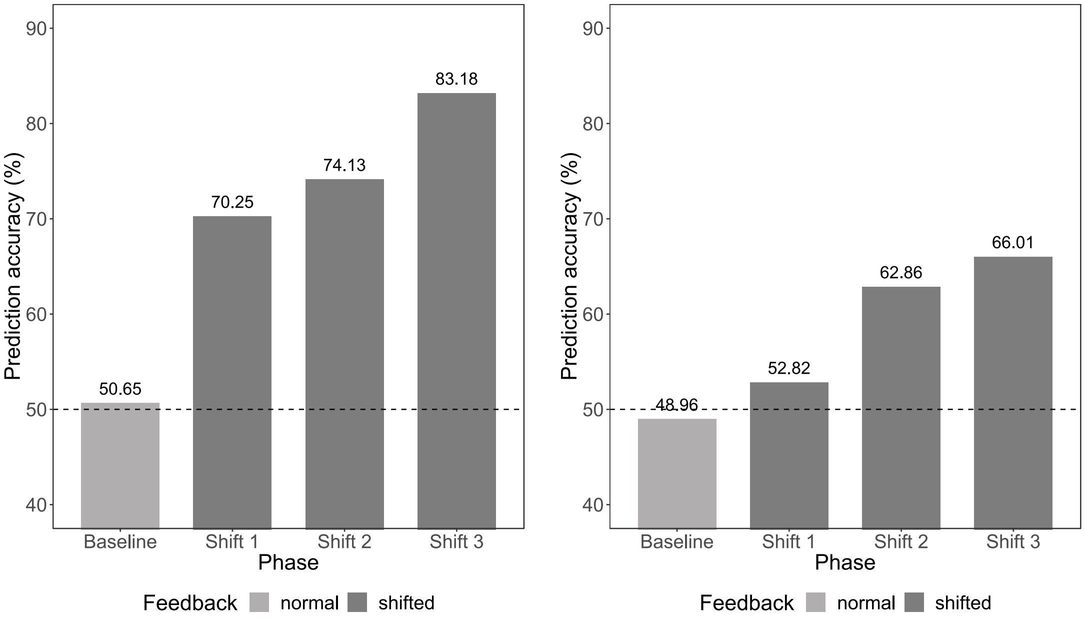

Figure 4.1: Summary of the prediction accuracy scores of the fitted RF models across all experimental phases with regard to the classification task of the applied perturbation direction for Experiment 1 (F2 perturbation; left panel) and Experiment 2 (fricative perturbation; right panel). The dashed line at 50% denotes the chance level.

As expected, the computed RF model was not able to distinguish between /ɨ/ tokens with regard to the (potential) perturbation direction during the baseline phase resulting in an accuracy score of 50.65 percent fluctuating around chance performance. As Experiment 1 progressed, there was a sudden jump in the accuracy score after the baseline phase followed by a steady increase from 70.25 percent in the first shift phase to 83.18 percent in the last shift phase. This pattern of accuracy scores indicates that speakers were able to develop two different compensatory strategies for the production of the target sound /ɨ/ which is consistent with GAMM analyses presented for single formant frequencies in section 2.3.5 on page 30.

For the baseline phase of Experiment 2, the data fed to the RF model consisted of 300 /s^j/ tokens potentially produced under downward or upward perturbation. For each of the three shift phases, the data included on average 250 /s^j/ tokens produced under downward and 250 /s^j/ tokens produced under upward perturbation. The accuracy scores for Experiment 2 are presented in Figure 4.1. These accuracy scores are overall consistent with accuracy scores computed for 19 speakers in section 3.3.6 on page 64.

Eyeballing the pattern of accuracy scores computed for Experiment 2, we arrive at similar conclusions as for Experiment 1. Overall, the accuracy scores increase steadily over the course of the experiment from chance level during the baseline phase to 66.01 percent during the last shift phase. However, the increase in accuracy scores occurs at a slower rate and the final magnitude of the observed accuracy scores is considerably lower compared to Experiment 1 (66.01 vs. 83.18 percent). This discrepancy between accuracy scores for Experiment 1 and Experiment 2 suggests that there was no relation between compensatory behavior in both studies. Since the same speakers completed both experiments, we can exclude any speaker-specific effects which might be responsible for the observed adaptation difference. In general, a quicker and a more complete adaptation to perturbations in vowels in contrast to fricatives is consistent with findings of previous oral-articulatory studies comparing perturbation across both sound types (e.g., McFarland & Baum, 1995).

To correlate adaptation magnitudes observed for each speaker during vowel and fricative perturbation, we calculated individual accuracy scores for the last phase of each experiment since we expected speakers to have achieved the highest degree of adaptation by the end of an experimental session. The data fed to individual RF models consisted of 50 /ɨ/ tokens for Experiment 1 and 30 /s^j/ tokens for Experiment 2 per speaker. Pearson's correlation coefficients revealed a weak, non-significant positive correlation between adaptation magnitudes observed across both experiments ($r = 0.22$, $p > .05$).

As can be seen from Figure 4.2, 13 out of 18 participants achieved an adaptation score of 90 percent or higher during Experiment 1 (f2, f4, f5, f6, f7, f8, f9, f10, f12, f13, f14 m3, and m4). On the other hand, there were only four participants who were able to achieve a score of 90 percent or higher in Experiment 2 (f1, f4, f11, and m4). While there were no speakers with an accuracy score lower than 80 percent in Experiment 1, there were two speakers who remained at chance level performance throughout the whole session in Experiment 2 (f3 and

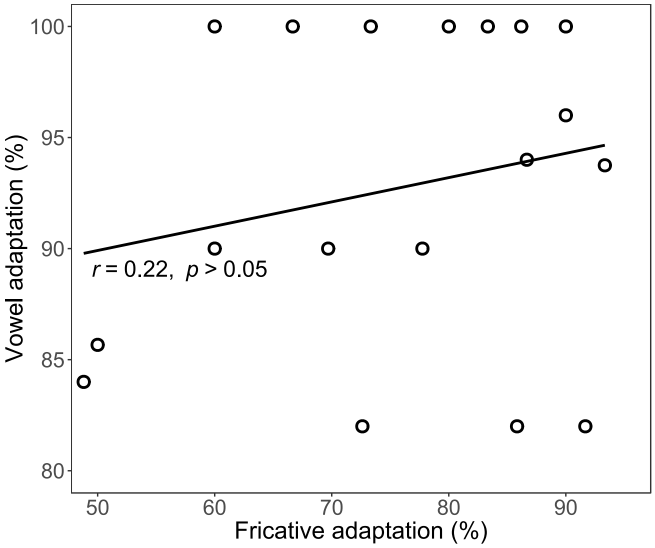

Figure 4.2: Correlation between individual accuracy scores computed for last shift phase of Experiment 1 (vowels) and Experiment 2 (fricatives), respectively.

m2). Furthermore, the accuracy scores observed in Experiment 2 had a considerably higher range compared to Experiment 1 (60 vs. 24 percent).

Some of the speakers who achieved a rather high accuracy score in Experiment 1, performed by at least 30 percent lower in Experiment 2 (f2, f3, f12, f14, and m2). There were only three participants who were able to achieve accuracy scores higher than 90 percent consistently across both experiments (f4, f10, and m4). However, for the remaining 16 speakers the prediction accuracy score they achieved in the first experiment did not predict the accuracy score achieved in the second experiment.

4.3.2 Adaptation strategies in both experiments

In this section, we examine closer the acoustic parameters involved during the adaptation process in Experiment 1 and Experiment 2 in attempt to understand the discrepancy observed in prediction accuracy scores between both experiments. This assessment is performed based on variable importance scores which underlie the RF models presented in the previous section. In Table 4.1, all variable importance scores are summarized for each of the three shift phases of Experiment 1.

While the importance score for all three formants remained close to the decision boundary during the baseline phase, all three parameters were deemed important during the first shift

Table 4.1: Acoustic parameters deemed important to classify whether a token of /ɨ/ was produced under downward or upward perturbation. Results of the variable importance computation performed on the data from the first, second, and third shift phases.

Shift 1		**Shift 2**		**Shift 3**	
Parameter	*Importance*	*Parameter*	*Importance*	*Parameter*	*Importance*
F2	50.61	F2	73.66	F2	125.70
F3	22.13	F3	26.44	F3	41.81
F1	7.59	F1	2.79	F1	10.43
shadowMax*	3.57	shadowMax*	2.27	shadowMax*	1.55

alpha = 0.01

* shadowMax is a “dummy” parameter computed by the Boruta algorithm to determine the importance decision boundary. See section 3.2.7 on page 52 for more details.

phase. However, at this point the importance score for F2 was already considerably higher compared to the scores for F1 and F3. As the experiment progressed, the importance scores for F2 and F3 continued to grow in an equivalent relation while the score for F1 remained relatively low. Overall, the F2 importance score was by far the highest among the three investigated parameters. In other words, the majority of speakers appears to have developed a dominant compensatory strategy in Experiment 1 by adjusting their F2 frequency. Consistent with findings on individual compensatory differences presented in section 2.3.3 on page 26, the increase of the F3 importance score suggests that some speakers adjusted this parameter as well in reaction to the applied F2 perturbation.

In Table 4.2, we summarized the variable importance scores for Experiment 2. In contrast to Experiment 1, none of the investigated acoustic parameters was deemed important during the baseline as well as the first shift phase. By the second shift phase, seven out of 13 investigated parameters were deemed important to classifying /s^{j}/ tokens produced under upward and downward perturbation. However, none of the corresponding importance scores assumed a value which was significantly higher compared to the remaining (unimportant) parameters. In the last shift phase, 10 out of 13 parameters were deemed important but remained, due to overall rather low importance scores, closely spaced. Consequently, many parameters traded their places from the second to the third shift phase and there was no parameter which distinctly separated itself from the larger group of 13 parameters. Put differently, the variable importance scores were characterized by a ceiling effect and remained mostly the same for

Table 4.2: Acoustic parameters deemed important to classify whether a token of /s^j/ was produced under downward or upward perturbation. Results of the variable importance computation performed on the data from the first, second, and third shift phases.

Shift 1		**Shift 2**		**Shift 3**	
Parameter	*Importance*	*Parameter*	*Importance*	*Parameter*	*Importance*
shadowMax*	3.80	Level$_{\text{Low}}$	8.91	SD	8.55
COG	-	F2	5.48	F1	7.60
SD	-	F1	5.32	Skewness	6.09
Skewness	-	Level$_{\text{Mid}}$	5.12	Level$_{\text{Low}}$	5.92
Freq$_{\text{Mid}}$	-	LevelD$_{\text{Mid-Low}}$	4.75	AmpD$_{\text{High-Mid}}$	5.73
AmpD$_{\text{Mid-MinLow}}$	0.93	Skewness	3.13	F2	4.78
AmpD$_{\text{High-Mid}}$	-	COG	2.99	LevelD$_{\text{High-Mid}}$	4.25
Level$_{\text{Low}}$	-	shadowMax*	2.82	COG	4.24
Level$_{\text{Mid}}$	-	SD	2.78	AmpD$_{\text{Mid-MinLow}}$	3.62
Level$_{\text{High}}$	-	Freq$_{\text{Mid}}$	-	Freq$_{\text{Mid}}$	3.38
LevelD$_{\text{High-Mid}}$	-	AmpD$_{\text{Mid-MinLow}}$	-	shadowMax*	2.73
LevelD$_{\text{Mid-Low}}$	-	AmpD$_{\text{High-Mid}}$	-	LevelD$_{\text{Mid-Low}}$	2.63
F1	-	Level$_{\text{High}}$	-	Level$_{\text{Mid}}$	-
F2	-	LevelD$_{\text{High-Mid}}$	-	Level$_{\text{High}}$	-

alpha = 0.01

* shadowMax is a “dummy” parameter computed by the Boruta algorithm to determine the importance decision boundary. See section 3.2.7 on page 52 for more details.

the second half of Experiment 2.

This suggests that speakers did not develop a dominant compensatory strategy in reaction to the applied spectral perturbation. This interpretation is consistent with individual variable importance scores computed for each speaker during the last shift phase of Experiment 2 and summarized in Table 3.10 on page 66.[2] As discussed in section 3.3.7 on page 65, participants' individual compensatory behavior could be rather categorized into four different groups.

Taking a look at individual variable importance scores computed for each speaker during the last shift phase of Experiment 1 (Table 4.3), we see that although a majority if speakers developed a main adaptation strategy focusing on F2 frequency, there was some variability across participants with regard to the acoustic parameter deemed most important to predict

Table 4.3: Overview of the importance decisions for all formant frequencies regarding the classification task of /ɨ/ tokens with regard to the applied perturbation direction. The variable importance computation was performed on speakers' individual data from the last shift phase of Experiment 1. For each speaker, one parameter is marked as being the most important (+).

Parameter	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	m1	m2	m3	m4	Total
F1															+				1
F2		+			+	+		+	+	+	+	+	+	+		+			11
F3	+		+	+			+										+	+	6

the direction of the applied perturbation.

Specifically, while F2 was deemed the most important parameter for 11 out of 18 speakers (f2, f5, f6, f8, f9, f10, f11, f12, f13, f14, and m2), F3 was the most important parameter for six speakers (f1, f3, f4, f7, m3, and m4). The individual importance scores further suggest that F1 frequency was rather unimportant for successful compensation during Experiment 1 as it was deemed important only for one participant (m1). These results are consistent with GAMM analyses on average compensatory behavior presented in section 2.3.5 on page 30.

4.4 Discussion

In this chapter, we compared results from two bidirectional auditory perturbation studies of the Russian close-central unrounded vowel /ɨ/ and the voiceless palatalized fricative /sʲ/. Both experiments were conducted with identical speakers. The results of the first experiment examining F2 perturbation are consistent with our findings reported in Chapter 1. The accuracy scores of the reported RF models suggest that speakers were able to adapt to the applied F2 perturbation with all participants reaching an accuracy score of at least 80 percent. The variable importance scores computed for the investigated acoustic parameters suggested that, on average, speakers developed a compensatory strategy which focused on the F2 frequency from the first shift phase and maintained it till the end of the experiment. Apart from that, the results suggest that F3 was involved in several speakers' adaptation strategies which is consistent with our previous categorization of speakers in symmetrical and asymmetrical adapters in section 2.3.3 on page 26. For more details on the potential relation between F2 and F3 adjustments see section 2.3.4 on page 29.

[2] Apart from speaker m5, all speakers who participated in Experiment 2 were also part of the current comparison.

In Experiment 2, we observed overall smaller accuracy scores compared to Experiment 1. That means that the adaptation degree achieved by speakers during F2 perturbation did not predict their compensatory behavior for spectral shifts in fricatives. This was confirmed by a correlation analysis of individual accuracy scores computed for each experiment. In contrast to Experiment 1, the accuracy scores improved only slowly over the course of Experiment 2. At the same time, the magnitude of variable importance scores remained comparable for most acoustic parameters deemed important for classifying /s^j/ tokens produced under upward and downward perturbation. This observation suggests that speakers were not able to identify a dominant compensatory strategy in order to adapt to the applied perturbation.

The number of participants who achieved an accuracy score of over 90 percent was 13 for the first, but only three for the second experiment. This discrepancy remains unexplainable either by participants' individual acoustic-to-articulatory mappings or feedback preferences. Rather, we believe that this discrepancy arises due to the difference in numbers of articulatory parameters speakers had to adjust to successfully compensate for the perturbations applied during the first and the second experiment. In other words, the articulatory complexity of the respective adaptation task.

While during Experiment 1 participants were merely required to change their tongue position in one dimension to compensate for the F2 shifts, compensatory articulations in Experiment 2 potentially required them to adjust three parameters, namely the constriction location, tongue grooving, and the jaw height (Brunner et al., 2011). The latter articulatory adjustments are less obvious and as a consequence the adaptation process may take longer and less speakers are able to identify the appropriate articulatory adjustments to compensate for the perturbation. This hypothesis is consistent with variable and relatively low variable importance scores observed in Experiment 2.

In summary, the presented line of reasoning suggests that the articulatory-acoustic relation in fricatives (and perhaps more generally in consonants) is less transparent compared to vowels as our participants were less successful in identifying the necessary articulatory adjustments to correct for acoustic errors during perturbation of fricatives than vowels. These findings are consistent with results of previous oral-articulatory perturbation studies (e.g., McFarland & Baum, 1995) and support the idea that the achievement of the speech production goals is dependent on sound-specific articulatory-acoustic relationships (Perkell, 2012).

Chapter 5

Articulatory implementation of compensatory adjustments

5.1 Introduction

Throughout the previous chapters, we recovered several pieces of evidence supporting the hypothesis that speakers might not use designated articulators to compensate for applied auditory perturbations. For instance, in our first experiment we observed that while the majority of investigated speakers compensated for F2 perturbation by adjusting F2 frequency in production, several participants additionally changed their F3 frequency. This finding was supported by GAMM as well as RF analyses (see sections 2.3.5 on page 30 and 4.3.2 on page 81, respectively). We hypothesized that speakers who adapted their F2 and F3 frequencies during the experiment might have been using their lips to assist in the adaptation process (see section 2.3.4 on page 29). Furthermore, in our second experiment we demonstrated that individual speakers adjusted a diverse set of acoustic parameters in reaction to spectral shifts of fricatives. The parameters affected as result of the perturbation were associated, among others, with jaw height, constriction degree and constriction location (see section 3.3.7 on page 65).

Subsequently, by means of a comparison of compensatory strategies observed across vowels and fricatives we came to the conclusion that the more parameters speakers have the necessity to adjust in reaction to perturbation, the more demanding the adaptation process and variable its outcome might become. We observed that while the majority of speakers was able to develop a common compensatory strategy in reaction to the F2 perturbation, the same speakers failed to do so in the case of fricative perturbation which manifested itself in lower variable importance scores in the second experiment compared to the first one (see section 4.3.2 on page 81).

By relying solely on acoustic parameters to assess the degree of adaptation, previous auditory perturbation studies implicitly, and maybe somewhat simplistically, assumed that the observed acoustic compensations are caused by specific articulatory changes. That is, certain articulatory adjustments, like different constriction locations, are expected to cause increases or decreases in specific acoustic frequencies. However, the possibility, or sometimes even necessity, to adjust a higher number of parameters to compensate for applied perturbation implies also that the ways how these adjustments may be implemented articulatorily are likely to vary across individual speakers. This hypothesis is consistent with the results of several previous studies which examined auditory perturbation in conjunction with articulatory data (e.g., Max et al., 2003; Neufeld, 2013). Therefore, it appears plausible to assume that the degree of adaptation across acoustic and articulatory dimensions might differ.

In the current chapter, we want to assess the degree of motor equivalence present during compensatory adjustments in reaction to auditory perturbation. To this end, we conducted an auditory perturbation study of the German sibilant fricative /s/ where we additionally collected participants' articulatory movements using EMA. The acoustic spectrum of the investigated sound was perturbed downwards in one word and kept unchanged in a control word, both embedded within a single stimulus sentence. By including both perturbation conditions into a single stimulus, we effectively increased the number of perturbed /s/-productions per each experimental session with the goal to grant speakers a longer adaptation period. Furthermore, we added a post-shift phase to the experiment to explore any long-term adaptation effects present after the shift phases. One of the goals of this study was to replicate and consolidate our findings presented in Chapter 3. For instance, we wanted to know whether speakers were able to adapt to different auditory perturbation conditions of a single target sound within the same utterance. Ultimately, having acoustic and articulatory data of participants' /s/-productions allowed us to examine how they articulatorily implemented their compensatory strategies.

During the data analysis, we relied on the methodology developed in Chapter 3 but adapted it to current concerns. First, using RF modeling we investigated the evolution of the adaptation process in acoustic and articulatory dimension, respectively. This involved a closer examination of acoustic and articulatory parameters which were driving the prediction accuracy scores of the computed RF models. We rounded out this analysis by comparing /s/-productions across the baseline, shift, and post-shift phases. Then, for individual speakers we aimed at identifying those acoustic parameters which were deemed most important to predict the perturbation condition under which an /s/ token was produced. Subsequently, we were able to uncover corresponding articulatory parameters which were deemed important to implement those acoustic adjustments. The advantage of this analysis is that its findings are not bound by any preconceptions regarding potential co-varying relations between certain articulators like previous studies investigating motor equivalence (e.g., Hughes & Abbs, 1976; Perkell et al., 1993). To break down the adaptation process into single compensatory strategies, we traced acoustic and articulatory changes that occurred over the course of the experiment using GAMM modeling.

5.2 Methods

5.2.1 Participants

Nineteen female, native monolingual speakers of German without reported speech, language, or hearing disorders participated in the study. The mean age of the group was 28.4 years. The study was approved by the ethics committee of the German Linguistic Society (Deutsche Gesellschaft für Sprachwissenschaft, DGfS) and all participants gave their written consent to participate in the experiment.

5.2.2 Equipment

The overall experimental set-up was identical to the one described in section 2.2.2 on page 16 except for the fact that participants' speech was recorded with a Sennheiser ME 62 omnidirectional microphone. In order to perturb the target segment /s/, we employed AUDAPTER's pitch shifting facilities. To identify the onset and offset of the fricative in real-time, AUDAPTER performed an analysis of the speech signal's short-time root-mean-square (RMS) and RMS ratio curves (see Figure 5.1A). When the RMS ratio curve crossed a threshold of 0.03 and kept on rising for the following 15 ms, suggesting the onset of the fricative, pitch shifting was activated. Subsequently, AUDAPTER deactivated pitch shifting when the RMS ratio curve fell below a threshold of 0.06 and kept on falling for the following 15 ms (Figure 5.1B). For further details, see p. 43. The audio signal was digitized and fed back to participants with a sampling rate of 32 kHz. The experimental stimulus contained two /s/ tokens, however, only the first occurrence was tracked during experimental trials.

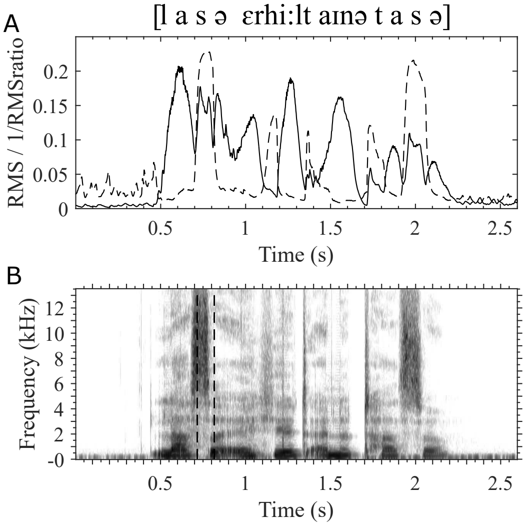

Figure 5.1: Example of a single experimental trial: (A) RMS (solid line) and RMS ratio curve (dashed line) of the speech signal. (A) Fricative onset and offset (dashed lines) tracked by AUDAPTER overlaid over a spectrogram of the speech signal.

To capture articulatory measurements, the AG501 3D electromagnetic articulograph (Carstens Medizinelektronik GmbH, Germany) was used. Articulatory sensors were placed midsagittaly on the tongue tip (TT), tongue mid (TM), tongue back (TB), upper lip (UL), lower lip (LL) and the lower incisors (JAW) with static head reference sensors placed on the gum above the upper incisors, the nasal bone and behind each ear on the mastoid process. The TT sensor was glued approximately 1 cm behind speakers' tongue tip in order to cause as little interference as possible when forming the constriction required for the production of /s/ (see Figure 5.2). Sensor movements were recorded at 1250 Hz.

To identify the neutral position of reference sensors, recordings from a static pose were made while participants bit down on a plastic plate and remained immobile for a few seconds. During post-processing, this reference position was used to correct for any head movements that occurred during the experiment and to translate the articulatory data into a coordinate system centered on the upper incisors, thus allowing for comparison of sensor displacements across all participants.

Before head-correction, to smooth out high-frequency noise, sensor signals were downsampled to 250 Hz and were filtered using Kaiser-Bessel windows with a cut-off frequency

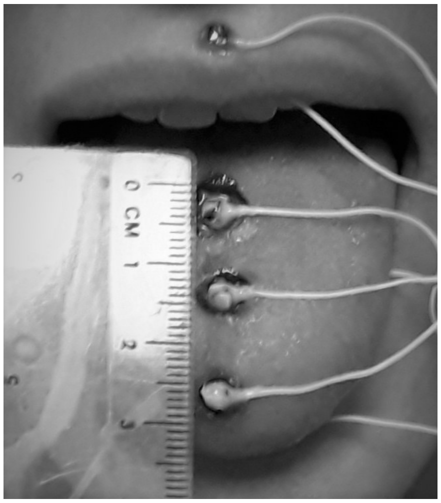

Figure 5.2: Placement of EMA tongue sensors.

of 40 Hz for the TT sensor, 20 Hz for the JAW, UL, LL, TM, and TB sensors, and 5 Hz for the static reference sensors. In all cases, the transition band was 10 Hz. The articulatory data was post-processed using MATLAB routines developed at the Institute of Phonetics and Speech Processing, LMU Munich.

5.2.3 Experimental procedure, speech stimuli, and manipulation

Each experimental session lasted for about 25 minutes and consisted of five experimental phases: baseline, shift 1, shift 2, shift 3, and post-shift phase. Before the actual experiment began, participants completed a familiarization block in order to accustom themselves to speaking with articulatory sensors attached. Focus was made on the production of the fricative /s/ which later served as the target segment for the auditory perturbation. Overall, each participant produced four repetitions of six sentences each containing four /s/ tokens resulting in overall 96 /s/-productions during a familiarization block. The sentences used for familiarization are summarized in Table 5.1. These /s/-productions were not analyzed.

In the course of the experimental session that followed the familiarization block, participants were asked to produce the sentence [lasə (ʔ)ɛɐ̯hiːlt aɪ̯nə tasə] (*Lasse erhielt eine Tasse*; Eng. *Lasse got a mug*) 160 times.

During the baseline phase, no auditory perturbation was applied to participants' speech. After the baseline, three shift phases followed during which the spectral properties of the /s/ sound contained in [lasə] were perturbed in near real-time. The pitch shift was applied

[vanɛsa ʃtiːs haɪ̯sn̩ kɛsl̩ ʊm]	[ʃtʁaʊ̯sə (ʔ)ɛsn̩ nasə nɛsl̩n]
’Vanessa stieß heißen Kessel um.’	’Strauße essen nasse Nesseln.’
’Vanessa knocked over a hot kettle.’	’Ostriches eat wet nettles.’
[klaʊ̯s zaːs ɪm vaɪ̯sn̩ zɛsl̩]	[diː maʊ̯s fʁɪst zyːsn̩ maɪ̯s]
’Klaus saß im weißen Sessel.’	’Die Maus frisst süßen Mais.’
’Klaus sat in a white armchair.’	’The mouse eats sweet corn.’
[laʁs fɛɐ̯gɔs aɪ̯n glaːs vasɐ]	[zaski̯a (ʔ)ʊnt iːnɛs giːsn̩ naʁt͡sɪsn̩]
’Lars vergoss ein Glas Wasser.’	’Saskia und Ines gießen Narzissen.’
’Lars spilled a glass of water.’	’Saskia and Ines water daffodils.’

Table 5.1: Sentences spoken during the familiarization block.

exclusively to the fricative contained in the target word and did not affect any of the other segments of the experimental stimulus. The fricative contained in [tasə] served as a control stimulus. A negative pitch shift of –6.5 semitones was applied to the target fricative such that its spectrum was shifted in its entirety towards lower frequencies resulting in an average decrease of its COG by 2 kHz. As can be seen in Figure 5.3, the applied pitch shift altered the overall spectral balance of the fricative /s/ affecting the amplitudes of the lower (1-3.5 kHz) and higher (6.5-13 kHz) frequency bands in a complementary fashion.

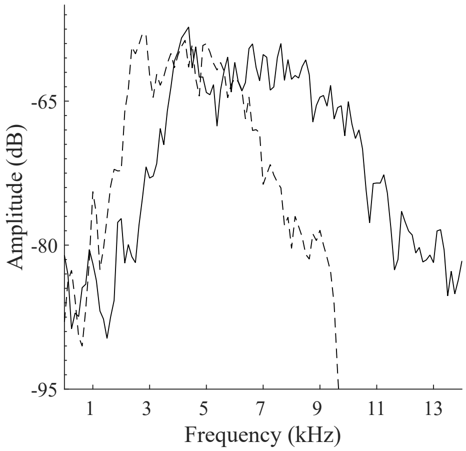

Figure 5.3: Example power spectra of the original (spoken by a participant; solid lines) and perturbed (heard by a participant; dashed lines) fricative segments during the shift phases.

Due to the manipulation, the frication noise of the target sound /s/ resembled that of a more low-frequency fricative. Although the amplitude in the middle frequency band (3.5-

6.5 kHz) was also prone to some minor changes, we think that these modifications were perceptually less salient compared to complementary perturbations that occurred in lower and higher bands (see Table 5.2).

Parameter	**shift effect** ([lasə])
$\text{Level}_{\text{Low}}$ (dB)	7.00 (4.68)
$\text{Level}_{\text{Mid}}$ (dB)	–4.00 (5.30)
$\text{Level}_{\text{High}}$ (dB)	–37.93 (7.67)
COG (Hz)	–2175 (706)
SD (Hz)	–571 (389)

Table 5.2: Mean differences between produced and perceived frication noise with respect to the amplitude of the low ($\text{Level}_{\text{Low}}$), mid ($\text{Level}_{\text{Mid}}$), and high ($\text{Level}_{\text{High}}$) frequency bands as well as the first two spectral moments (COG and SD). Standard deviation is given in parentheses. For more details on the role of the reported acoustic parameters for fricative production see section 3.2.6.

The shift effects were also observable by means of the first (COG) and second (SD) spectral moment. As expected, the downward pitch shift caused a significant decrease in COG and SD. The magnitude of the applied perturbation remained constant throughout the three shift phases and resulted in average shift effects summarized in Table 5.2 based on the data of the first shift phase.

In order to test for potential after-effects of the adaptation, each experimental session ended with a post-shift phase where the perturbation applied during the three shift phases was turned off again.

Participants were asked to produce the experimental stimuli with a neutral intonation and in a moderate speech tempo in order to improve the online tracking of the fricative. To keep the speech amplitude equal across all phases, participants were provided with a real-time graphic display of the microphone gain and asked to maintain a relatively constant volume throughout the experiment.

5.2.4 Data pre-processing

All recordings of 19 participants amounted to 3040 trials. For each trial, the acoustic onsets and offsets of both fricative segments contained in [lasə] and [tasə] were first labeled automatically employing the signal's RMS ratio curve in combination with a set of heuristic rules and then corrected manually using Praat (Boersma & Weenink, 2019). Subsequently, based on the labeled land marks, the experimental measurements were extracted from the acoustic and articulatory signals of each response. To this end, the acoustic signal was first high pass filtered at 600 Hz to exclude any potential influence of accidentally occurring voicing. Then, power spectral densities (PSD) were computed with MATLAB's pwelch() function over the middle 50 percent of the fricative. Finally, the spectral measures were computed for the time averaged power spectra. In the articulatory dimension, positional data of three tongue (TB, TM, TT), both lips (UL, LL), and the jaw sensor were extracted in anterior-posterior (y-axis) and vertical (z-axis) axes. Overall, 11 acoustic and 12 articulatory measurements were extracted per trial to investigate acoustic and articulatory adaptation in participants' speech (for details on used acoustic parameters see section 3.2.6 on page 49).

To determine whether AUDAPTER was able to track the fricative, and therefore to successfully deliver perturbation, each trial was visually inspected and all trials where the interval for pitch shifting did not correspond with the speaker's /s/-production were discarded from the analysis. Furthermore, the data of one speaker had to be excluded completely from the analysis due to recurrent detaching of all tongue sensors in the course of the experimental session. The final data set consisted of 5720 acoustic and 5434 articulatory pairs of perturbed and unperturbed /s/ tokens. The number of articulatory pairs was lower since additional trials had to be discarded due to sensor failure or detachment.

5.2.5 Statistical analyses

All analyses were performed in R (version 3.4.1; R Core Team, 2017). To understand the involvement of acoustic and articulatory dimensions in the adaptation process, we applied the RF algorithm to responses produced by participants under perturbed vs. unperturbed feedback using the implementation provided in the *randomForest* package (version 4.6-14). For more details on RF modeling see section 3.2.7 on page 52. All modeling procedures were applied to normalized parameter values which were obtained by subtracting each participants' mean of each acoustic and articulatory parameter produced during the baseline phase in the respective stimulus word. By means of this normalization, the average values of all investigated parameters (acoustic and articulatory) for [lasə] and [tasə] were set at zero for the baseline phase productions.

Based on the acoustic and articulatory variables deemed most important for the adaptation, we provide an overview of the average compensatory effects by comparing /s/ tokens produced during the baseline phase with tokens produced during the last shift as well as the post-sift phase by means of pairwise t-tests, respectively. To control for the use of multiple comparisons, the p-value was adjusted applying the Bonferroni correction.

To further investigate the relation between acoustic and articulatory dimensions, we computed individual variable importance scores of the investigated acoustic and articulatory parameters for the last shift phase of the experiment. Then, we cross-examined speakers adaptation strategies in the acoustic dimension with how they implemented these changes in the articulatory dimension. Finally, we analyzed the magnitude and temporal evolution of speakers' compensatory responses focusing on individually relevant acoustic and articulatory parameters.

To analyze the adaptation process over time we employed GAMMs. We fitted models with normalized parameter values averaged across all participants and all experimental trials as dependent variables using the *mgcv* package (version 1.8-19) by Wood (2017b). All GAMM models were evaluated, interpreted, and visualized by means of the *itsadug* package (version 2.3) by van Rij et al. (2017). In the model structure, we included random factor smooths with an intercept split for the perturbation condition (perturbed vs. unperturbed) in order to assess individual compensation magnitude differences over the course of the experiment. The model also included a fixed effect which assessed the 'constant' effect of the perturbation direction independently from the temporal variation. By extracting the fitted parameter curves estimated individually for each participant by the GAMM models, we were able to classify participants' compensatory behavior into different groups with regard to acoustic and articulatory changes.

5.3 Results

We begin our review of the results by summarizing adaptation magnitude across articulatory and acoustic dimensions as well as the parameters which speakers adjusted the most in reaction to the applied perturbation. This allows us to compare average /s/ productions across different experimental phases with respect to most relevant acoustic and articulatory parameters. Then, building upon the most important acoustic parameters across all speakers, we attempt to identify the corresponding articulatory strategies employed to achieve the acoustic compensation. Finally, we examine the evolution of the adaptation process for individual speakers across the whole experiment.

5.3.1 Adaptation magnitude across acoustic and articulatory dimensions

In this section, we provide prediction accuracy scores computed based on acoustic and articulatory parameters for all experimental phases. To acquire the accuracy scores, we ran 5 acoustic and 5 articulatory RF models each fed on average with 2285 acoustic or 2175 articulatory /s/ tokens (potentially) produced under perturbed or unperturbed feedback. The accuracy scores computed for the acoustic and articulatory dimensions are summarized in Figure 5.4.

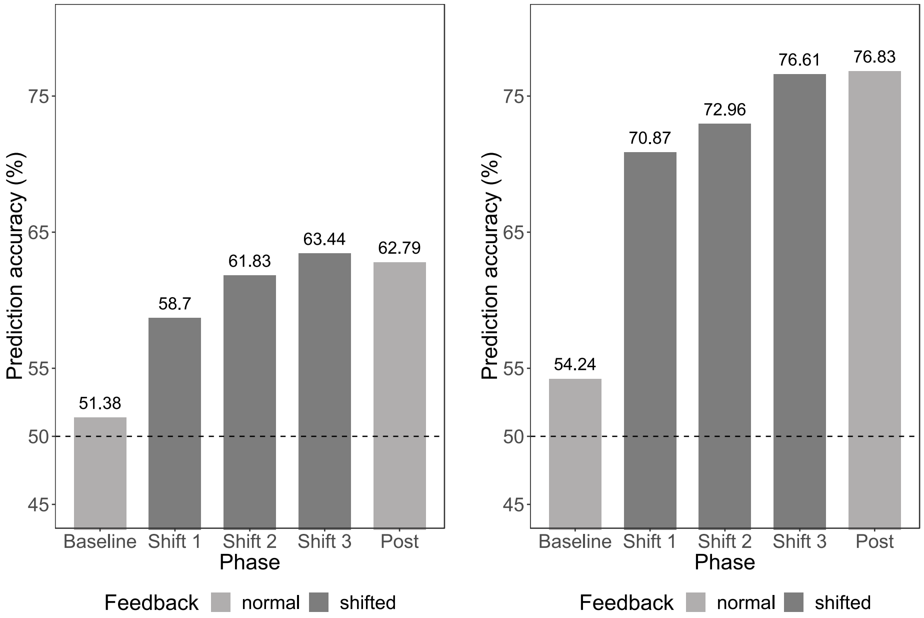

Figure 5.4: Summary of the prediction accuracy scores of acoustic (left panel) and articulatory (right panel) RF models across all experimental phases with regard to the classification task of the perturbation condition. The dashed line at 50% denotes the chance level.

As expected, since no auditory perturbation was applied during the baseline phase the corresponding acoustic RF model was not able to distinguish between /s/ tokens produced in [lasə] and [tasə]. The accuracy score amounted to 51.38 percent fluctuating around chance performance. In the articulatory dimension, however, the baseline accuracy score amounted to 54.24 percent indicating subtle articulatory differences between /s/ tokens produced in [lasə] and [tasə]. In combination, baseline accuracy scores for both dimensions suggest that these articulatory differences did not result in any significant acoustic differences and were probably due to different segmental contexts between the two stimulus words.

In the course of the three shift phases, the prediction accuracy scores based on acoustic parameters increased incrementally to 63.44 percent hinting at progression of the adaptation process in the acoustic dimension. In other words, as the shift phases continued, speakers adjusted the acoustic make-up of the perturbed /s/ tokens in [lasə]. The prediction accuracy based on articulatory parameters increased to 70.87 percent after the baseline phase and reached 76.61 percent in the third shift phase. This suggests that applied perturbation immediately impacted speakers' articulatory movements resulting in different production strategies for perturbed and unperturbed /s/ tokens. However, in contrast to the acoustic dimension, the accuracy scores in the articulatory dimension were overall higher during the shift phases, even when considering the higher baseline accuracy score. Furthermore, the articulatory accuracy scores were characterized by abrupt changes rather than incremental increases. This observation suggests that speakers were actively exploring the articulatory space during the shift phases.

During the post-shift phase, when auditory perturbation was no longer applied, the accuracy score in the acoustic dimension decreased only slightly to 62.79 percent suggesting the presence of long-term adaptation effects. Intriguingly, the prediction accuracy score in the articulatory dimension increased slightly to 76.83 during the post-shift phase. Possibly, this was a result of articulatory overshoot during the production of adapted /s/ tokens after the auditory perturbation ceased for the post-shift phase.

5.3.2 Importance of acoustic and articulatory parameters

Here, we present the results of the variable selection procedure for all 11 acoustic and 12 articulatory parameters underlying the RF models presented in the previous section. All importance decisions made by the Boruta algorithm are summarized in Tables 5.3 and 5.4 for acoustic and articulatory parameters, respectively.

During the baseline phase, one of the acoustic (AmpD$_{\text{Mid-MinLow}}$) and two (UL_Z and TM_Y) of the articulatory parameters were deemed important in classifying whether an /s/ token was produced as part of [lasə] or [tasə]. However, the variable importance scores for most of the listed parameters remained approximately within one unit away from the importance decision boundary which amounted to 2.99 for the acoustic and to 2.55 for the articulatory dimension. Solely the vertical displacement of the upper lip (UL_Z) reached a slightly higher importance score of 6.19. Overall, however, the Boruta procedure was not able to identify predictor parameters of substantial importance to discriminate between /s/ tokens produced in [lasə] or [tasə] during the baseline phase.

Table 5.3: Acoustic parameters deemed important to classify whether an /s/ token was produced under normal or auditorily perturbed feedback. Results of the variable importance computation performed on the data from shift 1, shift 2, shift 3, and post-shift phases.

Shift 1		**Shift 2**		**Shift 3**		**Post-shift**	
Parameter	*Score*	*Parameter*	*Score*	*Parameter*	*Score*	*Parameter*	*Score*
Freq_{Mid}	9.29	Skewness	11.52	Skewness	10.26	$\text{Level}_{\text{Mid}}$	11.44
$\text{LevelD}_{\text{Mid-Low}}$	7.75	Freq_{Mid}	6.87	$\text{LevelD}_{\text{Mid-Low}}$	9.39	$\text{AmpD}_{\text{High-Mid}}$	9.78
$\text{Level}_{\text{Low}}$	7.19	COG	6.47	$\text{AmpD}_{\text{Mid-MinLow}}$	7.57	Skewness	9.48
COG	5.17	$\text{LevelD}_{\text{Mid-Low}}$	6.44	COG	5.24	$\text{Level}_{\text{High}}$	7.14
$\text{LevelD}_{\text{High-Mid}}$	4.77	$\text{Level}_{\text{Low}}$	6.33	$\text{LevelD}_{\text{High-Mid}}$	4.72	$\text{Level}_{\text{Low}}$	6.93
$\text{AmpD}_{\text{High-Mid}}$	4.52	$\text{LevelD}_{\text{High-Mid}}$	5.96	Freq_{Mid}	4.69	SD	6.26
Skewness	3.32	SD	4.42	$\text{AmpD}_{\text{High-Mid}}$	4.34	$\text{LevelD}_{\text{High-Mid}}$	6.15
$\text{Level}_{\text{High}}$	3.21	$\text{Level}_{\text{High}}$	3.87	SD	4.24	COG	5.85
$\text{Level}_{\text{Mid}}$	3.12	$\text{AmpD}_{\text{High-Mid}}$	3.71	$\text{Level}_{\text{Mid}}$	4.24	$\text{LevelD}_{\text{Mid-Low}}$	5.45
shadowMax*	2.54	$\text{Level}_{\text{Mid}}$	3.71	$\text{Level}_{\text{Low}}$	3.68	$\text{AmpD}_{\text{Mid-MinLow}}$	5.33
SD	2.36	shadowMax*	2.50	$\text{Level}_{\text{High}}$	3.14	Freq_{Mid}	2.74
$\text{AmpD}_{\text{Mid-MinLow}}$	-	$\text{AmpD}_{\text{Mid-MinLow}}$	1.75	shadowMax*	2.54	shadowMax*	2.57

alpha = 0.01

* shadowMax is a “dummy” parameter computed by the Boruta algorithm to determine the importance decision boundary. See section 3.2.7 on page 52 for more details.

During the first shift phase, Freq_{Mid}, $\text{LevelD}_{\text{Mid-Low}}$, and $\text{Level}_{\text{Low}}$ were the three most import variables for the classification of /s/ tokens. While the $\text{LevelD}_{\text{Mid-Low}}$ score remained relatively constant across all three shift phases, importance scores for the other two variables decreased in significance during the second and the third shift phase. On the other hand, spectral skewness rapidly became the parameter with the highest importance score for the last two shift phases. COG remained among the five most important parameters across all shift phases but did not increase in its significance in the course of the experiment.

For the duration of all three shift phases, $\text{Level}_{\text{Mid}}$ and $\text{Level}_{\text{High}}$ exhibited importance scores close to the importance decision boundary, however, during the post-shift phase these parameters abruptly became some of the most important along with $\text{AmpD}_{\text{High-Mid}}$. On the other hand, the scores of all parameters deemed important during the last shift phase, decreased during the post-shift phase (e.g., skewness and $\text{LevelD}_{\text{Mid-Low}}$). This suggests that speakers were no longer controlling for these parameters in the same degree after the perturbation was no longer applied.

Table 5.4: Articulatory parameters deemed important to classify whether an /s/ token was produced under normal or auditorily perturbed feedback. Results of the variable importance computation performed on the data from shift 1, shift 2, shift 3, and post-shift phases.

Shift 1		**Shift 2**		**Shift 3**		**Post-shift**	
Parameter	*Score*	*Parameter*	*Score*	*Parameter*	*Score*	*Parameter*	*Score*
JAW_Z	11.08	LL_Z	22.01	LL_Z	25.34	LL_Z	32.10
TT_Z	10.08	TT_Z	19.76	JAW_Z	24.10	TT_Z	21.23
JAW_Y	9.19	TT_Y	19.16	TT_Y	21.23	UL_Z	18.23
LL_Y	8.49	TB_Z	15.74	TT_Z	18.17	TT_Y	17.21
TB_Z	8.41	JAW_Z	13.40	JAW_Y	15.48	TB_Z	15.49
TB_Y	7.87	TM_Y	12.65	LL_Y	15.18	JAW_Z	15.42
TT_Y	7.82	TB_Y	10.46	TB_Z	14.72	JAW_Y	13.89
TM_Y	7.78	JAW_Y	10.00	UL_Z	13.57	LL_Y	12.03
TM_Z	7.37	LL_Y	9.44	TB_Y	13.07	TM_Z	11.90
LL_Z	7.07	UL_Z	8.35	TM_Y	11.54	UL_Y	9.50
UL_Y	5.77	UL_Y	8.13	TM_Z	9.56	TM_Y	9.41
UL_Z	4.62	TM_Z	6.36	UL_Y	9.54	TB_Y	8.92
shadowMax*	2.53	shadowMax*	2.15	shadowMax*	2.65	shadowMax*	2.35

alpha = 0.01

* shadowMax is a "dummy" parameter computed by the Boruta algorithm to determine the importance decision boundary. See section 3.2.7 on page 52 for more details.

In the articulatory dimension, importance scores of all investigated parameters remained closely spaced during the first shift phase. Nonetheless, parameters describing vertical tongue tip and anterior-posterior jaw displacements were deemed most important to classify /s/ tokens with regard to the perturbation condition. During the second shift phase, vertical displacements of the lower lip achieved the highest importance score which grew even further during the third shift phase resulting in a considerate gap separating it from the remaining scores. Furthermore, the anterior-posterior displacements of the tongue dorsum including TT, TM, and TB sensors were characterized by relatively high importance scores during the second shift phase.

The importance scores corresponding to the vertical as well as anterior-posterior tongue tip and jaw displacements remained among the highest over the course of all shift phases. These displacements were likely a consequence of a constriction location change between alveolar and post-alveolar regions. On the other hand, parameters describing the upper lip displacements exhibited very low importance values independent of the shift phase.

5.3.3 Baseline, shift, and post-shift /s/-productions

In this section, we provide a brief overview of participants' /s/ production before any auditory perturbation was applied and compare it to productions observed in the last and the post-shift phase. Based on variable importance scores computed for the acoustic and articulatory parameters in the previous section, we restrain ourselves to reporting average values of the spectral skewness, LevelD$_{\text{Mid-Low}}$, AmpD$_{\text{Mid-MinLow}}$, and COG for the acoustic dimension, as well as LL_Z, TT_Y, JAW_Z, TT_Z, JAW_Y, and TB_Z for the articulatory dimension. For the current overview, we chose variables with highest unique importance scores.

The average values of all selected acoustic and articulatory parameters for all 18 analyzed participants are summarized in Table 5.5 and 5.6, respectively.

Table 5.5: Average values of acoustic parameters with highest unique variable importance scores for the baseline, third shift, and the post-shift phases. The data are split by the perturbation condition (perturbed vs. unperturbed). Standard deviation is given in parentheses.

Parameter	perturbed /s/ ([lasə])			unperturbed /s/ ([tasə])		
	baseline	*shift 3*	*post-shift*	*baseline*	*shift 3*	*post-shift*
Skewness	0.56 (0.62)	0.52 (0.62)	0.53 (0.63)	0.69 (0.61)	0.77 (0.60)*	0.77 (0.70)*
LevelD$_{\text{Mid-Low}}$ (dB)	8.06 (5.62)	8.76 (5.45)*	8.91 (4.96)**	6.08 (4.99)	5.85 (5.77)	6.34 (5.08)
AmpD$_{\text{Mid-MinLow}}$ (dB)	42.63 (6.57)	42.35 (6.13)	42.65 (5.97)	42.05 (6.9)	41.56 (6.82)	41.74 (6.19)
COG (Hz)	7624 (876)	7753 (1045)**	7775 (1005)**	7197 (794)	7184 (908)	7312 (886)**

*p < .05
**p < .025

In the acoustic dimension, we observed significant differences between /s/ tokens spoken in [lasə] during the baseline and the last shift phase for LevelD$_{\text{Mid-Low}}$ (0.70 dB, $t = 2.155$, $p < .05$) and COG (129.10 Hz, $t = 2.270$, $p < .025$). These differences were also significant between the baseline and the post-shift phase suggesting that speakers kept speech adaptations acquired over the course of the shift phases even after the applied perturbation ceased. The difference for LevelD$_{\text{Mid-Low}}$ amounted to 0.85 dB ($t = 2.721$, $p < .025$) and the difference for COG to 150.88 Hz ($t = 2.694$, $p < .025$). On the other hand, for /s/ tokens produced in [tasə] we did not observe significant changes neither in LevelD$_{\text{Mid-Low}}$ nor in COG during the last shift phase. However, there was a significant change in COG of /s/ tokens produced in [tasə] during the post-shift phase (114.82 Hz, $t = 2.298$, $p < .025$). Possibly, it was due to some kind of interaction between production strategies of adapted and non-adapted /s/ tokens which occurred after the perturbation was no longer present during the post-shift phase. The standard deviation for COG was overall higher during shift and post-shift phases, which is a sign of higher degree of production variability for /s/. Furthermore, there were significant changes of spectral skewness of /s/ tokens produced in [tasə] compared across baseline, shift, and post-shift phases. This provides additional support for the idea that the applied perturbation influenced the production of unperturbed /s/ tokens.

Table 5.6: Average values of articulatory parameters with highest unique variable importance scores for the baseline, third shift, and the post-shift phases. The data are split by the perturbation condition (perturbed vs. unperturbed). Standard deviation is given in parentheses.

Parameter	perturbed /s/ ([lasə])			unperturbed /s/ ([tasə])		
	baseline	*shift 3*	*post-shift*	*baseline*	*shift 3*	*post-shift*
LL_Z	–23.42 (2.45)	–23.84 (2.28)**	–23.92 (2.20)**	–22.73 (2.42)	–22.80 (2.29)	–22.81 (2.26)*
TT_Y	12.88 (4.15)	13.01 (3.66)	13.32 (3.47)	13.17 (3.98)	13.61 (3.55)	14.02 (3.27)**
JAW_Z	–19.86 (2.88)	–20.11 (3.35)	–20.08 (3.32)	–19.72 (2.86)	–20.02 (3.18)	–19.98 (3.15)
TT_Z	–4.25 (2.89)	–4.52 (3.20)	–4.28 (3.14)	–4.03 (2.67)	–3.95 (2.83)	–3.66 (2.73)
JAW_Y	1.40 (3.16)	1.81 (3.44)*	2.01 (3.48)**	1.40 (3.24)	1.91 (3.50)**	2.07 (3.52)**
TB_Z	–1.84 (5.11)	–2.40 (4.61)	–2.27 (4.67)	–2.01 (5.18)	–2.31 (4.91)	–2.33 (4.88)

*p < .05
**p < .025

In the articulatory dimension, we observed significant differences with regard to vertical displacements of the lower lip for /s/ tokens produced in [lasə] during the baseline and the third shift phase (–0.42 mm, *t* = –2.915, *p* < .025) as well as during the baseline and the post-shift phase (–0.50 mm, *t* = –3.546, *p* < .025). Additionally, there were significant changes in the anterior-posterior jaw displacements between the baseline and the third shift phase (0.41 mm, *t* = 2.032, *p* < .05), as well as between the baseline and the post-shift phase (0.61 mm, *t* = 3.002, *p* < .025). On the other hand, for /s/ tokens produced in [tasə] we observed changes of anterior-posterior jaw displacements (0.51 mm, *t* = 2.502, *p* < .05 and 0.67 mm, *t* = 3.275, *p* < .025, respectively). This suggests that there was some kind of articulatory interaction between perturbed and unperturbed /s/ tokens during the experiment. Additionally, there were significant changes in /s/ tokens produced in [tasə] with regard to anterior-posterior tongue tip displacements during the post-shift phase.

5.3.4 Relations between acoustic and articulatory changes

In this section, we cross-examine adapting strategies in acoustic and articulatory dimensions for individual speakers. To this end, we first conducted the variable selection procedure of acoustic parameters for the third shift phase of the experiment separately for each of the 18 participants. These results are summarized in Table 5.7.

As can be seen in Table 5.7, the number of acoustic parameters deemed important for the prediction of the perturbation condition varied considerably across participants ranging from one (f1, f5) to nine (f14) variables. There was also one participant for whom no parameter was identified as important (f2). These findings are consistent with results reported for /s^j/ adaptation in section 3.3.7 on page 65.

In overall 21 cases, one of the three most important variables was either LevelD$_{\text{Mid-Low}}$(f3, f5, f6, f7, f11, f16, f17), AmpD$_{\text{Mid-MinLow}}$(f1, f4, f8, f10, f13, f16, f17), or Freq$_{\text{Mid}}$(f3, f6, f9, f12, f14, f15, f17). The only participant whose production could not be classified based on at lest one of these three parameters was speaker f18. For her, the most important parameters were Level$_{\text{Low}}$, Level$_{\text{Mid}}$, and Level$_{\text{High}}$. The spectral moments COG, SD, and skewness were slightly less often among the three individually most important variables.

For each of the groups created based on the parameters LevelD$_{\text{Mid-Low}}$, AmpD$_{\text{Mid-MinLow}}$, and Freq$_{\text{Mid}}$ we compiled the corresponding articulatory variables deemed most relevant to predict the perturbation condition under which an /s/ was spoken. All variable importance decisions for each sub-group are summarized in Tables 5.8, 5.9, and 5.10.

For speakers for whom the most important acoustic parameter was LevelD$_{\text{Mid-Low}}$, the corresponding articulatory adaptations involved anterior-posterior (f3, f5, and f16) and vertical (f3, f5, f6, f11, and f17) tongue sensor displacements. Comparably often, speakers adjusted their the vertical position of the lower lip (f6 and f16; to a lesser degree also f5, f11, and f17). On the other hand, changes of the jaw position were deemed far less often as most important in this sub-group. Furthermore, upper lip displacements occurred quite rarely overall (f17).

For speakers for whom the most important acoustic parameter was AmpD$_{\text{Mid-MinLow}}$, the corresponding articulatory adaptations involved either vertical tongue tip (f4, f10, and f17) or lower lip (f8 and f16) displacements accompanied by changes in the vertical jaw position (f8 and f13). Compared to the sub-group for whom LevelD$_{\text{Mid-Low}}$ was the most important acoustic parameter, speakers of the current group displayed even less changes in the TB (f16) and TM (f13) sensors. On the other hand, there was a substantial number of anterior-posterior

Table 5.7: Overview of the importance decisions for all acoustic measures regarding the classification task of /s/ tokens with regard to the perturbation condition (perturbed vs. unperturbed). The variable importance computation was performed on speakers' individual data from the last shift phase of the experiment. Each parameter is marked either as being among the three most important (++) or important (+) parameters.

Parameter	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	Total ++	Total +
COG (Hz)				++		++	+	+			++	+		+	++	++	+	+	5	6
SD (Hz)								+	++	++		++		++			+	+	4	3
Skewness			++	++			++	++		++		+	++						6	1
Freq$_{\text{Mid}}$(Hz)			++			++			++			++		++	++		++		7	-
AmpD$_{\text{Mid-MinLow}}$ (dB)	++			++				++		++			++	+		++	++	+	7	2
AmpD$_{\text{High-Mid}}$ (dB)														+					-	1
Level$_{\text{Low}}$ (dB)							++	++				++	++	+	+			++	5	2
Level$_{\text{Mid}}$ (dB)														++			+	++	2	1
Level$_{\text{High}}$ (dB)						+									++		+	++	2	2
LevelD$_{\text{High-Mid}}$ (dB)												+		+					-	1
LevelD$_{\text{Mid-Low}}$ (dB)			++		++	++	++			+	++		+	+		++	++		7	3
Total	1	-	3	3	1	4	4	5	2	4	2	6	4	9	4	3	7	6		

Table 5.8: Overview of the importance decisions for articulatory measures regarding the classification task of /s/ tokens with regard to the perturbation condition (perturbed vs. unperturbed). The table contains only data of speakers with $\text{LevelD}_{\text{Mid-Low}}$ being the most important acoustic parameter. The variable importance computation was performed on speakers' individual data from the last shift phase of the experiment. Each parameter is marked either as being among the three most important (++) or important (+) parameters.

Parameter	f3	f5	f6	f7	f11	f16	f17	Total ++	Total +
TB_Y	+	+	+	+	+	++	+	1	6
TB_Z		++	++		+			2	1
TM_Y	+	++	+	+	+	+	+	1	6
TM_Z		+	++		++			2	1
TT_Y	++	+		+		+	+	1	4
TT_Z	++	+	+			+	++	2	3
UL_Y		+	+			+	++	1	3
UL_Z	+	+				+		-	3
LL_Y			+	++	++		+	2	2
LL_Z		+	++		+	++	+	2	3
JAW_Y	+		+	+			+	-	4
JAW_Z				++	+	+	+	1	3
Total	6	9	9	6	7	8	9		

(f1 and f17) and vertical (f1 and f10) displacements of the upper lip.

For speakers for whom the most important acoustic parameter was Freq_{Mid}, the corresponding articulatory adaptations were very often characterized by anterior-posterior tongue mid (f12 and f14) and tongue tip displacements (f3, f12, f14, and f15). Additionally, articulatory changes in this group included vertical lower lip (f6 and f15) and jaw (f9, f14) displacements.

5.3.5 Evolution of acoustic and articulatory adjustments

To investigate compensatory adjustments over the duration of the whole experiment, we computed GAMM models for the most relevant acoustic and articulatory parameters. In the acoustic dimension, models were computed for parameters $\text{LevelD}_{\text{Mid-Low}}$, $\text{AmpD}_{\text{Mid-MinLow}}$, Freq_{Mid}, and $\text{Level}_{\text{Low}}$. Then, to depict the most salient changes for each speaker, we extracted individual participant smooths for the corresponding acoustic parameter which was deemed most important to predict the perturbation condition under which the speaker pro-

Table 5.9: Overview of the importance decisions for articulatory measures regarding the classification task of /s/ tokens with regard to the perturbation condition (perturbed vs. unperturbed). The table contains only data of speakers with AmpD$_{\text{Mid-MinLow}}$ being the most important acoustic parameter. The variable importance computation was performed on speakers' individual data from the last shift phase of the experiment. Each parameter is marked either as being among the three most important (++) or important (+) parameters.

Parameter	f1	f4	f8	f10	f13	f16	f17	Total ++	Total +
TB_Y			+			++	+	1	2
TB_Z								-	-
TM_Y		+			+	+	+	-	4
TM_Z				+	++			1	1
TT_Y		+	+	++		+	+	1	4
TT_Z		++		++	+	+	++	3	2
UL_Y	++					+	++	2	1
UL_Z	++		+	++		+		2	2
LL_Y	+	+	+		++		+	1	4
LL_Z			++	+		++	+	2	2
JAW_Y	+	++		+	+		+	1	4
JAW_Z		+	++		++	+	+	2	3
Total	4	6	6	6	6	8	9		

duced an /s/ token (Figure 5.5). In Table 5.11, we also summarized parameter values for each speaker as fitted by the GAMM models for the third shift phase.

In the first group (Figure 5.5A), which consisted of six speakers for whom LevelD$_{\text{Mid-Low}}$ was determined as the most relevant parameter, for four speakers (f5, f7, f11, and f14) the values of this parameter increased over the course of the experimental session which suggests that they produced perturbed /s/ tokens either with a lower amplitude of the low band (600-5500 Hz) or a higher amplitude of the mid band (5500-11000 Hz) frequencies. In other words, these speakers tried to adapt to the applied perturbation by balancing the acoustic energy in low and mid frequency regions. The remaining two participants of this group (f10, f13) decreased their LevelD$_{\text{Mid-Low}}$ values over the course of the experiment. Remarkably, five out of six speakers of this group (f5, f7, f11, f13, and f14) adjusted the parameter in question not only for the perturbed but also for the unperturbed /s/ tokens. However, the magnitude of changes across both perturbation conditions was different for most speakers (significant for f5, f10, f13, and 14; see Table 5.11).

Three speakers of the second group (Figure 5.5B), produced the sound /s/ with lower

Table 5.10: Overview of the importance decisions for articulatory measures regarding the classification task of /s/ tokens with regard to the perturbation condition (perturbed vs. unperturbed). The table contains only data of speakers with Freq_{Mid} being the most important acoustic parameter. The variable importance computation was performed on speakers' individual data from the last shift phase of the experiment. Each parameter is marked either as being among the three most important (++) or important (+) parameters.

Parameter	f3	f6	f9	f12	f14	f15	f17	Total ++	Total +
TB_Y	+	+	++		+	+	+	1	5
TB_Z		++	+					1	1
TM_Y	+	+	+	++	++	+	+	2	5
TM_Z		++	+			+		1	2
TT_Y	++		+	++	++	++	+	4	2
TT_Z	++	+				+	++	2	2
UL_Y		+			+		++	1	2
UL_Z	+		+					-	2
LL_Y		+	+	+			+	-	4
LL_Z		++		+	+	++	+	2	3
JAW_Y	+	+	++	+	+		+	1	5
JAW_Z			++	+	++	+	+	2	3
Total	6	9	9	6	7	7	9		

$\text{AmpD}_{\text{Mid-MinLow}}$values during the third shift phase compared to the baseline phase. As in the first group, although observed changes occurred in both perturbation conditions, there was mostly a difference in $\text{AmpD}_{\text{Mid-MinLow}}$ values between perturbed and unperturbed /s/ tokens (significant for f1, f8, and f16; see Table 5.11). It appears that the speakers of the second group adapted to the applied perturbation in the same manner as the speakers of the first group, by balancing the acoustic energy in low and mid frequency regions, however adjusting the parameters in question in the opposite way.

The findings for Freq_{Mid} values were less consistent (Figure 5.5C). Particularly, three speakers of the third group produced the /s/ sound with lower Freq_{Mid} values during the third shift phase compared to the baseline (f9, f12, and f17). For f3 and f6, on the other hand, the Freq_{Mid} values increased during the third shift phase. Furthermore, the difference in Freq_{Mid} values across perturbed and unperturbed conditions was significant for f12, f17, and f3 (see Table 5.11). For f15, the observed changes in Freq_{Mid} were rather inconsistent.

A single participant (f18 in Table 5.11) reacted to the applied perturbation by decreasing the amplitude of lower frequencies (600-5500 Hz) across both perturbation conditions.

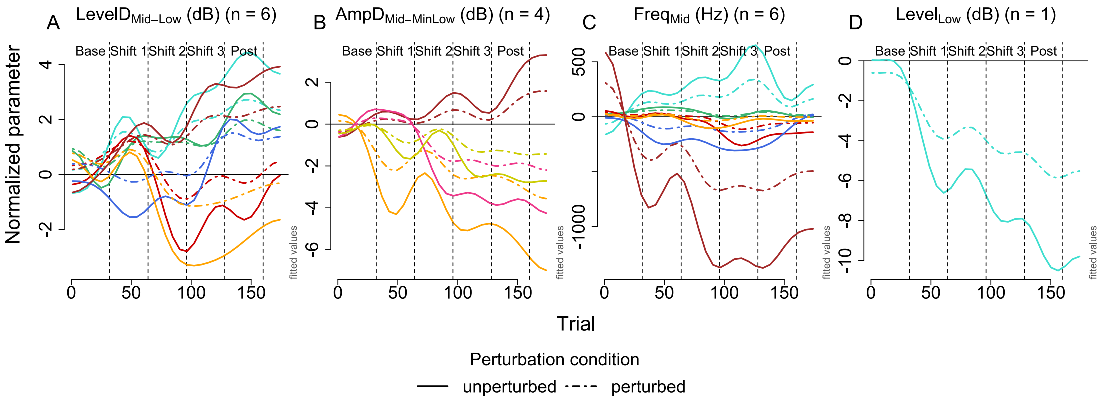

Figure 5.5: Random smooths of the GAMM models fitted for $\text{LevelD}_{\text{Mid-Low}}$, $\text{AmpD}_{\text{Mid-MinLow}}$, Freq_{Mid}, and $\text{Level}_{\text{Low}}$ plotted across baseline and shift phases of the experiment; lines are color-coded for single participants.

Table 5.11: Mean changes in the individually relevant acoustic parameters between the baseline and the last shift phase as fitted by the GAMM models. The parameter values are given in dB ($\text{LevelD}_{\text{Mid-Low}}$, $\text{AmpD}_{\text{Mid-MinLow}}$, and $\text{Level}_{\text{Low}}$) or Hz ($\text{Freq}_{\text{Mid}}$).

ID	**Parameter**	**Perturb.**		**Diff.**	**ID**	**Parameter**	**Perturb.**		**Diff.**
		+	–				+	–	
f1	$\text{AmpD}_{\text{Mid-MinLow}}$	–2.44**	–4.75**	2.31**	f10	$\text{LevelD}_{\text{Mid-Low}}$	–0.05	–1.12**	1.07**
f2	-	-	-	-	f11	$\text{LevelD}_{\text{Mid-Low}}$	1.36**	1.71**	–0.37
f3	Freq_{Mid}	342**	651**	–308**	f12	Freq_{Mid}	–129**	–289**	160**
f4	$\text{AmpD}_{\text{Mid-MinLow}}$	0.20	0.53	–0.34	f13	$\text{LevelD}_{\text{Mid-Low}}$	–0.99**	–2.99**	1.99**
f5	$\text{LevelD}_{\text{Mid-Low}}$	2.24**	3.45**	–1.21**	f14	$\text{LevelD}_{\text{Mid-Low}}$	2.13**	3.25**	–1.12**
f6	Freq_{Mid}	38	43	–6	f15	Freq_{Mid}	4**	–23**	27
f7	$\text{LevelD}_{\text{Mid-Low}}$	1.38**	1.75**	–0.37	f16	$\text{AmpD}_{\text{Mid-MinLow}}$	–1.29**	–2.44**	1.15**
f8	$\text{AmpD}_{\text{Mid-MinLow}}$	–1.95**	–3.76**	1.82**	f17	Freq_{Mid}	–663**	–1359**	697**
f9	Freq_{Mid}	–95	–222	125	f18	$\text{Level}_{\text{Low}}$	–4.57**	–7.90**	3.34**

**$p < .05$

However, the magnitude of this change was significant between perturbed and unperturbed /s/ tokens.

In the articulatory dimension, we computed GAMM models for the parameters LL_Z, JAW_Z, TT_Y, and TT_Z. Then, to depict most salient changes for each speaker, we extracted individual participant smooths for the corresponding articulatory parameter which was deemed most important to predict the perturbation condition under which the speaker

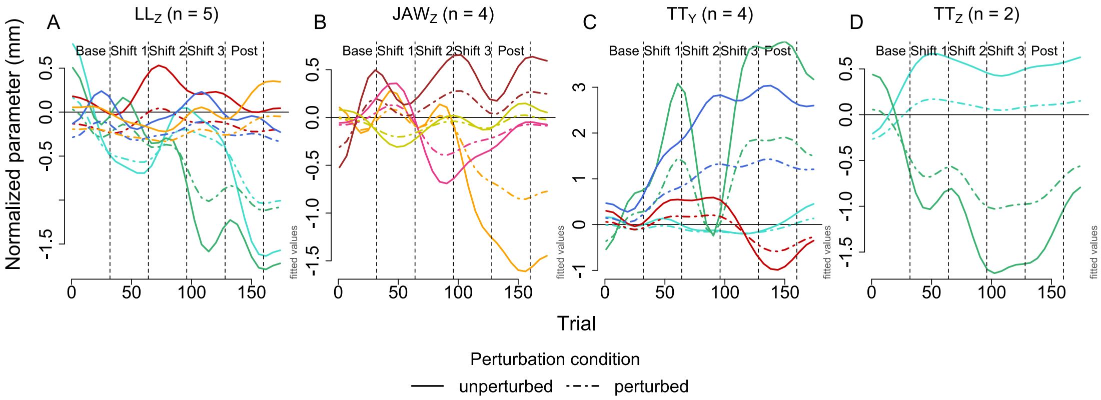

Figure 5.6: Random smooths of the GAMM models fitted for LL_Z, JAW_Z, TT_Y, and TT_Z plotted across baseline and shift phases of the experiment; lines are color-coded for single participants.

produced an /s/ token (Figure 5.6). In Table 5.12, we also summarized parameter values for each speaker as fitted by the GAMM models for the third shift phase.

In the first group (Figure 5.6A), which consisted of five speakers for whom LL_Z was determined as the most relevant parameter, for all speakers (f2, f6, f8, f15, and f16) the values of vertical lip displacements decreased over the course of the experimental session. This suggests that during shift phases these participants produced /s/ tokens with wider lip opening. For three speakers (f6, f8, and f16), the LL_Z values were significantly different between perturbed and unperturbed /s/ tokens, while two speakers (f15 and f16) kept the vertical lip displacement constant in unperturbed /s/ tokens (see Table 5.12).

Two (f7 and f13) out of four speakers of the second group (Figure 5.6B), produced the sound /s/ with lower JAW_Z values during the third shift phase compared to the baseline phase. As in the first group, this suggests that these speakers spoke with a wider mouth opening in reaction to the perturbation. For two speakers (f7 and f9), the vertical jaw displacements were significantly different across both perturbation conditions (see Table 5.12).

For two (f3 and f12) out of four speakers of the third group, for whom TT_Y was deemed the most relevant articulatory parameter, the values of the anterior-posterior tongue tip displacements decreased during the shift phases which suggests that these speakers pushed their tongue tip forwards in reaction to the applied perturbation. In the case of f12, the tongue tip displacements were significantly bigger for unperturbed /s/ tokens. This was also true for the other two speakers (f10 and f18), although the overall magnitude of the anterior-posterior

Table 5.12: Mean changes in the individually relevant acoustic parameters between the baseline and the last shift phase as fitted by the GAMM models. The parameter values are given in mm.

ID	Parameter	Perturb.		Diff.	ID	Parameter	Perturb.		Diff.
		+	–				+	–	
f1	UL_Y	–0.09**	–0.15**	0.06	f10	TT_Y	1.86**	3.90**	–2.04**
f2	LL_Z	–0.39**	–0.34**	–0.04	f11	TM_Z	–0.23**	–0.24**	0.01
f3	TT_Y	–0.17	–0.17	0.00	f12	TT_Y	–0.42**	–0.65**	0.23**
f4	TT_Z	0.08	0.49	–0.41**	f13	JAW_Z	–0.22**	–0.34**	0.12
f5	TB_Z	–1.20**	–0.95**	0.25	f14	JAW_Z	–0.11	–0.12	0.01
f6	LL_Z	–0.88**	–1.32**	–0.43**	f15	LL_Z	–0.23**	–0.01	–0.22
f7	JAW_Z	–0.67**	–1.24**	0.57**	f16	LL_Z	–0.26**	–0.07	–0.19**
f8	LL_Z	–0.12	0.21	–0.32**	f17	TT_Z	–0.98**	–1.63**	0.65
f9	JAW_Z	0.05	0.19	–0.14**	f18	TT_Y	1.37**	2.92**	–1.56**

**p < .05

tongue tip displacements increased in these speakers.

Two speakers (f4 and f17) reacted to the applied perturbations by adjusting their vertical tongue tip displacements (see Table 5.12). The difference in TT_Z values between /s/ tokens produced under the perturbed and unperturbed condition was significant for both speakers, although one speaker increased her values (f4) and the other decreased them (f17).

Three speakers reacted to the applied perturbation by decreasing values either of UL_Y (f1), TB_Z (f5), or TM_Z (f11) parameters. For none of the three speakers, the difference in these changes was significant across both perturbation conditions (see Table 5.12).

5.4 Discussion

Although auditory perturbation studies constitute a rather mature and established line of research, we are aware only of a handful investigations which examine actual articulatory movements resulting from adaptation to the applied perturbation (Max et al., 2003; Feng et al., 2011; Neufeld, 2013; Trudeau-Fisette et al., 2017). Instead, the overwhelming majority of auditory perturbation studies rely on acoustic changes to interpret speakers' compensatory behavior. This approach is implicitly based on the assumption that changes in the acoustic dimension are driven by specific changes in speakers' articulation. However, there are reasons to believe that this is an over-simplification.

To investigate this question further, in the current chapter, we conducted an auditory perturbation experiment of the German fricative /s/ and collected articulatory movements by means of EMA along with the acoustic signal. During the experiment, speakers had to produce 160 repetitions of the sentence *Lasse erhielt eine Tasse* while the /s/ in [lasə] was perturbed but remained unaltered in [tasə].

Overall, the acoustic results presented here are in line with the findings reported in Chapter 3 with respect to the spectral perturbation of /s^j/. Namely, the observed changes were characterized by a high amount of inter-speaker variability with respect to the acoustic parameter which was deemed most relevant to classify the perturbation condition (perturbed vs. unperturbed) under which an /s/ token was produced. Specifically, for about 50 percent of speakers their adjustments resulted in changes of the balance of acoustic energy between low and mid frequency bands. For additional 20 percent of speakers, the peak frequency of the mid frequency band shifted in their production. For the remaining 30 percent of speakers, acoustic adjustments were mostly related to changes in the amplitude of low frequencies, COG, and spectral SD.

The number of modified acoustic parameters varied across speakers and experimental phases. However, the prediction accuracy of acoustic RF models formulated to classify perturbed and unperturbed /s/ tokens increased steadily over the course of the experiment. This suggests that, on average, perturbed and unperturbed /s/ tokens were produced differently by the speakers.

Taking a look at speakers' articulatory movements that occurred during perturbation, we established that speakers immediately reacted to the auditory shifts. Specifically, for 40 percent of speakers articulatory movements most relevant to differentiate between perturbed and unperturbed /s/ tokens were observed either as anterior-posterior tongue tip or vertical lower lip displacements. For further 20 percent of speakers, the most relevant articulatory movements were vertical jaw displacements. For the remaining 40 percent of speakers, most relevant articulatory changes occurred across anterior-posterior, vertical, and horizontal tongue dorsum (TB and TM), upper lip and jaw position adjustments.

Generally, the prediction accuracy of articulatory RF models increased over the course of the experiment although the classification pattern suggests that changes in the articulatory dimension occurred in a more erratic fashion in contrast to the acoustic dimension. In other words, there was no congruency between the magnitude of adaptation effects observed across acoustic and articulatory dimensions such that it appears to be impossible to infer the degree of adaptation in one dimension from the adaptation effects observed in the other dimension.

The observed divergence of RF classification patterns across acoustic and articulatory dimensions provides some support for the hypothesis that speakers' production was characterized by motor equivalent adjustments. On the one hand, our findings suggest that the applied perturbation had an immediate impact on speakers' articulatory strategies causing measurable articulator displacements. On the other hand, the results demonstrate that it took a certain amount of experimental exposure to the applied perturbation before speakers articulatory adjustments affected the acoustic dimension.

There are additional pieces of evidence supporting that there is a mismatch between acoustic and articulatory dimensions. Indeed, even in the baseline phase, before any perturbation was applied, although the acoustic RF model was not able to differentiate between /s/ tokens produced in [lasə] and [tasə], the articulatory RF model already achieved a higher than chance performance. This strongly suggests that articulatory differences might still result in the same acoustic output. After the perturbation ceased for the post-shift phase, we observed long-term adaptation effects in the acoustic dimension since the corresponding prediction accuracy decreased only slightly. On the other hand, the prediction accuracy in the articulatory dimension increased. A plausible explanation for this, rather unexpected, pattern is the idea that speakers enforced their compensatory strategy, they adopted during the shift phases, even stronger once their auditory feedback has changed during the post-shift phase again.

Furthermore, a comparison of individually important parameters across the acoustic and articulatory dimensions demonstrated that a diverse set of articulatory adjustments resulted in comparable changes in the acoustic dimension. For instance, speakers who balanced the amplitude across low and mid frequency bands were able to achieve it by changing either their tongue tip, lower lip, or jaw position. Taken together, the many-to-one relations between articulatory and acoustic parameters, the divergence of RF classification patterns as well as steady improvements in the acoustic dimension and, in contrast, erratic improvements in the articulatory dimension suggest that speakers were exploring their articulatory space in order to adjust the acoustic make-up of their speech towards a certain goal.

In summary, current findings are consistent with the idea discussed in section 3.4 that the compensatory variability observed in the case of fricatives can be attributed to the production stage rather than being a consequence of diminished role of auditory feedback for fricative production. This also provides further support to the hypothesis that speakers ability to successfully adapt to a perturbation is strongly dependent on the articulatory complexity of the posted task as discussed in section 4.1.

Chapter 6

Conclusion

Proponents of most recent speech production theories generally agree on the notion that speech sounds are multidimensional entities defined across acoustic and somatosensory dimensions (e.g., Guenther & Hickok, 2015; Hickok, 2012; Houde & Nagarajan, 2011; Schwartz et al., 2012). However, since the broad evidence to support such an inclusive view stems from rather independent, longstanding streams of research, i.e., oral-articulatory and auditory perturbation, motor equivalence, and neuroimaging studies, which themselves were often conceived based on competing theoretical stances, several questions regarding the relation between acoustic and articulatory dimension of speech sounds remain unanswered. The overall goal of the current dissertation was to better understand some of the functional aspects of the acoustic-articulatory relation of speech sounds. This chapter will briefly review the results of the four previous chapters with respect to the main study questions formulated in the introductory chapter. We will conclude by providing a general discussion of our findings and a brief outlook of potential future work.

In Chapter 2, we conducted a bidirectional F2 perturbation study of the close central unrounded vowel /ɨ/. The F2 values were perturbed in opposite directions depending on the consonant preceding the target vowel (/d/ vs. /g/). In one experimental group, the perturbation was applied in such a way that corresponding compensatory adjustments had to interfere with speakers' initial coarticulatory relation between the experimental stimuli /dɨ/ and /gɨ/. Our findings demonstrate that the compensatory magnitude was neither affected by the high degree of linguopalatal contact in the target vowel nor the coarticulatory relations between the produced sounds. Extending on previous insights by Feng et al. (2011) and Rochet-Capellan and Ostry (2011), these results provide further evidence that congruency between specific auditory and somatosensory targets is not crucial for vowel production even under more restricting articulatory conditions. On the other hand, consistent with previous research by Niziolek and Guenther (2013), our findings suggest that the phonemic category of the perturbed sound may have an impact on speakers' compensatory magnitude.

In Chapter 3, we investigated bidirectional auditory perturbation of the fricative /s^j/ to investigate the hypothesis that somatosensory dimension may prevail over the acoustic for consonants (e.g., Guenther et al., 1998). Our results demonstrate that speakers developed two different compensatory strategies for applied perturbations providing strong support for the relevance of auditory feedback for fricatives production. However, in contrast to the first experiment, we observed lower compensatory magnitude for the perturbation of fricatives. Following the insights from previous oral-articulatory perturbation studies of fricatives (e.g., McFarland et al., 1996), we ascribed this finding to a higher number of acoustic parameters which speakers had to adjust in contrast to vowels in order to successfully adapt to the applied perturbation. By examining changes that appeared across different acoustic parameters, we were able to recover speaker-specific compensatory strategies.

In Chapter 4, we re-examined and compared the results of both perturbation experiments presented in the first two chapters of the dissertation. As expected based on the idea of task-dependent articulatory complexity of adaptation, we observed that individual speakers compensated by different amounts during the two experiments. This, indeed, suggests that the more complex articulatory strategy which is required to produce fricatives may be responsible for the discrepancy observed in comparison with vowels with regard to the compensatory magnitude. In addition to providing an explanation of our experimental data, this hypothesis appears to be a general and self-sufficient explanation for the compensatory variability observed regularly across different oral-articulatory and auditory perturbation studies.

In Chapter 5, we conducted a simultaneous EMA and auditory perturbation study of the fricative /s/ with the goal to assess the magnitude of adaptation across acoustic and articulatory dimensions. In particular, our investigation was guided by previous findings that speakers are able to produce acoustically equivalent goals by employing different articulatory strategies during perturbed and unperturbed speech production (for an overview see Perrier & Fuchs, 2015). Indeed, we observed a discrepancy between adaptation across acoustic and articulatory dimensions with respect to its magnitude and temporal evolution. Additionally, although it appears that there were some general tendencies, there was high degree of inter-individual variability regarding which articulators speakers employed in order to compensate for the applied auditory perturbation. We believe that these findings provide further, more direct evidence for the hypotheses that successful adaptation to a perturbation task is dependent on its articulatory complexity and that congruency between specific acoustic and articulatory goals is not crucial for the production of fricatives.

In summary, the results presented in this dissertation provide strong arguments for the hypothesis that auditory feedback is essential for controlling speech movements even in cases in which speakers have to maintain a high somatosensory contact during their compensatory adjustments. It also appears that the mapping between acoustic and somatosensory targets is quite flexible, while the degree to which this flexibility is constrained is not dependent on specific goals in the articulatory dimension or congruency between acoustic or articulatory goals, but rather on the functional requirement to produce certain speech output. These observations fit with the definition of speech sounds as perceptuo-motor units comprising of articulatory movements which are shaped by perceptual properties (Schwartz et al., 2012). This description of speech sounds also fits with the ideas formulated previously from the perspective of speech acquisition (e.g., Guenther, 1994) and speech development across the life span (e.g., Perkell et al., 1997).

Although our results demonstrate that auditory feedback has equally impacted the production of vowels and fricatives, we observed additional articulatory and auditory (perceptual) requirements that influenced speakers' compensatory abilities to different degrees in the case of vowels and fricatives. In the case of vowels, on the one hand, we observed that a phonemic contrast between the perturbed and unperturbed sound increased the compensatory magnitude. In the case of fricatives, on the other hand, we observed that compensatory magnitude was decreased due to articulatory complexity of the target sound. Following the idea that representations of speech sounds are defined in a multidimensional auditory-articulatory space, it appears plausible that both dimensions can have an influence on speakers' ability to compensate for perturbations.

The current dissertation demonstrated advantages of studying effects of auditory perturbation in combination with articulatory recordings. Having both, acoustic and articulatory signals, enabled us to gain additional insights into acoustic-articulatory relations. Furthermore, employing machine learning models allowed us to deal with highly complex data characterized by a high degree of inter-speaker variability in a more systematic fashion. Thus, we believe that phonetic researchers should seriously consider to use these analysis tools in their research.

Bibliography

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51*(4), 705–723.

Aubin, J., & Ménard, L. (2006). Compensation for a labial perturbation: An acoustic and articulatory study of child and adult french speakers. In *Proceedings of the 7th international seminar on speech production* (pp. 209–216).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boersma, P., & Weenink, D. (2019). *Doing phonetics by computer.* Version 6.0. 46.

Bolla, K. (1981). *A conspectus of Russian speech sounds*. Böhlau Verlag.

Bondarko, L. V. (2005). Phonetic and phonological aspects of the opposition of 'soft' and 'hard' consonants in the modern Russian language. *Speech Communication*, *47*(1), 7–14.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth, Belmont, CA.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(2), 201–251.

Brunner, J., Hoole, P., & Perrier, P. (2011). Adaptation strategies in perturbed /s/. *Clinical Linguistics & Phonetics*, *25*(8), 705–724.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America*, *128*(4), 2033–2048.

Casserly, E. D. (2011). Speaker compensation for local perturbation of fricative acoustic feedback. *The Journal of the Acoustical Society of America*, *129*(4), 2181–2190.

Caudrelier, T., & Rochet-Capellan, A. (2019). Changes in speech production in response to formant perturbations: An overview of two decades of research. *Speech production and perception: Learning and memory. Peter Lang Publisher*, 15–75.

Degenhardt, F., Seifert, S., & Szymczak, S. (2017). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics*, *20*(2), 492–503.

Evers, V., Reetz, H., & Lahiri, A. (1998). Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics*, *26*(4), 345–370.

Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679.

Flege, J. E., Fletcher, S. G., & Homiedan, A. (1988). Compensating for a bite-block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data. *The Journal of the Acoustical Society of America*, *83*(1), 212–228.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, *18*(1), 207–220.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of the Acoustical Society of America*, *84*(1), 115–123.

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. *Phonetica*, *37*(5-6), 306–326.

Gay, T., Lindblom, B., & Lubker, J. (1981). Production of bite-block vowels: Acoustic equivalence by selective compensation. *The Journal of the Acoustical Society of America*, *69*(3), 802–810.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., … Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America*, *128*(5), 3079–3087.

Guenther, F. H. (1994). Skill acquisition, coarticulation, and rate effects in a neural network model of speech production. *The Journal of the Acoustical Society of America*, *95*(5), 2924–2924.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105*(4), 611.

Guenther, F. H., & Hickok, G. (2015). Role of the auditory system in speech production. In *Handbook of Clinical Neurology* (Vol. 129, pp. 161–175). Elsevier.

Hamlet, S. L., & Stone, M. (1976). Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *Journal of Phonetics*, *4*(3), 199–218.

Hamlet, S. L., & Stone, M. (1978). Compensatory alveolar consonant production induced by wearing a dental prosthesis. *Journal of Phonetics*, *6*(3), 227–248.

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, *82*(398), 371–386.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422.

Honda, M., Fujino, A., & Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics*, *30*(3), 281–302.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82.

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica*, *33*(3), 199–221.

Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, *129*(2), 944–954.

Jesus, L. M., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics*, *30*(3), 437–464.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of f0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246–1251.

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, *113*(1), 532–543.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308.

Kelso, J. S., & Tuller, B. (1983). Compensatory articulation under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech, Language, and Hearing Research*, *26*(2), 217–224.

Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 812.

Klein, E., Brunner, J., & Hoole, P. (2018). Which factors can explain individual outcome differences when learning a new articulatory-to-acoustic mapping? In Q. Fang, J. Dang, P. Perrier, J. Wei, L. Wang, & N. Yan (Eds.), *Studies on speech production* (pp. 158–172). Cham: Springer International Publishing.

Klein, E., Brunner, J., & Hoole, P. (2019a). The influence of coarticulatory and phonemic relations on individual compensatory formant production. *The Journal of the Acoustical Society of America*, *146*(2), 1265–1278.

Klein, E., Brunner, J., & Hoole, P. (2019b). The relevance of auditory feedback for consonant production: The case of fricatives. *Journal of Phonetics*, *77*, 100931.

Klein, E., Brunner, J., & Hoole, P. (2019c). Spatial and temporal variability of corrective speech movements as revealed by vowel formants during sensorimotor learning. *Speech production and perception: Learning and memory. Peter Lang Publisher*, 76–107.

Kochetov, A. (2017). Acoustics of Russian voiceless sibilant fricatives. *Journal of the International Phonetic Association*, *47*(3), 321–348.

Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1175–1189.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1–13.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*(27), 9351–9358.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, *7*(2), 147–161.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608.

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068.

Max, L., Wallace, M. E., & Vincent, I. (2003). Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1053–1056).

McFarland, D. H., & Baum, S. R. (1995). Incomplete compensation to articulatory perturbation. *The Journal of the Acoustical Society of America*, *97*(3), 1865–1873.

McFarland, D. H., Baum, S. R., & Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *The Journal of the Acoustical Society of America*, *100*(2), 1093–1104.

Mitsuya, T., MacDonald, E. N., Munhall, K. G., & Purcell, D. W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of America*, *138*(1), 413–424.

Mitsuya, T., Samson, F., Ménard, L., & Munhall, K. G. (2013). Language dependent vowel representation in speech production. *The Journal of the Acoustical Society of America*, *133*(5), 2993–3003.

Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, *125*(1), 384–390.

Nasir, S. M., & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology*, *16*(19), 1918–1923.

Neufeld, C. (2013). *Multimodal targets in speech production: Acoustic, articulatory and dynamic evidence from formant perturbation* (Unpublished master's thesis). University of Toronto.

Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, *33*(29), 12090–12098.

Padgett, J., & Żygis, M. (2007). The evolution of sibilants in Polish and Russian. *Journal of Slavic Linguistics*, *15*(2), 291–324.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382–407.

Perkell, J. S., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech communication*, *22*(2-3), 227–250.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study. *The Journal of the Acoustical Society of America*, *93*(5), 2948–2961.

Perrier, P., & Fuchs, S. (2015). Motor equivalence in speech production. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 225–247). Wiley Online Library.

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977.

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www. R-project. org/.

Rahman, M. S., & Shimamura, T. (2013). A study on amplitude variation of bone conducted speech compared to air conducted speech. In *Proceedings of the 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (pp. 1–5).

Reilly, K. J., & Dougherty, K. E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*, *134*(2), 1314–1323.

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory–motor transformations in speech. *Journal of Neuroscience*, *31*(7), 2657–2662.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333–382.

Savariaux, C., Boë, L.-J., & Perrier, P. (1997). How can the control of the vocal tract limit the speaker's capability to produce the ultimate perceptive objectives of speech? In *Proceedings of the fifth european conference on speech communication and technology.*

Savariaux, C., Perrier, P., & Orliaguet, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, *98*(5), 2428–2442.

Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336–354.

Shadle, C. H. (1990). *Articulatory-acoustic relationships in fricative consonants*. Springer.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113.

Skalozub, L. G. (1963). *Palatogrammy i rentgenogrammy soglasnych fonem russkogo literaturnogo jazyka [Palatograms and x-ray images of Russian consonants].* Izdatel'stvo Kievskogo universiteta.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866.

Trudeau-Fisette, P., Tiede, M., & Ménard, L. (2017). Compensations to auditory feedback perturbations in congenitally blind and sighted speakers: Acoustic and articulatory data. *PloS one*, *12*(7), e0180300.

van Rij, J., Wieling, M., Baayen, R., & van Rijn, D. (2017). *itsadug: Interpreting time series and autocorrelated data using GAMMs.* R package, version 2.3.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306–2319.

Westbury, J. R., Hashi, M., & Lindstrom, M. J. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication*, *26*(3), 203–226.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116.

Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, *62*(4), 1025–1036.

Wood, S. N. (2017a). *Generalized additive models: An introduction with R.* Chapman and Hall/CRC.

Wood, S. N. (2017b). *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation.* R package, version 1.8–19.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/. *The Journal of the Acoustical Society of America*, *123*(6), 4466–4481.

Appendices

Appendix A: Experimental data and code

For each chapter, anonymized, pre-processed and cleaned experimental data as well as associated R code can be downloaded from https://data.mendeley.com/datasets/x8zcpscb8h/1.

Appendix B: Previous publications

The author of this thesis has previously been among the authors of the following works:

Klein, E., Brunner, J., & Hoole, P. (2016). Relation between articulatory and acoustic information in phonemic representations. In C. Draxler & F. Kleber (Eds.), *Proceedings of Phonetik & Phonologie 12* (pp. 90–93). 12–14 October 2016, Ludwig-Maximilians-Universität München, Germany. Retrieved from `https://epub.ub.uni-muenchen.de/29405/`

Klein, E., Brunner, J., & Hoole, P. (2017a). Degree of fexibility of the acoustics to articulation mapping in vowels and consonants. In S. Fuchs, J. Cleland, A. Rochet-Capellan, & O. Maky (Eds.), *Proceedings of the 5th International Winterschool: Speech Production and Perception* (pp. 15–16). 9–13 January 2017, Chorin, Germany.

Klein, E., Brunner, J., & Hoole, P. (2017b). Flexibility of the acoustics-to-articulation mapping: evidence from a bidirectional perturbation study. In *Proceedings of Phonetics and Phonology in Europe 2017*. 12–14 June 2017, Cologne, Germany.

Klein, E., Brunner, J., & Hoole, P. (2017c). How fexible are speakers' internal representations of the articulatory-to-acoustic mapping? In *7th International Conference on Speech Motor Control Groningen: Abstracts* (p. 56). 5–8 July 2017, Groningen, Netherlands. Retrieved from `http://sstp.nl/article/view/29527`

Klein, E., Brunner, J., & Hoole, P. (2017d). Real-time auditory perturbation of fricative spectra. In M. Belz, C. Mooshammer, S. Fuchs, S. Jannedy, O. Rasskazova, & M. Zygis (Eds.), *Proceedings of Phonetik & Phonologie 13* (pp. 105–108). 28–29 September 2017, Humboldt-Universität zu Berlin, Germany. Retrieved from `https://edoc.hu-berlin.de/handle/18452/19531`

Klein, E., Brunner, J., & Hoole, P. (2017e). Which factors can explain individual outcome differences when learning a new articulatory-to-acoustic mapping? In *Proceedings of the 11th International Seminar on Speech Production (ISSP)*. 16–19 October 2017, Tianjin, China.

Klein, E., Brunner, J., & Hoole, P. (2018). Which factors can explain individual outcome differences when learning a new articulatory-to-acoustic mapping? In Q. Fang, J. Dang, P. Perrier, J. Wei, L. Wang, & N. Yan (Eds.), *Studies on Speech Production* (pp. 158–172).

Cham: Springer International Publishing.

Klein, E., Brunner, J., & Hoole, P. (2019a). The infuence of coarticulatory and phonemic relations on individual compensatory formant production. *The Journal of the Acoustical Society of America*, 146 (2), 1265–1278.

Klein, E., Brunner, J., & Hoole, P. (2019b). The relevance of auditory feedback for consonant production: The case of fricatives. *Journal of Phonetics*, 77, 100931.

Klein, E., Brunner, J., & Hoole, P. (2019c). Assessing acoustic and articulatory dimensions of speech motor adaptation with random forests. In *Proceedings of Interspeech* 2019 (pp. 899—903). 15–19 September 2019, Graz, Austria.

Klein, E., Brunner, J., & Hoole, P. (2019d). Spatial and temporal variability of corrective speech movements as revealed by vowel formants during sensorimotor learning. In S. Fuchs, J. Cleland, & A. Rochet-Capellan (Eds.), *Speech Production and Perception: Learning and Memory* (pp. 76–107). Peter Lang Publisher.