Reports in Phonetics, University of the Saarland
Berichte zur Phonetik, Universität des Saarlandes

PHONUS

Edited by: B. Möbius & J. Trouvain

No. 19, July 2017

Phonetik
Fachrichtung Sprachwissenschaft und Sprachtechnologie
Universität des Saarlandes
D-66041 Saarbrücken, Mai 2017

ISSN 0949-1791

VORWORT DER HERAUSGEBER

Die PHONUS-Reihe setzt die Veröffentlichung von Doktorarbeiten von Mitgliedern der Phonetik-Gruppe an der Universität des Saarlandes fort. Der vorliegende Band, PHONUS 19, präsentiert Jeanin Jüglers Doktorarbeit mit dem Titel *The Impact of Training Procedures on the Pronunciation of Stops in Second Language Acquisition: The Case of German Learners of French.* In ihrer Dissertation konzentriert sich Jeanin Jügler auf deutsche Lerner des Französischen und untersucht, inwiefern verschiedene Trainingsmethoden helfen, die Aussprache der Plosive /b p d t g k/ im Französischen zu verbessern. Aufgrund der unterschiedlichen phonetischen Realisierung dieser Laute im Deutschen und Französischen weisen deutsche Lerner des Französischen üblicherweise Probleme mit einer akzeptablen Aussprache in der Fremdsprache auf. Die Arbeit zeigt, dass verschiedene Trainingsmethoden einen Einfluss auf unterschiedliche Aspekte der Produktion von französischen Plosiven haben. Weiterhin zeichnet sich ab, dass die langanhaltende Verbesserung der Aussprache von Plosiven einen komplexen Prozess darstellt, die nicht durch ein einmaliges Ausspracheтraining erreicht werden kann.

Saarbrücken, im Juli 2017 — Bernd Möbius & Jürgen Trouvain

EDITORS' FOREWORD

The PHONUS series continues to publish doctoral theses by members of the Phonetics group at Saarland University. The current volume, PHONUS 19, presents Jeanin Jügler's PhD dissertation, entitled *The Impact of Training Procedures on the Pronunciation of Stops in Second Language Acquisition: The Case of German Learners of French*. In her thesis, Jeanin Jügler focuses on German Learners of French and investigates how different training procedures influence the pronunciation of the French stops /b p d t g k/. German and French mark the distinction between these stops differently, which usually results in foreign accented productions. This thesis shows that different pronunciation training procedures affect different aspects of the production of the French stops. Furthermore, it becomes clear that improving the pronunciation of French stops is a rather complex process that cannot be achieved without recurring pronunciation exercises.

Saarbrücken, July 2017 — Bernd Möbius & Jürgen Trouvain

The Impact of Training Procedures on the Pronunciation of Stops in Second Language Acquisition: The Case of German Learners of French

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultäten
der Universität des Saarlandes

vorgelegt von

Jeanin Jügler

aus Naumburg (Saale)

Dekan:	Univ.-Prof. Dr. Roland Marti
Berichterstatter:	Prof. Dr. Bernd Möbius Dr. Ingmar Steiner
Tag der letzten Prüfungsleistung:	03.03.2017

Short summary

In this thesis the effect of different training procedures on the pronunciation of French stops by German native speakers was investigated. It is known that German learners of French show difficulties in producing French voiced and voiceless stops correctly, which results from interferences of the native and non-native phonological and phonetic systems.

In a first experiment it was shown that the manipulation of the learner's speech on the basis of a native speaker's suprasegmental features, is rated as less accented by native listeners. However, manipulation runs the risk of making the speech sound artificial and needs to be treated with caution. The following experiment examined the effect of exposure to a French native speaker and the learner's manipulated voice on the pronunciation of French stops by German learners of French. There is evidence that both training groups improved their voiceless stops after a production training with one of the provided methods, but learners of the native speaker group improved to the level of the reference speaker. However, this improvement was not perceived by native French listeners. Another experiment investigating the exposure to several native speakers in a perceptual training showed an improvement in the perception of voiceless stops as well as the pronunciation of voiced stops. Although the improvement in perception was maintained after three months, the improvement in the pronunciation of voiced stops was not sustainable.

Kurzzusammenfassung

Diese Arbeit untersuchte den Effekt von verschiedenen Trainingsverfahren auf die Aussprache von französischen Plosiven durch deutsche Muttersprachler. Es is bekannt, dass deutsche Lerner des Französischen Probleme mit der korrekten Aussprache von französischen stimmhaften und stimmlosen Plosiven haben. Diese Probleme lassen sich auf Interferenzen des muttersprachlichen und fremdsprachlichen phonologischen und phonetisch Systems zurückführen.

Ein erstes Experiment zeigte, dass die Manipulation der Stimme des Lerners auf Grundlage von suprasegmentalen Eigenschaften eines Muttersprachlers dazu führte, dass die Lerner mit einem geringeren fremdsprachlichen Akzent wahrgenommen wurden. Allerdings kann die Manipulation der Stimme dazu führen, dass diese ihre Natürlichkeit verliert. Aus diesem Grund sollten Manipulationsprozesse mit Vorsicht behandelt werden. Das folgende Experiment untersuchte den Effekt eines Verfahrens, in dem deutsche Lerner des Französischen entweder mit ihrer eigenen, manipulierten Stimme oder mit Aufnahmen eines französischen Muttersprachlers trainierten. Es konnte gezeigt werden, dass sich beide Gruppen nach dem Ausspracheetraining in der Produktion von stimmlosen Plosiven verbesserten. Jedoch verbesserte sich dich Aussprache von Lernern, die mit Aufnahmen eines Muttersprachlers trainierten so stark, dass sie ein muttersprachliches Niveau aufzeigten. Allerdings konnte diese Verbesserung von französischen Hörern nicht wahrgenommen werden.

In einem weiteren Experiment wurde untersucht, ob sich ein rein perzeptives Training mit mehreren französischen Muttersprachlern positiv auf die Wahrnehmung und ebenfalls auf die Produktion von französischen Plosiven auswirkt. Obwohl die Verbesserung in der Wahrnehmung auch nach drei Monaten nachgewiesen werden konnte, war die Verbesserung in der Produtkion von stimmhaften Plosiven nicht langanhaltend.

Zusammenfassung

In der heutigen Zeit ist es für jeden Schüler in Deutschland verpflichtend mindestens eine Fremdsprache zu belegen. Es gibt jedoch weitere Gründe eine Fremdsprache zu erlernen, beispielsweise um sich mit im Ausland lebenden Familienmitgliedern oder Freunden zu unterhalten, für Reise- oder Arbeitszwecke oder einfach, weil man Freude am Fremdsprachenlernen hat. Die meisten Menschen weisen jedoch einen Akzent auf, wenn sie eine Fremdsprache sprechen. Dieser Akzent entsteht aus Interferenzen des muttersprachlichen und fremdsprachlichen phonologischen und phonetischen Systems (e.g., Best, 1994; Flege, 1995; Kuhl and Iverson, 1995; Escudero, 2005). In dieser Arbeit werden einige der wichtigsten Modelle des Fremdspracherwerbs mit Bezug auf die Wahrnehmung und Produktion von fremdsprachlichen Lauten beschrieben (Perceptual Assimilation Model, (Best, 1994), Speech Learning Model, (Flege, 1995), Second Language Linguistic Perception Model, (Escudero, 2005; van Leussen and Escudero, 2015), Native Language Magnet Model, (Kuhl and Iverson, 1995; Kuhl et al., 2008)). Diese Modelle versuchen zu erklären, welche Prozesse während des Fremdspracherwerbs durchlaufen werden, warum die meisten Lerner einen fremdsprachlichen Akzent aufweisen und weshalb der Erwerb der muttersprachlichen Aussprache so schwierig für Lerner ist.

Im Allgemeinen ist es so, dass Aussprachetraining in den meisten Fremdsprachkursen eher eine geringere Rolle einnimmt, obwohl es allgemein bekannt ist, dass eine korrekte Aussprache einen essentiellen Faktor in der Kommunikation darstellt (Setter and Jenkins, 2005). Aus diesem Grund sollte der Fokus nicht auf das akzentfreie Sprechen gelegt werden, sondern vielmehr auf Verständlichkeit. Leider bietet der Lehrplan kaum Gelegenheit für ein umfassendes Aussprachetraining mit pädagogisch wertvollen Aufgaben. Und zusätzlich haben viele Lehrer keine umfassende phonetische Ausbildung (Baker, 2011), um Schülern die richtige Aussprache zu vermitteln und vorallem hilfreiches Feedback zu geben.

Butler and Winne (1995) zeigen in ihrem Modell des selbstregulierenden Lernens, wie wichtig Feedback für den Lernprozess ist. Das Modell bezieht sich auf das Lernen im Allgemeinen und beschreibt die mentalen Prozesse, die durchlaufen werden, um eine gestellte Aufgabe zu bewältigen. Es wird weiterhin erklärt wie Feedback helfen kann, die Leistung des Lerners zu verbessern. Hattie and Timperley (2007) beschäftigten sich weiterhin mit der Frage nach der Rolle von Feedback innerhalb des Lernprozesses und diskutieren in ihrem Feedback Modell, auf welchen verschiedenen Ebenen des Lernens Feedback wirkt. Einen Schritt weiter gehen Ball and Rahilly (2013), die verschiedene Feedbackmechanismen beschreiben, die dem Lerner während des Sprechens helfen sich selbst zu kontrollieren und zu verbessern.

Eine Studie von Lyster and Ranta (1997) beschäftigte sich mit den verschieden Arten und der Quantität von gegebenem Feedback in fremdsprachlichen Kursen. Die folgenden sechs Feedbackarten wurden dabei identifiziert: 1) Explizite Verbesserung, 2) Umformung (Recast), 3) Nachfrage, 4) Metalinguistisches Feedback, 5) Herauslockung (Elizitation), 6) Wiederholung. Es konnte gezeigt werden, dass in etwa 55 % aller Fälle, in denen Feedback gegeben wurde, eine Umformung (Recast) des Fehlers verwendet wurde. Ein Nachteil an Umformungen ist jedoch, dass sie kaum eine Reaktion der Schüler hervorruft. Eine Umformung wird dadurch definiert, dass sie die korrekte Reformulierung des Fehlers durch den Lehrer beinhaltet, ohne jedoch explizit darauf einzugehen, dass der Schüler einen Fehler gemacht hat. Das hat den weiteren Nachteil, dass der Lerner möglicherweise gar nicht wahrnimmt, dass er einen Fehler produziert und der Lehrer seine Aussage verbessert und umformuliert hat. Im Vergleich reagierten Schüler in 43 % der Fälle, in den Elizitation verwendet wurde. Ein Vorteil von Elizitationen ist, dass Lerner impliziert darauf hingewiesen werden, dass ein Fehler gemacht wurde, den es zu verbessern gilt.

Computer-unterstützte Aussprachetrainingssysteme (computer-assisted pronunciation training (CAPT) systems) können sehr hilfreich für Lerner von Fremdsprachen sein. Sie können ihr eigenes Tempo wählen und so viele Übungen und Wiederholungen zu bestimmten Ausspracheproblemen machen, wie sie möchten. Vor allem Lerner, die zurückhaltender sind und sich nicht viel in den Unterricht einbringen, können ihre Aussprache angstfrei trainieren. Und insofern die automatische Sprach- und Fehlererkennung eine gute Erkennungsrate aufweisen, können diese Systeme sehr wertvolles Feedback geben. Allerdings zeigte eine Evaluierung von 20 kommerziellen und nicht kommerziellen Systemen, dass viele CAPT Systeme immer noch Feedback verwenden, das nur sehr schwer interpretierbar und aus diesem Grund nicht sehr hilfreich für die Lerner ist. Für die Interpretation des Sprachsignals (Oszillogramm) und auch des Spektrogramms von Aufnahmen eines Muttersprachlers und des Lernenden ist ein Hintergrundwissen in akustischer Phonetik von Nöten. Weiterhin vermittelt es den Benutzern der Software die Vorstellung, dass diese Darstellungen für eine korrekte Aussprache übereinstimmen müssten. Allerdings wird dies nicht einmal der Fall sein, wenn der Muttersprachler die gleiche Äußerung ein weiteres Mal aufnimmt. Die Verwendung von Oszillogramm und Spektrogramm hat selbstverständlich den Vorteil, dass sie schnell und einfach abgebildet werden können und zugleich schick und wissenschaftlich aussehen. Es ist jedoch wichtig, dass das verwendete Feedback einfach zu interpretieren ist und dem Lernenden alle wichtigen Informationen mitteilt, um sich verbessern zu können. Zum Beispiel, welcher Fehler gemacht wurde, wo dieser Fehler gefunden werden kann (bei längeren Äußerungen) und wie der Fehler behoben werden kann. Ohne diese Informationen hat der Lerner keine Möglichkeit das Feedback in den richtigen Kontext zu setzen und wird sich nicht verbessern können (Kulhavy, 1977; Hattie and Timperley, 2007).

Eine Art von implizitem Feedback, das auf spielerische Art und Weise den Lernern helfen soll deutsche Vokale korrekt zu produzieren, wurde mit der Trainingssoftware ‘Hüpf’ verfolgt und wird in Kapitel 2 beschrieben. Die Arbeit an der Software ist derzeit noch nicht abgeschlossen und benötigt an einigen Stellen weitere Aufarbeitung. Die Software erhielt bewusst einen spielerischen Charakter, um den Benutzern einerseits ein Programm zu bieten, das Freude beim Lernen bereitet. Andererseits hilft es, Lernern ohne phonetisches Wissen einen Einblick in die Wichtigkeit verschiedener Aspekte des Vokaltrainings zu geben. Das Training bezieht sich derzeit auf das Sprechen von deutschen Einzelwörtern, deren betonter Vokal automatisch analysiert (erster, zweiter Formant und Dauer) und mit Referenzwerten abgeglichen wird, um die Korrektheit bzw. Übereinstimmung zu ermitteln. Die Vokalqualität wird dabei durch einen Frosch visualisiert, der in einem Teich auf Seerosen springen muss. Eine Seerose stellt dabei den Vokal dar, der trainiert werden soll. Wird der Vokal mit den korrekten Formantwerten produziert, landet der Frosch auf der entsprechenden Seerose. Sind die Formantwerte jedoch zu weit von den Referenzwerten entfernt, landet er im Teich. Die Seerosen sind dabei so auf dem Bildschirm dargestellt, dass die x- und y-Achse der horizontalen und vertikalen Zungenposition entsprechen. Die Idee dahinter ist, dass der Lerner durch mehrfache Produktionen der Beispielwörter implizit den Zusammenhang zwischen den Zungenpositionen lernt und dementsprechend anpasst. Eine zusätzliche Darstellung der Dauer des Vokals als Balken um die Seerose zeigt dem Lerner an, ob der Vokal zu kurz oder zu lang produziert wurde. Die Trainingserfolge werden vom Programm gespeichert und können in einer Gesamtstatistik betrachtet werden. Eine Evaluation des Programms bezüglich Benutzerfreundlichkeit und

Nutzen steht jedoch noch aus.

Anschließend wird in Kapitel 4 auf das Lernerkorpus eingegangen, das innerhalb des Projektes IFCASL (Individualized Feedback in Computer-Assisted Second Language Learning) entwickelt wurde und Aufnahmen von etwa 50 deutschen und 50 französischen Sprechern in der Muttersprache und der Fremdsprache (Französisch bzw. Deutsch) beinhaltet. Der Korpus wurde unter Berücksichtigung von in der Literatur diskutierten Ausspracheproblemen sowohl auf Seiten der deutschen Französischlerner und der französischen Deutschlerner erstellt. Die wichtigsten Ausspracheprobleme werden in diesem Kapitel besprochen und mit Beispielen unterlegt. Um zu sehen, welche Ausspracheprobleme sich mit zunehmendem Kenntnisstand der Fremdsprache reduzieren und welche auch noch mit einem fortgeschrittenen Level immer noch vorhanden sind, wurden sowohl Anfänger (A1-B1) als auch fortgeschrittene Lerner (B2-C2) aufgenommen. Um die Sprache der Lerner so variabel wie möglich zu gestalten, wurden verschiedene Sprechkonditionen aufgenommen (einzelne vorgelesene Sätze, Fokusbedingungen, Geschichte "Die drei kleinen Schweinchen"). Der komplette Korpus wurde mit Hilfe eines forced-alignment Tools (Jouvet et al., 2011; Fohr and Mella, 2012) automatisch segmentiert und annotiert. Studentische Hilfskräfte kontrollierten und korrigierten anschließend einen großen Teil des Korpus auf falsch gesetzte Grenzen und Lautabweichungen.

Kapitel 5 beschreibt das erste Experiment der Arbeit, das mit Aufnahmen der Geschichte "Die drei kleinen Schweinchen" des IFCASL Korpus durchgeführt wurde. Das Experiment untersuchte den Einfluss von prosodischen Manipulationen auf den wahrgenommenen fremdsprachlichen Akzent und die Natürlichkeit der Aufnahmen. In der Literatur wurde gezeigt, dass Sprecher in der Fremdsprache (L2) im Allgemeinen langsamer sprechen als in der Muttersprache (L1) (Munro and Derwing, 1995b; Guion et al., 2000 mit Bezug auf Äußerungsdauer und Raupach, 1980; Gut, 2009; Baese-Berk and Morrill, 2015; Trouvain and Möbius, 2014b hinsichtlich Sprech-/Artikulationsgeschwindigkeit). Weiterhin konnten Untersuchungen zu Grundfrequenzunterschieden zeigen, dass Lerner oft Wort- und/oder Phrasenakzente falsch positionieren und ebenfalls einen reduzierten Grundfrequenzumfang in der L2 aufweisen (e.g., Mennen, 1998; Ullakonoja, 2007; Hincks and Edlund, 2009; Busà and Urbani, 2011; Busà and Stella, 2012; Zimmerer et al., 2014). Mit Bezug auf die Wahrnehmung von fremdsprachlichen Akzenten wurde die Rolle der Prosodie jedoch oft außenvorgelassen. Boula de Mareüil and Vieru-Dimulescu (2006) untersuchten den Einfluss des Transfers prosodischer Merkmale (Phonemdauer und Grundfrequenzkontur) des Spanischen auf italienische Aufnahmen und umgekehrt. Ein Perzeptionsexperiment zeigte, dass Prosodie neben den segmentalen Eigenschaften einen wichtigen Faktor in der Einschätzung des Grades der Akzentuierung einnimmt. Allerdings bezog sich diese Manipulation nur auf muttersprachliche Äußerungen. Andere Untersuchungen mit fremdsprachlichen Äußerungen zeigten, dass der Transfer von muttersprachlichen prosodischen Eigenschaften auf die fremdsprachlich akzentuierten Aufnahmen die Wahrnehmung des fremdsprachlichen Akzents reduzierte (e.g., Winters and Grantham O'Brien, 2013; Rognoni and Busà, 2013; Ulbrich and Mennen, 2015).

Aufnahmen von "Die drei kleinen Schweinchen" von französischen Lernern des Deutschen wurden mit Hilfe der in Boula de Mareüil and Vieru-Dimulescu (2006) beschriebenen Technik manipuliert. Jeweils eine deutsche weibliche Sprecherin und ein deutscher männlicher Sprecher wurden ausgewählt und dienten als Referenzsprecher für die Manipulation. Dabei wurde die Dauer jeder Silbe anhand deutscher Vergleichswerte automatisch angepasst. Dies ermöglichte gleichzeitig eine korrekte Übertragung der Grundfrequenzkontur. Phonetische Analysen zu Sprechgeschwindigkeit und Grundfrequenzumfang wurden sowohl für eine Gruppe deutscher Muttersprachler, als auch für die französischen Sprecher in der Fremdsprache und Muttersprache angefertigt. Es konnte gezeigt werden, dass die französischen Sprecher in der Muttersprache schneller sind als die deutschen Sprecher in der Muttersprache. In der L2 reduziert sich jedoch die Sprechgeschwindigkeit der französischen Sprecher um die Hälfte. Ebenso zeigte sich, dass es einen Unterschied im Grundfrequenzumfang zwischen deutschen und französischen Sprecherinnen gibt. Französische Sprecherinnen weisen einen geringeren Umfang in der Muttersprache als deutsche Sprecherinnen auf. Allerdings erhöhen die französischen Sprecherinnen ihren Umfang entsprechend, wenn sie Deutsch sprechen. Einen Unterschied zwischen männlichen Sprechern konnte für beide Sprachen nicht gefunden werden.

In einem Wahrnehmungsexperiment mit deutschen Muttersprachlern konnte gezeigt werden, dass die Prosodietransplantation hilft, die Stärke des wahrgenommenen Akzents to reduzieren. Allerdings resultierte die Manipulation von Sprachdaten in teilweise sehr verzerrten und unnatürlich klingenden Aufnahmen. Dass der Akzent der manipulatierten Produktionen dennoch als relativ stark eingeschätzt wurde, lässt sich einerseits durch die reduzierte Qualität erklären. Andererseits aber auch durch die nicht-muttersprachlichen Produktionen auf segmentaler Ebene, welche nicht manipuliert worden waren. Insgesamt konnte jedoch gezeigt werden, dass manipulierte nicht-muttersprachliche Aufnahmen als weniger akzentuiert durch Muttersprachler eingeschätzt wurden.

In Kapitel 6 wird ein weiteres Experiment beschrieben, das untersuchte, ob manipulierte Aufnahmen der Lerner dabei helfen die Aussprache von französischen stimmhaften und stimmlosen Plosiven von deutschen Sprechern zu verbessern. Beide Sprachen weisen die gleichen Plosivinventare /b p d t g k/ im Phonemsystem auf. Allerdings werden die Laute aus phonetischer Sicht unterschiedlich realisiert. Die deutschen phonologisch stimmhaften Laute werden dabei stimmlos mit einer kurzen Voice Onset Time (VOT) realisiert und die phonologisch stimmlosen Plosive mit einer langen VOT (aspiriert). Im Vergleich werden die stimmhaften französischen Plosive mit einer stimmhaften Verschlussphase ohne VOT nach Verschlusslösung produziert, und die stimmlosen Plosive werden mit einer stimmlosen Verschlussphase und kurzer VOT gebildet. Aufgrund dieser Unterschiede in der Aussprache weisen deutsche Lerner des Französischen Probleme auf die französischen Laute korrekt zu produzieren (e.g., Flege, 1995; Kuhl and Iverson, 1995; Escudero, 2005).

In vorherigen Studien konnte gezeigt werden, dass das Trainieren mit der eigenen manipulierten Stimme hilfreich im Aussprachetraining ist (e.g., Nagano and Ozawa, 1990; Bissiri and Pfitzinger, 2009; De Meo et al., 2016). Insgesamt wurden zwei Konditionen getestet: Training mit der manipulierten Stimme (Manipulationsgruppe) und mit Aufnahmen eines französischen Muttersprachlers (Native Speaker Gruppe). Die Manipulation bezog sich dabei ausschließlich auf die Manipulation der VOT auf Basis von Durchschnittswerten eines französischen Muttersprachlers. Zusätzlich wurde auch eine Kontrollgruppe getestet, die zu keinem Zeitpunkt ein Training erhielt. In einem Vortest wurden zunächst Minimalpaare, eingebettet in kurze französische Sätze mit gleichem vorhergehenden Kontext, von jeweils 10 deutschen Muttersprachlern pro Gruppe aufgenommen. Das Training und der Nachtest wurde am Folgetag durchgeführt, da die Manipulation der Aufnahmen manuell vorgenommen wurden. Die Teilnehmer wurden vor dem Training darauf aufmerksam gemacht, dass es sich um ein Training für Plosive handelt, um den Fokus auf die eingebetteten Minimalpaare zu legen. Das Training bestand aus einem kurzen Imitationstraining mit drei französischen Sätzen pro Laut, entweder mit den manipulierten Aufnahmen oder den Aufnahmen eines Muttersprachlers. Anschließend folgte eine Transferphase, in der sechs französische Sätze pro Laut produziert werden sollten, ohne dass sich manipulierte oder muttersprachliche Aufnahmen vorweg angehört werden konnten.

Hinsichtlich der VOT konnte gezeigt werden, dass sich beide Experimentgruppen in der Produktion der stimmlosen Plosive verbesserten. Allerdings war die Verbesserung der Manipulationsgruppe nicht stark genug, um einen Unterschied zur Kontrollgruppe zu zeigen. Die VOT der Native Speaker Gruppe wurde jedoch so stark reduziert, dass sie statistisch gesehen mit den Referenzwerten des französischen Muttersprachlers übereinstimmten. Eine Verbesserung der stimmhaften Plosive konnte jedoch für keine der beiden

Gruppen gezeigt werden. Ein Grund dafür könnte sein, dass die Reduzierung der VOT für stimmlose Plosive einfacher umzusetzen war, da dieses phonetische Merkmal die deutschen stimmhaften Plosive charakterisiert. Allerdings scheint es schwieriger für die Lernenden zu sein, die VOT für die stimmhaften Plosive noch weiter zu reduzieren.

Kapitel 7 beschreibt ein anschließendes Perzeptionsexperiment, in dem untersucht wurde, ob die Verbesserung der stimmlosen Plosive der Native Speaker Gruppe von französischen Hörern wahrgenommen werden konnte. Aus den Produktionen der drei Gruppen wurden die jeweiligen Wörter herausgeschnitten und den französischen Hörern individuell präsentiert. Diese mussten entscheiden, ob sie die Variante mit dem stimmhaften oder stimmlosen Plosiv gehört hatten. Zusätzlich machten sie Angaben über den Akzentuierungsgrad der Produktion (1 = muttersprachlich, 7 = stark akzentuiert). Die Ergebnisse zeigten, dass die Produktionen der Folgetests häufiger falsch identifiziert wurden. Dieser Abfall von korrekten Identifikationen wurden dabei durch die stimmlosen Aufnahmen der Native Speaker Gruppe bedingt. Einerseits zeigen diese Ergebnisse, dass die Sprecher dieser Gruppe Änderungen in der Aussprache der Plosive im Folgetests vorgenommen haben. Andererseits führte diese Veränderung zu einer verschlechterten Wahrnehmung durch französische Hörer. Diese Ergebnisse können beispielsweise durch das individuelle Verhalten der Sprecher erklärt werden. Einige Sprecher reduzierten die VOT der stimmlosen Plosive so extrem, dass sie mit den produzierten stimmhaften Plosiven übereinfielen. Dies macht die korrekte Identifizierung der Plosive sehr viel schwieriger. Bezüglich des Grades der Akzentuierung kann gesagt werden, dass die Sprecher insgesamt mit einem milden Akzent beurteilt wurden. Ebenso wurden die Aufnahmen des Folgetests beider Experimentgruppen tendenziell mit einem geringeren Akzent wahrgenommen.

In Kapitel 8 wird das letzte Experiment dargestellt, dass den Einfluss eines puren perzeptiven Trainings auf die Perzeption und Produktion von französischen stimmhaften und stimmlosen Plosiven von deutschen Muttersprachlern untersucht. Da das Training Wörter von verschiedenen französischen Sprechern beinhaltet, wird es auch High Variability Training genannt. In einem Vortest wurde die Perzeption mit Hilfe eines Identifikationsexperiments und die Produktion von stimmhaften und stimmlosen Plosiven getestet. Teilnehmer der Experimentgruppe absolvierten anschließend ein dreiwöchiges Training mit zehn Trainingseinheiten. Das Training bestand aus Identifikationstests, in denen die Lerner visuelles Feedback erhielten, ob die getroffene Entscheidung richtig war. Sie hörten dafür jeweils ein französisches Wort und mussten entscheiden, ob sie die stimmhafte oder stimmlose Variante des Wortes gehört hatten. Das Training bezog sich jedoch nur auf Wörter, die mit einem bilabialen Plosiv begannen. Nach drei Wochen wurde ein Folgetest durchgeführt, um zu sehen, ob sich die Lerner nach dem Training verbessert hatten. Weiterhin wurden drei Generalisierungstests durchgeführt, um zu überprüfen, ob sich mögliche Vebesserungen auf andere Artikulationsorte (alveolar, velar) und Wortpositionen (medial, final) übertragen hatten. Nach drei Monaten wurden die Teilnehmer erneut getestet, um zu überprüfen, ob mögliche Effekte langanhaltend sind. Um sichergehen zu können, dass eine etwaige Verbesserung dem Training zuzuschreiben ist, wurden ebenfalls Lerner einer Kontrollgruppe getestet. Die Teilnehmer dieser Gruppe erhielten jedoch kein Training.

Es konnte gezeigt werden, dass die Lerner im Allgemeinen sehr gut darin waren zu erkennen, ob es sich um einen stimmhaften oder stimmlosen Plosiv handelt. Allerdings wurden Wörter mit einem stimmhaften Plosiv öfter richtig erkannt, als Wörter mit einem stimmlosen Plosiv. Die Teilnehmer der Experimentgruppe konnten sich jedoch hinsichtlich der Identifikation von stimmlosen Plosiven verbessern. Diese Verbesserung konnte auch noch nach drei Monaten gezeigt werden. Hinsichtlich der Produktion zeigte sich, dass die Lerner in der Lage waren, die Aussprache von stimmhaften Plosiven zu verbessern, obwohl sie nur die Perzeption trainierten. Allerdings konnte diese Verbesserung nach drei Monaten nicht mehr nachgewiesen werden. Ebenfalls wurden keine Generalisierungen zu anderen Artikulationsorten und Wortpositionen gefunden.

Es lässt sich festhalten, dass die Verbesserung der Aussprache in einer Fremdsprache mit Hilfe von speziell entwickelten Feedbackmethoden kein einfaches Unterfangen darstellt. Ob eine Feedbackmethode wirkungsvoll ist, kann nicht pauschal gesagt werden. Eine Methode kann die Aussprache eines bestimmten Problems verbessern, könnte sich auf ein anderes Problem jedoch negativ auswirken. Anhand der beiden dargestellten auditiven Trainingsmethoden konnte gezeigt werden, dass die Imitation der eigenen Stimme zwar dabei hilft die Aussprache zu verbessern, das Training mit den Aufnahmen eines Muttersprachlers jedoch stärkere Verbesserungen hervorrief. Dies kann jedoch auch damit zusammenhängen, dass die Lerner die eigene Stimme irritierend fanden und sich deshalb nicht 100%ig auf das Training konzentrieren konnten. Allerdings zeigte sich auch, dass diese Art von Training nur eine Verbesserung der französischen stimmlosen Plosive hervorrief. Das zweite durchgeführte Training verbesserte jedoch die Produktion stimmhaften Plosive. Man kann also erkennen, dass die Art des Trainings ebenfalls einen Einfluss auf die Ergebnisse des Aussprachetrainings haben kann. Aus diesem Grund ist es wichtig, unterschiedliche Feedbackmethoden für verschiedene Lautklassen zu testen und weiterzuentwickeln. Für manche Phänomene scheint eine alleinige Methode nicht ausreichend zu sein. Weiterhin hat sich gezeigt, dass eine Verbesserung der Aussprache nicht zwangsweise erhalten bleibt und dass man regelmäßig trainieren muss, um seine Aussprache zu verbessern und die Verbesserungen beizubehalten.

Acknowledgements

First of all, I would like to thank Prof. Dr. Bernd Möbius for giving me chance to work in such an obliging and encouraging research group. Thank you especially for the inspiring discussions and critical comments that helped me tremendously with creating and performing the experiments and finishing this thesis. I would also like to thank Dr. Ingmar Steiner as well as Prof. Dr. Yves Laprie for the valuable feedback and the support throughout the last years.

A special thanks goes to Dr. Frank Zimmerer who always had an open door whenever I encountered a problematic situation or just whenever I needed a sympathetic ear, and to Dr. Jürgen Trouvain who cheered me up especially when the Saarland weather was damping my spirits.

During the last four years, I really enjoyed writing my dissertation in such a great working atmosphere. Thanks to all the former and current colleagues for creating an environment in which I enjoyed working. And thank you to all the colleagues from the IFCASL project who supported me in different situations throughout the last years.

I would like to further thank Dr. Christoph Draxler for the support, suggestions, and implementations of the online experiments in Percy.

Thanks to my proof readers Meryem Ariman, Eva Lasarcyk, and François Conrad who gave useful comments and made suggestions on the chapters of this thesis. Without you, this work would not be in the shape it is today.

A special thanks goes to my family and friends who supported me over the last few years. Last but not least I would like to thank Alexander who supported me throughout this process with all his love, discussions, and motivational speeches.

Contents

Chapter 1

Introduction

1.1 Motivation and Phonetic Research Aims

Learning at least one foreign language has become obligatory for students in school. But there are many other reasons why someone can choose to learn another language. To be able to talk to family or friends abroad, for vacation or work purposes, or simply for fun. However, when learning a foreign language most people retain a foreign accent which results from interferences of the native and non-native phonological and phonetic systems (e.g., Best, 1994; Flege, 1995; Kuhl and Iverson, 1995; Escudero, 2005). Unfortunately, most foreign language classes do not focus on pronunciation training, although it is common knowledge that the correct pronunciation is essential for communication (Setter and Jenkins, 2005). A poor pronunciation will make it harder to be understood by native speakers of the foreign language. For this reason, it is argued that pronunciation training should not focus on being able to speak native-like, but being intelligible. Unfortunately, most teachers do not have time to focus on detailed pronunciation training and include pedagogically valuable exercises. In addition, most teachers do not have the appropriate phonetic training (Baker, 2011) in order to teach the correct pronunciation or give helpful feedback.

That pronunciation training or at least improving in pronunciation is something that learners want, is illustrated in Figure 1.1. A couple of years ago, at the start of this thesis, this inquiry was found on the black board at Saarland University: "Looking for someone (preferably a French native speaker), who can teach me to speak French without an accent in a short time.". Funnily enough, someone else added: "Magician preferred.", which gets to the heart of the problem. Learning the correct pronunciation, and

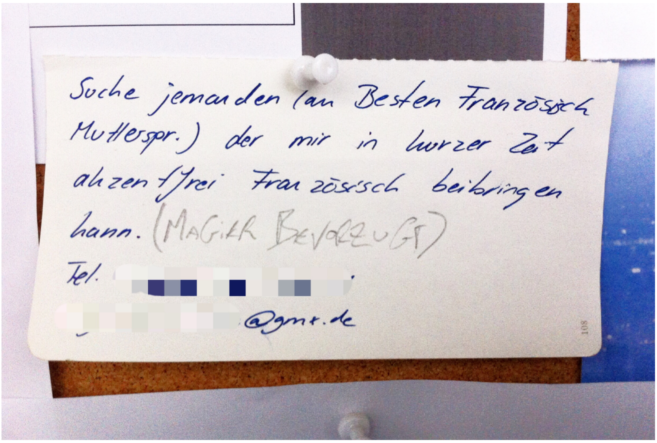


Figure 1.1: Inquiry on the black board at Saarland University: "Looking for someone (preferably a French native speaker), who can teach me to speak French without an accent in a short time." The addition in pencil says: "Magician preferred".

in this example a native-like pronunciation is desired, is not something that can usually be acquired in a short time. A lot of effort has to be put into the training.

For this reason, computer-assisted pronunciation training (CAPT) systems can be helpful for learners to train on specific aspects in pronunciation in the second language. learners can train at a self-defined pace with as many repetitions or exercises as desired. Furthermore, the learner does not have to be self-conscious when practicing pronunciation. And, if the speech recognition and error detection work at a high level, the learner can benefit from valuable feedback in order to improve specific problems in the given foreign language.

Although there are some commercial and non-commercial CAPT systems available, many of the systems do not give appropriate feedback. Some even present the learner with feedback methods, such as oscillograms or spectrograms, which are not only difficult to interpret without any knowledge of acoustic phonetics, but also give the wrong impression that these visual displays need to look exactly the same for the learner and the 'teacher'. Not only do feedback methods included in these systems need to be easily interpretable, but they also need to be explicit for specific pronunciation problems.

This thesis focused on the production of French stops by German learners of French. Both languages share the same set of stops /b p d t g k/. However, the phonetic realization of the sounds differ in German and French. For this reason, German learners of French have difficulties to produce French stops correctly. Two experiments aim at testing the effect of different training procedures on improving the pronunciation of these stops in French. It is important to note that not all exercises and feedback methods are suitable for all pronunciation problems. For this reason it is important to investigate if specific procedures have a positive impact on the pronunciation behavior. Most importantly, it cannot be ruled out that certain training procedures can be counterproductive and may even result in a degradation of pronunciation.

1.2 Structure of the Thesis

In chapter 2, four important models of second language acquisition with respect to perception and production are described and a short comparison highlights the crucial differences of the models. Since feedback is an important factor of influence when learning a second language, the role of feedback in general and in the context of second language learning is discussed. Highlighted problems with regard to pronunciation training in L2 classrooms lead to a discussion of the advantage of computer-assisted pronunciation training systems. Overall, 20 different systems were assessed in terms of the applied feedback methods. Subsequently, the vowel training system 'Hüpf' is discussed in chapter 3, shortly describing the functionalities and training procedure.

Chapter 4 describes the IFCASL (Individualized Feedback in Computer-Assisted Second Language Learning) corpus – a learner corpus of native and non-native speech for the language pair French-German. The corpus was designed to examine which pronunciation problems German learners of French and French learners of German face when speaking in the foreign language, respectively. The main problems are discussed and illustrated with examples from the corpus.

Chapter 5 describes the first of four conducted experiments. The first experiment investigated the effect of transplanting prosodic features (syllable duration and pitch contour) from German native speakers onto accented recordings by French learners of German on perceived foreign accentedness and naturalness.

Since the first experiment showed promising results regarding a reduction of perceived foreign accentedness, chapter 6 discussed the second experiment which investigated two training methods on improving pronunciation for French voiced and voiceless stops by German learners of French: Training with a manipulated version of the learner's own voice and training with recordings of a French native speaker. Subsequently, chapter 7 discussed the perceptual evaluation of the training materials from the second experiment by French native listeners.

Chapter 8 outlined the last experiment that investigated the effect of an auditory training method on the perception and production of French voiced and voiceless stops by German learners of French. Lastly, chapter 9 depicts a summary of the thesis and a discussion of the conducted experiments.

1.3 Preliminary Remarks

Some parts of this thesis are based on revised and extended proceedings articles that are listed in the Curriculum Vitae at the end of this thesis.

The evaluation of some of the listed CAPT systems in section 2.2.3 'Feedback in (Computer-Assisted) Pronunciation Training' was part of an assessment by colleagues working in the IFCASL project. The detailed assessment was published in Trouvain et al. (2016c), which not only evaluated feedback with regard to pronunciation training. It also included an evaluation of available instructions, the quality and type of the exercises, scoring mechanisms as well as the pedagogical structure of the systems. The section 'Feedback in (Computer-Assisted) Pronunciation Training' in this thesis is mostly based on an evaluation of feedback mechanisms used in an extended set of 20 commercial and uncommercial CAPT systems.

Chapter 4 'IFCASL Corpus' is based on the three papers by Fauth et al. (2014), Trouvain et al. (2013) and Trouvain et al. (2016a) which describe the corpus that was developed and recorded within the IFCASL (Individualized Feedback for Computer-Assisted Spoken Language Learning) project. The papers were collaborative work of the colleagues of LORIA, Nancy, France and Saarland University, Saarbrücken, Germany. Additional examples and references were included.

Chapter 5 'Experiment I: Prosody Transplantation' is based on the conference paper by Jügler et al. (2016). It was extended by an acoustic analysis of speaking rate and pitch range for French native speakers as well as a more detailed analysis of the perception experiment.

The following chapter 6 'Experiment II: Exposure to a Native Speaker and Modified Voice' is based on the proceedings article by Jügler and Möbius (2015) which was extended by a more thorough analysis. And chapter 8 'Experiment IV: High Variability Training' is based on the conference article by Jügler et al. (2015) which reported preliminary results for six participants. Within the scope of this dissertation, ten participants per group were analyzed and the acoustic analysis was carried out to a greater extent.

Chapter 2

Theoretical Background

2.1 Models of Second Language Learning

This section aims at giving a short overview of some of the most important models of second language acquisition with respect to the perception and/or production of (novel) foreign sounds. The following models try to explain why most speakers of a second language retain a foreign accent and why it is difficult to become more native-like. A summary of these models is helpful in order to interpret and explain the results of the conducted experiments testing the effect of different training procedures on the pronunciation of French stops by German learners of French. Finally, a short comparison of the systems concludes the section.

2.1.1 Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) (Best, 1994, 1995) aims at explaining non-native speech perception and is based on the perception of articulatory gestures of the speech sounds. Depending on the gestural similarity or difference, non-native sound contrasts should be more easy or more difficult to perceive. Six patterns describe in which way a foreign sound contrast is assimilated to native sounds:

1. *Two Category Assimilation* (TC Type): The gestures of two sounds of a non-native contrast are similar to two different native phones.
   E.g., voiced and voiceless French stops /b p d t g k/ vs. voiced and voiceless German stops /b p d t g k/ which are phonetically similar but not identical.

2. *Single Category Assimilation* (SC Type): Two sounds of a non-native contrast are assimilated to one phone in the native language.
   E.g., /l/-/r/ contrast for Japanese natives.

3. *Category Goodness Difference* (CG Type): Two sounds of a non-native contrast are assimilated to one phone in the native language. However, one non-native sound is more similar to the native sound than the other sound.
   E.g., English /ɛ/-/æ/ contrast for German natives. English /ɛ/ is more similar to German /ɛ/ than English /æ/.

4. *Both Uncategorizable* (UU Type): Both non-native sounds are not considered to fall within any native category.
   E.g., French nasal vowel contrast /ã/-/ɛ̃/-/õ/ for German natives.

5. *Uncategorized vs Categorized* (UC Type): While one sound of a non-native sound contrast is assimilated to a non-native category, the other is not.
   E.g., French oral vs. nasal vowel contrast /ɛ/-/ɛ̃/ for German learners of French. The nasal sound does not assimilate to any native category.

6. *Non-Assimilable* (NA Type): Non-native sounds are too dissimilar from any native phone and are not perceived as speech sounds.
   E.g., Xhosa clicks for any language that does not have clicks in its phoneme inventory.

Learners should not show any problems in distinguishing between sounds of the TC type, since they are assimilated to two individual phones in the native language. Learners are predicted to be still considerably good at discriminating sounds of the CG type. Although these sounds are assimilated to one phone in the native language, learners are still able to perceive a difference between both foreign sounds, since one non-native sound is more similar to the native category than the other sound. In the case of the NA type, discrimination depends on how similar the sounds are to each other, even when they are not perceived as speech sounds. With respect to the SC type, discrimination between the two non-native speech sounds is considered to be difficult, since both non-native sounds are assigned to one native category. Furthermore, in terms of the UU type, discrimination can be either good or poor, depending on the similarity of the two non-native sounds to be compared. In the context of the UC type, discrimination is expected to be very good.

PAM argues that with exposure to the L2 and increasing linguistic experience, learners will be able to perceive a non-native contrast that was not distinguishable before, which is enabled by splitting L1 categories.

2.1.2 Speech Learning Model (SLM)

The objective of the Speech Learning Model (SLM) (Flege, 1995) is to establish to what extent individuals are able to learn the foreign production and perception of non-native sounds. According to SLM, the difficulty of acquiring non-native sounds can be predicted by the perceptual similarity between L1 and L2 sounds. The SLM proposes three sound categories which differ in perceptual similarity of non-native sounds to the L1 sound and, hence, the difficulty in acquiring these sounds:

1) *identical*
   Non-native sounds that are identical to sounds of the native language will be perceived as the appropriate counterparts of the native language. Therefore, identical sounds are easy to acquire.
   E.g., French nasal consonants /m n/ by German natives.

2) *similar*
   Acoustic differences between similar native and non-native sounds are perceived, however, the perceptual difference is not distinct enough. The L2 sound will be replaced by the L1 sound that is closest to the non-native sound. Establishing new perceptual categories is very difficult for the learner and a foreign accent is likely to result.
   E.g., voiced and voiceless French stops /b p d t g k/ vs. voiced and voiceless German stops /b p d t g k/ which are phonetically similar but not identical.

3) *new*
   New non-native sounds are easily differentiated from any native sounds since there is no equivalent available in the L1 phonological system. New perceptual categories have to be established and it is assumed that learners will easily acquire a correct pronunciation.
   E.g., /h/ for French natives.

Flege (1995) described four postulates as well as seven hypotheses which form SLM. The *first postulate* states that the mechanisms and processes used in order to acquire the native language can be applied to the acquisition of foreign languages throughout life, meaning that there is no age limit for learning new languages. The *second postulate* argues that phonetic categories describe language-specific characteristics of speech sounds and influence the way speech sounds are perceived and processed. The *third postulate* says that phonetic categories of the native language are not fixed and can evolve over time, influenced by any non-native sounds. This means, that with increasing linguistic experience, category boundaries can shift. And lastly, the *forth* *postulate* argues that bilingual speakers try to keep phonetic categories of the L1 and L2 separate in order to make clear distinctions between L1 and L2 sounds.

Furthermore, the following hypotheses were made: Native and non-native sounds are perceptually discriminated by means of phonetic features in order to identify phonetic categories, rather than abstract phonemic features. This means, that perception is based on acoustic characteristics of the speech sounds and not on higher-level distinctive features (*H1*). Also, a new category for an L2 sound can be established when the learner perceives acoustic significant differences between an L1 sound and the L2 sound (*H2*). Which means, that the greater the phonetic difference between the native and non-native sound, the more likely it is that sounds will be perceptually discriminated (*H3*). In other words, the establishment of a new category of the L2 sound may be blocked, if perceptually close L1 and L2 sounds fall into the same category of the native language due to their phonetic similarity (*H5*). However, the later a learner acquires a language, the more difficult it is to perceive the phonetic differences between sounds of the L1 and L2, and between several L2 sounds that fall in the same category in the L1 (*H4*). Differences may be found between categories for L2 sounds between monolinguals and bilinguals, which can have two reasons. First, bilinguals try to maintain and distance the different categories from each other. Second, different features or feature weights might play a role in the establishment of the categories (*H6*). Finally, based on the phonetic category specifications, the L2 sound will be articulated accordingly. This means that when an L2 sound is associated with a category of the L1, it will be produced as the corresponding sound in the native language. However, if a new category is established for the L2 sound (e.g., with increasing linguistic experience) that matches the category of a native speaker of that language, the production of the sound will change accordingly to a (more) native-like production (H7).

2.1.3 Second Language Linguistic Perception Model (L2LP)

The Second Language Linguistic Perception model (L2LP) (Escudero, 2005; van Leussen and Escudero, 2015) tries to explain the acquisition of L2 speech perception. The origin of the L2LP model is based on the assumption that learners of an L2 are influenced by phonological rules and the phonological system of their native language (L1) when learners perceive and produce sounds of the L2. However, with increasing knowledge and exposure to the L2, the perception shifts from a full copying mechanism (complete transfer of the native phonological system and its rules) to a perception that is adjusted to the L2.

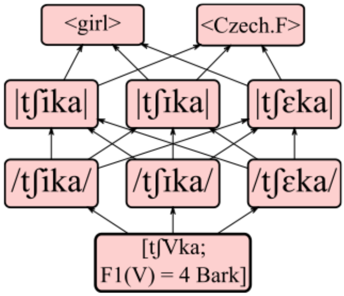


Figure 2.1: Example of different pathways for the perception procedure within the L2LP (van Leussen and Escudero, 2015: p. 4).

The L2LP model consists of four levels: the *acoustic* level, which represents the acoustic information that is perceived, the *phonetic* level, in which the acoustical information is processed and recognized as speech sounds (or possible variants) based on the L1, the *phonemic* level, representing the canonical forms of larger units, like morphemes or words, and the *lexical* level, in which the phonemic forms are connected to lexical forms. van Leussen and Escudero (2015) illustrate the perception process of the model by means of an example of vowel perception (see Figure 2.1).

The example in Figure 2.1 illustrates a Spanish vowel minimal pair contrast (*chica* (girl) vs. *checa* (Czech female)). By varying the frequency of the first formant of the vowel and setting it to a value of 4 Bark, the vowel can be perceived as either one of the two words. The example is based on Dutch learners of Spanish. Since the Dutch phoneme system includes the three front vowels /i/, /ɪ/, and /ɛ/, three phonetic representations are activated (full copying), which connect to three phonemic forms, that can be associated with either one of the two Spanish lexical items. As can be seen in Figure 2.1, any connection from the acoustic to the lexical form is possible within the perception process. The optimal pathway[1] guides the perception from the acoustic to the lexical level. Without any or little knowledge of the L2, this pathway is strongly influenced by the L1 (see Figure 2.2, left

[1] The optimal pathway is defined as the pathway with the 'strongest weakest' connection of all possible pathways. This argumentation is based on Optimal Theory (OT). Following OT, there are different possibilities to realize any linguistic expression. All realizations that are not allowed in (or are constrained by) the given grammar of a language, will be excluded (or marked as violating the grammar). The realization without any constraints or violations is considered to be the optimal candidate. If no optimal candidate without any constraints exists, the realization with the least violations is chosen (cf. Prince and Smolensky, 1993).

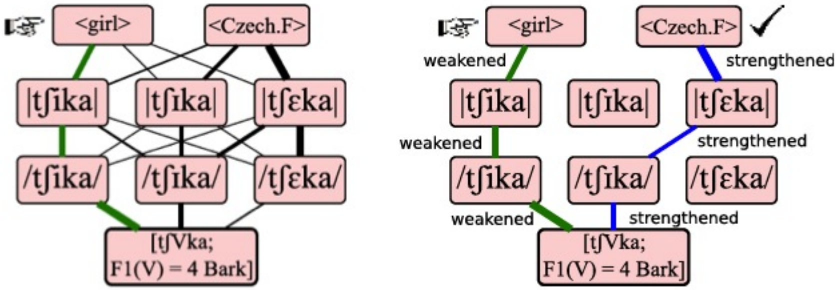


Figure 2.2: Example of finding the optimal path. The thicker the lines the stronger the connection between the different levels. Without any or much linguistic experience, the optimal path is illustrated in the left image which is influenced by the L1. The right image illustrates the path as it is expected in the L2. This path is strengthened with increasing linguistic knowledge (van Leussen and Escudero, 2015: pp. 5-6).

image). However, with increasing knowledge and exposure to the L2, the optimal pathway is reevaluated and changed to a perception that is more appropriate for L2 perception (see Figure 2.2, right image). van Leussen and Escudero (2015) argued that this reevaluation process is highly influenced by errors. If the initial perception, based on L1 knowledge, leads to misunderstandings, reevaluations and readjustments of the path are necessary.

Based on the L2LP, L2 sound contrasts that do not exist in the native language are difficult to acquire and result in erroneous perception and production. The L2LP calls this a *new* scenario which is difficult to learn. In order to acquire the L2 contrast, the learner needs to create a new category or divide and adjust the existing single L1 category. A *similar* scenario is described by an L2 contrast that is acoustically similar to two individual L1 sounds. This scenario is considered to be less difficult to learn, since L1 categories will be maintained and their boundaries will be adjusted to the L2 boundaries. The third scenario is called *subset* and refers to the mapping of an individual L2 sound to two L1 sounds ('multiple category assimilation'). It is predicted that this scenario is less problematic than the *new* scenario, since it does not require to create a new category.

2.1.4 Native Language Magnet Model (NLM)

The Native Language Magnet model (Kuhl and Iverson, 1995; Kuhl et al., 2008) is based on evidence that at birth, infants are able to discriminate between sounds by means of the acoustic differences of speech sounds. However, exposure to the native language in the first year of life results in the generalization of speech presentations that are characteristic for the respective L1 (e.g., Kuhl, 1992; Kuhl et al., 2006). This results in loss of phonetic boundaries and the ability to distinguish between certain sounds that merged into the same category. This is also called the language-specific *magnet effect*. *Perceptual magnets* thereby describe prototypes of sounds. Distinguishing between a prototype and sounds of the same category or that are 'magnetically' drawn to the prototype, lead to a poor discrimination. The learned perceptual patterns also lead to language-specific speech production as imitation of these patterns are stored in memory (Kuhl et al., 2008).

With respect to L2 acquisition, NLM argues that L2 sounds that are similar to the L1 category are difficult to discriminate. And that L2 sounds that are dissimilar to an L1 category are easier to distinguish from the native sound. The closer sounds of the second language are to the prototype of the L1, the more likely they are assimilated or drawn to this prototype.

2.1.5 Comparison of the Models of Second Language Acquisition

This section shortly summarizes the discussed second language acquisition models which differ in certain aspects from each other (see Table 2.1). While PAM argues that perception is based on L2 articulatory gestures which are compared with gestures of the native language, perception in SLM and NLM is based on the comparison of acoustic features or the similarity of phonetic categories. In contrast, the L2LP is based on comparisons on the level of phonological categories. All models argue that with increasing linguistic experience and exposure to the second language, learners are able to improve their perception and pronunciation of foreign sounds by readjusting or recreating categories (phonetic, phonological, gestural categories depending on the model).

Figure 2.3 illustrates the different predictions made by the models with regard to the difficulty of L2 perception. While PAM distinguishes between a number of perceptual outcomes, NLM differentiates between similar and not similar foreign sounds in comparison to the L1, but rather on a continuous scale. While SLM and NLM describe the difficulty to perceive a foreign sound based on a comparison of an individual L2 sound to a single sound of the native language, PAM and L2LP refer to a comparison of L2 sound contrasts to L1 sound contrasts (cf. Two Category Assimilation, PAM and

Table 2.1: Comparison of the L2 acquisition models: Perceptual Assimilation Model (PAM), Speech Learning Model (SLM), Second Language Linguistic Perception Model (L2LP), Native Language Magnet Model (NLM).

Model	**Comparison based on**	**Influenced by**	**L2 acquired by**
PAM	articulatory gestures	L1 gestures	splitting categories
SLM	acoustic features/ phonetic categories	L1 categories	creating or merging categories
L2LP	phonological categories	L1 categories and grammar	creating categories
NLM	acoustic features/ phonetic categories	L1 categories	creating categories

similar, L2LP) and to single L1 sounds (cf. Single Category, PAM and *new*, L2LP). In addition, L2LP further includes the condition *multiple category assimilation* which describes the opposite case – mapping of an individual L2 sound to two L1 sounds.

In order to help overcome phonetic and phonological difficulties learners face while learning a second language, feedback can be helpful to become aware of the deviations between the foreign accented productions and productions by native speakers. The role of feedback in a more general learning environment and with respect to pronunciation training in second language learning is discussed in the following section.

2.2 The Role of Feedback in Second Language Learning

This section discusses the main functions of feedback. For this reason, the general concept of feedback and self-regulated learning outside a language learning environment is discussed in the beginning, taking different models of feedback and self-regulated learning and the role of feedback within these models into consideration. Later, different types of feedback in classroom interaction are discussed. Because L2 classroom interactions capture more than pronunciation training, feedback in (computer-assisted) pronunciation training will also be discussed in detail.

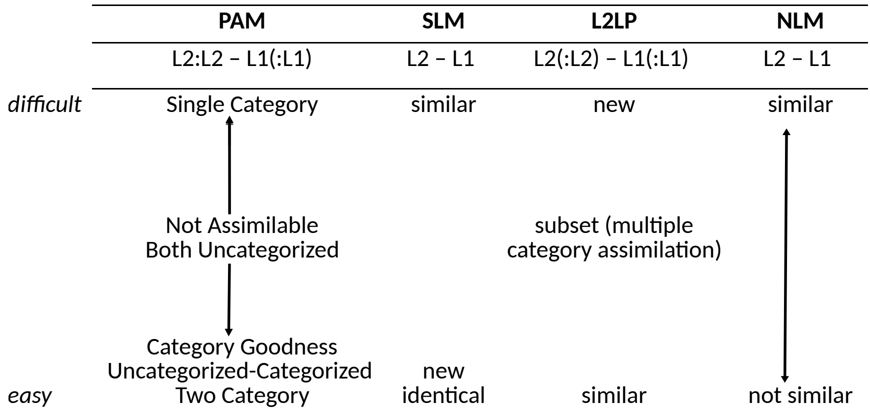


Figure 2.3: Overview of the predictions made by the Perceptual Assimilation Model (PAM), Speech Learning Model (SLM), Second Language Linguistic Perception Model (L2LP), Native Language Magnet Model (NLM).

2.2.1 Feedback and Self-Regulated Learning

One definition of feedback is the conceptualization of "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie and Timperley, 2007: 81). Feedback is only helpful if embedded in an appropriate learning context (Hattie and Timperley, 2007) and without any knowledge about the studied material, feedback only has little impact. The student has no way to make a connection between the provided feedback and the unfamiliar material (Kulhavy, 1977). Sadler (1989) argued that the learner needs to be aware of three things in order to benefit from feedback: 1) how good performance is determined, 2) if and how the learner's performance differs from the (expected) good performance, and 3) how to reduce the gap between own and good performance.

Butler and Winne (1995) described five functions that feedback serves. Feedback *confirms* that students' conceptual understanding or beliefs are consistent with instructional objectives, it *adds* information when they are missing, it *replaces* or *overwrites* incorrect knowledge, it *tunes* correct understandings, and it *restructures* information. However, Kulhavy (1977) showed that feedback can be accepted, modified, or rejected and that the usefulness of feedback depends on many factors (e.g., motivation, knowledge, beliefs). Hattie (1999) created an overview of over 500 meta-analyses which indicated a large variability regarding the usefulness of feedback mechanisms. For example, learners found it most helpful to receive feedback about a task and how to increase their efficiency. Also, feedback that was presented in form of video, audio, or was computer-assisted, and that provided additional information, was also effective. Also, it was shown that feedback was more useful when addressing correct rather than incorrect responses and that effectiveness was highly influenced by the complexity of goals and tasks. There are various types of feedback that can be used to help learners to recognize, evaluate, and address the mistakes being made, which are discussed in detail in 2.2.2 Feedback in L2 Classroom Interaction and 2.2.3 Feedback in (Computer-Assisted) Pronunciation Training.

To take a closer look on how feedback helps in a learning environment and how the different feedback functions are applied, a model of self-regulated learning (SRL) was proposed by Butler and Winne (1995) (see Figure 2.4). This model of SRL was created for a universal learning context and not explicitly for second language learning. However, this model can easily be adapted for this purpose. In order to start the SRL loop, a task has to be specified. The task encourages learners to make use of knowledge that is already available as well as beliefs for interpretation purposes, to set goals (about content, timing), as well as develop strategies and tactics in order to accomplish the set goals. By applying these strategies and tactics, specific outcomes/products are generated. A monitoring process helps the learner to validate the outcomes in relation to knowledge, beliefs, goals, and strategies, by creating internal feedback. This process might lead the learner to set new goals, to discard specific tactics and strategies or to develop new procedures. It is believed that learners who are more self-regulated, are more effective in accomplishing the set goal and are more successful at learning (Zimmerman and Schunk, 2001). External feedback can not only confirm and add to the learners performance but it can also conflict. If it conflicts with beliefs, knowledge, goals, or strategies, the task can be reinterpreted and the loop can be re-entered in order to generate an adjusted performance (Butler and Winne, 1995).

Figure 2.5 illustrates the SRL model in the context of L2 pronunciation. As a task, the pronunciation of the French word <pain> [pɛ̃] (bread) was chosen. Imagine a German learner of French in an L2 classroom who has to pronounce this word. The learner already knows that French voiceless stops are produced with a short Voice Onset Time (VOT) – in contrast to German, where voiceless stops are aspirated. This knowledge can be associated with *domain knowledge*. For *strategy knowledge* the learner knows the stop has to be produced with less intensity than in German which results in the production of a shorter VOT. The learner's *motivational beliefs* are that these

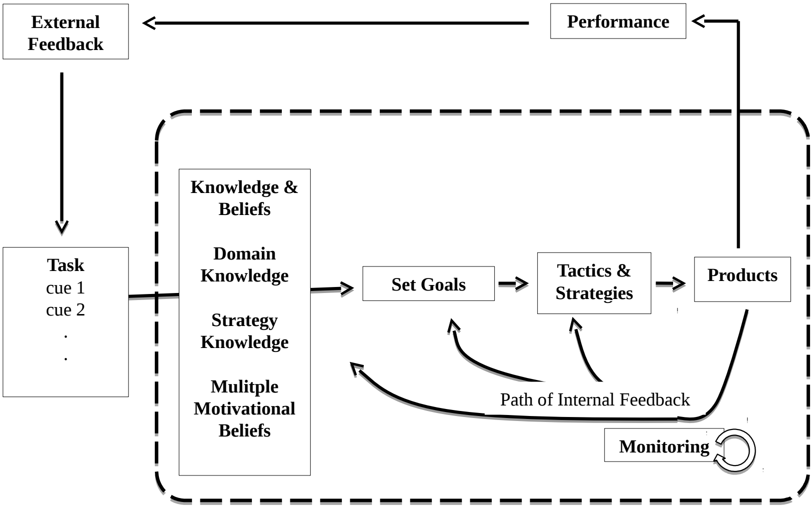


Figure 2.4: Model of self-regulated learning (Butler and Winne, 1995: p. 248).

phonetic changes might reduce the perceived foreign accent and the learner's *goal* is exactly that: reduce the foreign accent to a more native-like level by successfully changing the acoustic parameters. As *products*, Levelt's model *blueprint of a speaker* (Levelt, 1989) can be applied: from generating the preverbal message to the creation of the surface structure and the phonetic plan. The actual execution of the plan can be associated with the learner's *performance*, which can be the correct French pronunciation [pɛ̃] or a more German-like production [pʰɛ̃]. According to the outcome, external feedback can be given, which can be positive or negative, corrective feedback.

Motivation is an important impact factor on the outcome. If the goal of the learner is not to sound native-like, the learner can still have the correct domain and strategy knowledge but chooses not to act on it. Corrective feedback might be noted but ignored in subsequent productions. An accented *performance* might also result from the fact that the learner does not have this kind of knowledge. In this case feedback has to go beyond the point of simply being corrective. The learner needs to receive further information on *why* the production was not correct and on *how* to correct it. This way, the loop can be re-entered with additional information and strategies in order to produce a better outcome.

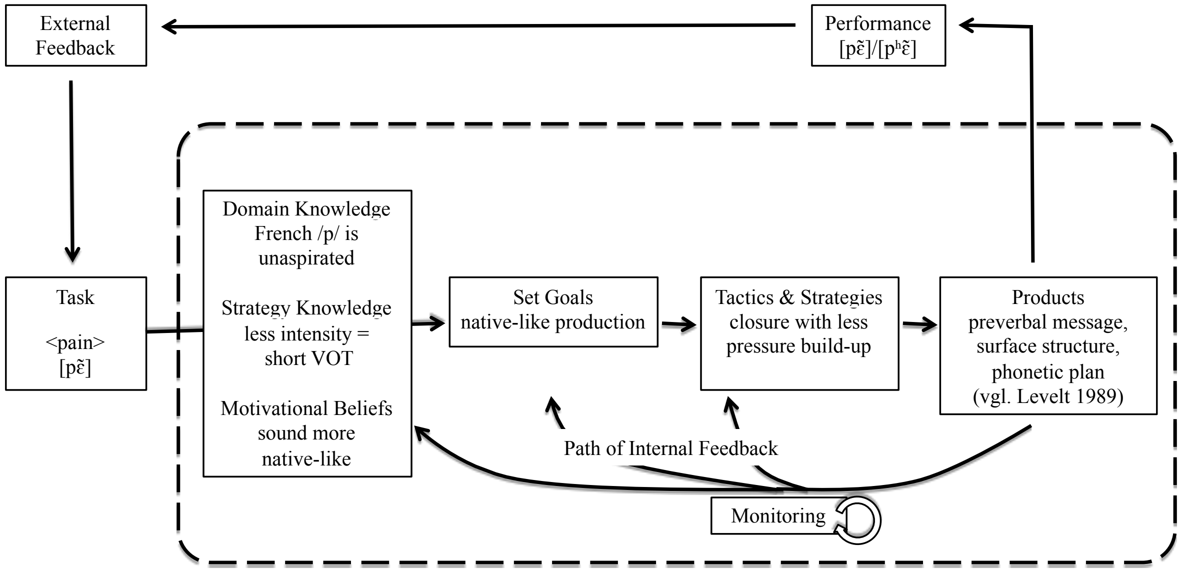


Figure 2.5: Adapted model of self-regulated learning (Butler and Winne, 1995) in the context of L2 pronunciation of the French word <pain> (bread).

To understand why some types of feedback are more effective and enhance learning, while others are not, Hattie and Timperley (2007) developed a model of feedback shown in Figure 2.6. Their claim is "that the main purpose of feedback is to reduce discrepancies between current understandings and performance and a goal" (p. 86). These discrepancies can be reduced by the students themselves, by applying more effective strategies, and making a greater effort to achieve the set goals, or by doing the opposite, e.g., discarding or lowering the set goals. The discrepancies can be reduced by the teacher, by setting specific goals, and providing appropriate feedback in order to help the students reach these goals.

Hattie and Timperley (2007) argue that efficient feedback works on three different levels: feed up ('Where am I going?' with respect to the set goals), feed back ('How am I going?'), and feed forward ('Where to next?'). These three feedback questions can help students to effectively work at four levels: task level, process level, self-regulation level, and self level. With regard to each of these levels, feedback needs to address specific needs in order to reach its full potential. In this respect feedback on a task is useful when it helps the learner to reevaluate or reinterpret specific issues of the task. Feedback is in no way helpful if it results from a lack of understanding. At the process level, feedback shows its full potential when it helps students to reformulate strategies and to reject faulty hypotheses. Regarding the self-regulation level, feedback is beneficial when it addresses learners' motivational beliefs and important regulatory processes that engage with the task. And lastly, feedback at the self level is considered to be mostly ineffective, as it rarely addresses the three feedback questions. However, Hattie and Timperley (2007) also note that in specific situations, instructional information can be more effective than feedback. Without any previous knowledge (cf. domain knowledge, Butler and Winne, 1995), feedback is mostly useless (cf. Kulhavy, 1977).

Nicol and Macfarlane-Dick (2006) suggested seven principles of good feedback practice in order to develop self-regulation processes. They integrated these principles in a model of formative assessment and feedback built on work by Butler and Winne (1995). Drawing on relevant research literature, seven principles of good feedback practice were identified which support and develop self-regulation: 1) clarify what good performance is, 2) facilitate self-assessment, 3) deliver high quality feedback information, 4) encourage teacher and peer dialogue, 5) encourage positive motivation and self-esteem, 6) provide opportunities to close the gap, and 7) use feedback to improve teaching.

In the context of speech, according to Ball and Rahilly (2013) different types of feedback can be used in order to self-regulate while speaking. First, auditory feedback is used to self-regulate phonetic accuracy regarding

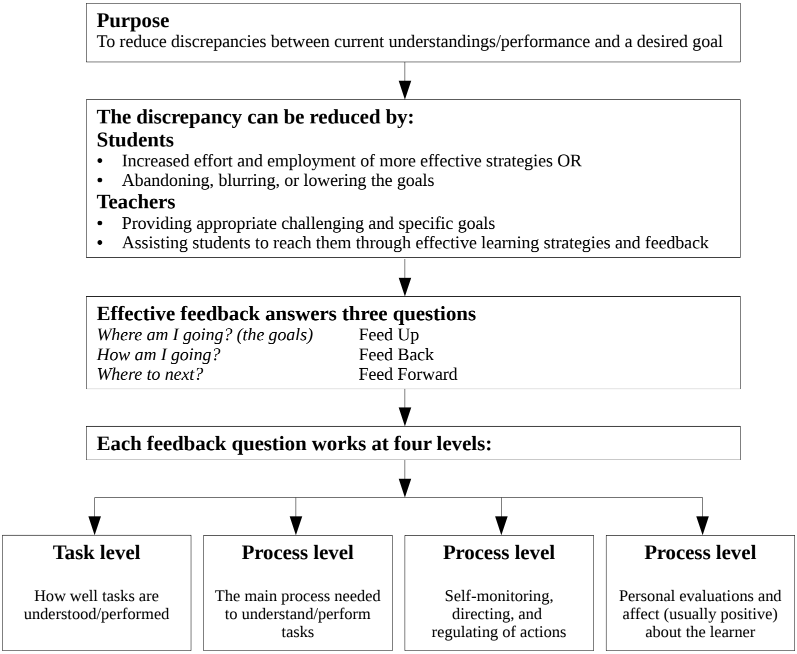


Figure 2.6: Model of feedback (Hattie and Timperley, 2007).

the correct pronunciation of sounds and prosodic aspects. However, auditory feedback is relatively slow as it takes about 160 to 250 ms until a produced sound is perceived and the speaker is ready to make any corrections. It was already discussed that speakers, in the context of L2 pronunciation, have difficulties to perceive their deviations to native speakers (Flege, 1995; Best, 1994; Escudero, 2005; Kuhl and Iverson, 1995). For this reason, self-regulation with the help of auditory feedback might not be sufficient. That auditory feedback is essential for speaking is shown in the literature. Delaying auditory feedback is known to reduce the degree of stammering for stutterers (e.g., Ingham and Andrews, 1971; Curlee and Perkins, 1963; Kalinowski et al., 1993, 1996) but disrupts continuous speech by speakers who do not stutter (e.g., Lee, 1950; Yates, 1963; Zanini et al., 1999; Stuart and Kalinowski, 2015).

Another feedback method available to the speaker while articulating is the tactile and kinaesthetic feedback, which describes the sense of touch and movement of the passive and active articulators. Therefore, speakers can self-regulate the movement and placement of their articulators while

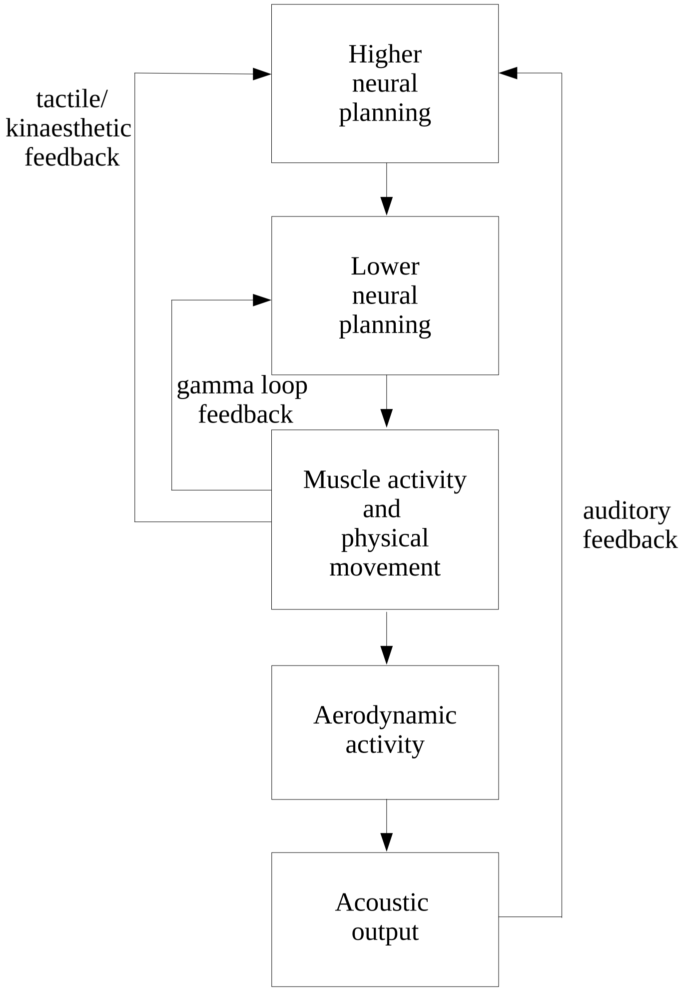


Figure 2.7: Model of speech production and feedback (Ball and Rahilly, 2013).

speaking. It was shown that tactile feedback can be helpful when combining it with visual feedback, for example in ultrasound studies (e.g., Tateishi and Winters, 2013; Pillot-Loiseau et al., 2015). Pillot-Loiseau et al. (2015) found that Japanese learners of French who received several ultrasound lessons while training the French vowel /u/ performed better after training than a control group. Tateishi and Winters (2013) showed that ultrasound training was also helpful in improving the pronunciation of /l/ by Japanese learners of English. However, it did not help to improve the pronunciation of /r/, which became even more accented.

Although the tactile/kinaesthetic feedback is faster than auditory feedback, it is still not sufficient to monitor fine muscle activity needed for certain sounds. The third type of feedback, and the fastest one, is called gamma loop feedback which is derived from knowledge about the physiology of the neuromuscular system. This method allows for comparisons at a lower neural activity level, which enables a really fast monitoring procedure. Based on these feedback mechanisms, Ball and Rahilly (2013) proposed a model of speech production and feedback which is illustrated in Figure 2.7. It illustrates the discussed types of feedback that are available during speaking which can help learners to self-regulate while speaking.

After discussing how feedback and self-regulated learning can be helpful for effective learning, performance, and self-regulation, the next section discusses how feedback is incorporated in second language classrooms.

2.2.2 Feedback in L2 Classroom Interaction

In one of the first reviews of error correction in classroom interaction, Hendrickson (1978) described that there was a noticeable change in methodology and material with regard to foreign language teaching. He noticed a shift from pure error prevention to using errors for learning purposes and raised the following issues:

(1) **Should** learner errors be corrected?

(2) **When** should learner errors be corrected?

(3) **Which** learner errors should be corrected?

(4) **How** should learner errors be corrected?

(5) **Who** should correct learner errors?

The following subsections discuss these issues with the help of the research literature.

Should learner errors be corrected?

There are many studies which support the opinion that the correction of errors is connected with negative effects (e.g., Horwitz et al., 1986; Truscott, 1996, 1999; Young, 1991). For example, Truscott (1996) argued that error correction should be abandoned with regard to grammar correction, since research failed to show evidence of grammar correction being effective. According to the author, it is even supposed to be harmful for the learner. Truscott reasoned that, since many studies on L1 learning did not show any improvements with respect to grammar, an effectiveness for L2 is ruled out. He claims that corrections interrupt, trigger negative emotions, are given inconsistently, are sometimes ambiguous (e.g., recasts) and are often not taken seriously.

In contrast, the *Noticing Hypothesis* by Schmidt (1990) suggests that a particular linguistic term can only be acquired, when the learner has been made aware of it. Corrective feedback can help to generate this awareness, especially with regard to discrepancies between the learner's output and the language to be learned (Long, 1996). Similarly, Gass (1991, 1997) argued that errors can not be learned correctly unless the learner is aware of its correct form.

With the help of feedback, the learner can focus on particular problems (Neri et al., 2002) and enable self-regulation and self-improvement (Butler and Winne, 1995). A similar approach is presented by the *Behavioristic Theory* by Watson (1926); Thorndike (1932); Skinner (1957). But within the scope of this theory, the learner is simply told if something is correct or incorrect and does not receive any additional information and, therefore, does not give the learner the possibility to self-evaluate. Ur (2012) conducted a survey with 500 students from primary and secondary school studying English in Israel. The questionnaire included questions on the preferred type of corrective feedback and whether they wanted to be corrected at all (both in oral and written form). The results showed that students want to be corrected, although this preference was stronger for feedback in written form. Furthermore, students want to be informed of the correct form (i.e. in form of a recast in oral situations or by giving the correct answer in written form). Levine (1975) reasoned that if an error is not corrected it will be considered correct and therefore learned incorrectly (i.e. also by other students) and the error will be maintained. Furthermore, several meta-analyses and state-of-the-art overviews were carried out, demonstrating the effectiveness of corrective feedback (e.g., Miller, 2003; Russel and Spada, 2006; Mackey and Goo, 2007; Li, 2010; Lister and Saito, 2010; Lyster et al., 2013; Kang and Han, 2015; Liu and Brown, 2015; Brown, 2016).

When should learner errors be corrected?

Many studies addressed the question whether to give corrective feedback immediately or after a certain time (delayed feedback). Regarding classroom interaction it is argued that giving students immediate feedback would interrupt the student (e.g., Long, 1977; Bartram and Walt, 1991; Harmer, 2001) and should, therefore, be avoided. In contrast, Kulik and Kulik (1988) conducted a meta-analysis of 53 studies including applied studies (i.e. class room observations) and experimental studies on feedback timing. 26 studies showed that delayed feedback had a beneficial effect over immediate feedback. However, the other 27 showed an advantageous effect for immediate feedback over delayed feedback. Overall it can be said that the efficacy of delayed feedback was shown for experimental research studies whereas the effectiveness of immediate over delayed feedback was shown for applied studies in the classroom. However, there was little research carried out with regard to immediate vs. delayed feedback after Kulik and Kulik's meta-analysis.

Quinn (2014) conducted a study with intermediate adult learners of English as a second language. Participants were assigned to three groups which received either immediate, delayed or no corrective feedback in one-on-one sessions. Several pre-tests tested their knowledge of English passive construction. Each participant received a short lesson on English passive construction and participated in three communicative language tasks to elicit the use of passive construction, in which they received corrective feedback (or no feedback for participants of the control group). Participants of the immediate feedback group received corrective feedback directly after an error occurred. In contrast, participants of the delayed feedback group received feedback at the end of each task. Following the tasks, all subjects completed a set of post-tests which were also repeated one week later ('delayed post-test'). Results showed that participants of all groups improved equally well. This means that with regard to acquiring passive construction in English, no corrective feedback is required.

Li et al. (2016) investigated the effect of immediate and delayed corrective feedback with respect to past passive construction in English by Chinese middle school students. The learners were assigned to four groups: immediate feedback, delayed feedback, no feedback, control. All learners of the experimental groups received a 2-hour instruction session in which they practiced to re-narrate a story to each other and later retelling the narrative to the class. Both feedback groups received the appropriate feedback as prompts to encourage self-correction. If the repair was still incorrect or no response was given by the learner, a recast was offered containing the correct form. Delayed feedback was given after the second task. Participants of all groups performed two tests: a grammaticality judgment test and an elicited imitation test (grammatical and ungrammatical sentences produced by a native speaker, which had to be rephrased (and corrected, if necessary) using the past passive construction). With regard to the elicited imitation test, no effect was found between feedback timing. However, immediate feedback was beneficial for the grammaticality judgement test. It seems as if the design of the experiment and the pre- and post-test conditions influence the outcome. Both Quinn (2014) and Li et al. (2016) tested the effect for English passive construction but received different results. One reason might also be the duration of the instruction session. Participants in the study by Li (2010) received a 2-hour session whereas the instruction session of the study by Quinn (2014) was much shorter (10 minutes).

Which learner errors should be corrected?

According to Neri et al. (2002), feedback should be given based on 1) error frequency, 2) error persistence, and 3) perceptual relevance. For errors that appear rarely, giving feedback will not have much impact on performance. However, when a particular error occurs quite frequently, feedback can help to reduce the appearance of this error and might therefore have an important impact on communication. Persistent errors that will not disappear with increasing exposure to the L2 should also be addressed. Otherwise, giving corrective feedback is not really necessary. In addition, errors that evidently hinder intelligibility should be tackled by giving feedback. In contrast, if the error is perceivable but does not necessarily hinder intelligibility, feedback can but does not have to be given. For example, French learners of German show problems with the correct pronunciation of /ç/ and tend to produce it as [ʃ]. However, there are regional varieties of German in which /ç/ is also realized as [ʃ]. In this regard, feedback can help the learner to sound more native-like (standard German) but it is not particularly necessary to correct this mistake, since it usually does not hinder intelligibility. Neri et al. (2002) argued errors should be corrected with a communicative approach in mind. Therefore, error correction should address mistakes that hinder intelligibility rather than working on the goal to sound more native-like. Overall, this argument is of course correct, however, a distinction between beginners and advanced speakers is necessary. Beginners are usually challenged by other, more fundamental problems. Therefore, focusing on specific issues that go beyond the scope of simply being intelligible should not be excluded a priori. The reason behind this approach is that corrective feedback should not be given in an overwhelming manner. This could discourage students and keep them from participating (Neri et al., 2002).

Ur (2012) posed the question if receiving corrective feedback might be humiliating to the student (cf. Truscott, 1996). She pointed out this risk exists and that teachers need to be aware of it. By humiliating the student, the teacher runs the risk that the student will stop participating in the lesson. However, Ur (2012) reasoned that feedback should be given publicly instead of privately. This way, not only individual learner will profit from it but the whole class. With regard to the amount of corrective feedback given, Ur (2012) agrees with Neri et al. (2002). She argued that too much corrective feedback can demoralize the student. However, she also pointed out that too little feedback can also lead to frustration or confusion. Feedback acknowledging correct answers or pronunciation is also very important in order to keep the student motivated.

How should learner errors be corrected?

To answer this question, one particular study is taken into consideration that described different types of feedback in detail and discussed the effect they had on the learning behavior of the learner. Lyster and Ranta (1997) conducted a classroom study in order to investigate the type and frequency of feedback used by teachers. The survey was carried out in six French immersion classrooms. Lyster and Ranta (1997) were interested in what manner feedback would be used to handle erroneous answers. Recordings were checked for errors and only those with errors regarding the language itself, not content, were considered to be faulty. They differentiated six types of feedback which were used during the examined lessons:

(1) Explicit correction

(2) Recasts

(3) Clarification request

(4) Meta-linguistic feedback

(5) Elicitation

(6) Repetition

Lyster and Ranta (1997) identified that about 60 percent of all erroneous utterances received some kind of feedback by the teacher but only 27 percent led to a repair by the student. With respect to the six feedback types (see Table 2.2), recast was the most preferred method, which was used in 55 percent of the time. Although recasts were used frequently, it was not very efficient in eliciting students uptake.

Lyster and Ranta (1997) define recasts as a corrected reformulation by the teacher. There is evidence that recasts benefit learning (e.g., Mackey and Philp, 1998), while others stress that recasts might not be helpful (e.g., Long, 1996; Lyster and Ranta, 1997; Lyster, 1998; Ellis et al., 2006; Ammar, 2008). Mackey and Philp (1998) investigated the effect of recasts as well as the nature and content of the responses to recasts by a group of learners in information-gap tasks. Additionally, a different group received different types of feedback, depending on the performance. Results showed that recasts rarely generated uptake by the learners. However, advanced learners who received recasts produced more higher-level morphosyntactic forms. With regard to less advanced learners, recasts were not as effective.

In contrast, elicitation was the most effective method to encourage students to repair the error (43%) in the study by Lyster and Ranta (1997). Ammar (2008) conducted an experiment with two experimental and one control group. Over a period of four weeks, participants of three intensive English as a second language classes received corrective feedback either as a) recasts without pushing the students to self-correct, b) prompts with encouragement to self-correct (elicitation, repetition, metalinguistic feedback), or c) no corrective feedback. The aim of the experiment was to test the impact of the different styles of giving feedback on the acquisition of English third person possessives determiners by francophone learners. Results showed that learners who received prompts, encouraging them to self-correct the mistakes made, helped the students to improve more than learners who received recasts and no corrective feedback. This was especially true for low-proficiency learners. Ellis et al. (2006) investigated the effect of recasts and metalinguistic feedback with regard to the acquisition of past tense <-ed>. Low-intermediate students of English as an L2 received one of the stated feedback types or no corrective feedback (control group) during two communicative tasks. Results indicated that learners who received metalinguistic feedback showed a larger improvement with regard to past tense <-ed>. Additionally, it was shown that learners receiving recasts did not show a difference in performance.

Table 2.2 shows hypothetical examples for the six identified feedback types regarding the erroneous pronunciation of the French phrase <les amis> by a German learner of French. The student pronounced this phrase as [leʔami] instead of [lezami]. Referring to Lyster and Ranta (1997), the learner does not have any reason to generate a repair in the case of *explicit correction* or *recast* because the teacher is already offering the student the correct answer. However, students might not be able to perceive the difference between the made mistake and the recast of the teacher. Since students have difficulties to hear the discrepancies of their own errors and the correct forms of the teacher, learners might also interpret the recast as a confirmation of a correct form (Long, 1996). It is often argued that errors can not be learned

Table 2.2: Hypothetical examples for the six types of corrective feedback found in a class room study (Lyster and Ranta, 1997) regarding the erroneous pronunciation of the French phrase <les amis> [leʔami] instead of [lezami] by German learner of French.

Feedback Type	**Example**
Explicit correction	"No, it's [lezami]."
Recast	"[lezami]."
Clarification request	"Pardon me?"
Elicitation	"[le]?"
Repetition	"[leʔami]?"
Metalinguistic feedback	"You have to pronounce the <s> from <les> because the word following <les> starts with a vowel. This is called liaison."

correctly unless the learner is aware of its correct form (e.g., Schmidt, 1990; Gass, 1991; Long, 1996; Gass, 1997). For this reason, recasts might not be the most appropriate method, but it seems to be the easiest and fasted method for a teacher in an L2 classroom. The remaining feedback methods clarification request, elicitation, repetition, and metalinguistic feedback do not include the correct form of the error and encourage the student to self-correct. These feedback methods can raise awareness of the error. However, if the uptake includes the same error or a different error, and the teacher chooses to continue with the topic, these false corrections might be regarded as correct and, therefore, learned incorrectly.

Who should correct learner errors?

Ur (2012) argued that students do not like to be corrected by other classmates because they do not rely on their corrections and suggestions. However, teachers sometimes do not recognize mistakes and give feedback inconsistently (cf. Truscott, 1996). Following Lyster and Ranta (1997), another approach of correcting mistakes is to let the learners correct themselves by using the appropriate feedback method. A fourth option would be to make use of computer-assisted language learning (CALL) or computer-assisted pronunciation training (CAPT) systems that are developed particularly to train specific grammar or pronunciation related problems (e.g., Eskenazi and Hansma, 1998; Neri et al., 2008a; Rosetta Stone, 2016; EnglishCentral, 2016).

2.2.3 Feedback in (Computer-Assisted) Pronunciation Training

The question arises why it is necessary to focus on pronunciation training when learning a foreign language, especially if a learner does not have the aim to sound native-like. However, a foreign accent can not only decrease intelligibility (e.g., Lane, 1963; Mennen et al., 2007; but see Munro and Derwing, 1995a; Hayes-Harb and Watzinger-Tharp, 2012) but also elicitate negative stereotypes. For example, learners can be perceived as less educated or less reliable than native speakers (e.g., Lambert et al., 1960; Bresnahan et al., 2002; Munro et al., 2006; Tsurutani, 2012). Neri et al. (2002) argue that pronunciation training should always focus on a communicative approach, unless the learner has special needs (i.e. the learner wants to focus on specific pronunciation problems). It should focus on speech intelligibility and not on the ability to speak native-like.

Different attempts have been made to make learners of a foreign language aware of pronunciation differences. Several studies addressed the issue if and which corrective feedback methods are beneficial to help learners improve in the pronunciation of a second language, both in classroom interaction and by using computer-assisted pronunciation training systems. The following subsections focus on feedback methods used in both contexts. Only a broad overview will be given since the following chapters of this thesis discuss more research in depth.

Feedback and Pronunciation Training in the L2 Classroom

Speaking a foreign language calls for more than in depth knowledge of grammar or vocabulary. A correct pronunciation on a segmental as well as a suprasegmental level (e.g., correct placement of pitch accent, pitch range, phone/syllable duration) is of vital importance for communication (Setter and Jenkins, 2005). A correct pronunciation is essential for speaking a foreign language, since a poor pronunciation will make it harder to be understood by other interlocutors (e.g., Mennen et al., 2007). The main challenge for the learners is to become aware of their own mistakes and to perceive the deviations in the native pronunciation of the respective L2 in comparison with their own realizations (Barry, 2007). It is especially problematic if the phonetic and phonological system of the native language (L1) interferes with the phonetic and phonological system of the foreign language (L2) (e.g., Flege, 1995; Best, 1994). Therefore, a learner is challenged by several phonological and phonetic differences between L1 and L2.

However, many L2 class room situations are characterized by minimal effort to teach pronunciation or include specific pronunciation and perception exercises to make learners aware of their deviations from native speakers (e.g., Setter and Jenkins, 2005; Hirschfeld and Trouvain, 2007). However, it is a challenge for teachers to visualize and explain pronunciation problems sufficiently without overwhelming the learners. Specially adapted exercises for specific phenomena need to be created, which is very time consuming. Most curricula do not reserve much time for pronunciation training and most effort is placed on grammar, vocabulary, or text comprehension, among other things, although pronunciation is known to be an important factor in second language learning (Setter and Jenkins, 2005). In the context of English as a second language, Setter and Jenkins (2005: p. 5) argue that "pronunciation needs to lose its isolated character and be treated pedagogically as part of communication and discourse". Listening tasks should include not only one variety of the learned language to increase exposure and variability. Furthermore, Baker (2011) argues that pronunciation is rarely taught in the education of L2 English teachers and that many teachers might not have adequate knowledge in order to teach pronunciation to students.

Furthermore, an utterance of the teacher who is often not a native speaker of the foreign language, might contain interferences itself. Therefore, even if the learners imitate the teacher correctly, they might acquire the teacher's interferences (Künzel, 1977; Hirschfeld and Trouvain, 2007). Hirschfeld and Trouvain (2007) argued that many teachers of German believe that imitation is the key to acquire the correct prosody, but that most learners are unable to produce correct imitations. For this reason, perception and production skills need to be developed with the help of appropriate exercises. Therefore, they propose the following methodological procedure[2]:

(1) Introducing the topic

(2) Perception exercise (discrimination and identification)

(3) Production exercise (imitation, individually and in chorus)

(4) Correction of erroneous pronunciations

(5) Repeat perception exercise

(6) More production exercises (imitation) with feedback

(7) Automation by repeating, reading, variation of speaking style

They also suggest various types of production and perception exercises that can be incorporated in the lesson.

[2] With regard to learning the correct prosody in German.

Hirschfeld (2012) debated that body movement can help the learners to understand certain problems and acquire the correct pronunciation of segmental and suprasegmental entities. She developed the pronunciation training procedure 'Bewegte Phonetik' (moving phonetics) that helped French learners of German and English to train pronunciation in both languages. For example, to illustrate that long German vowels are produced with a too short duration, movements of the hand can help to see, hear, and even feel the difference. Clapping one's hands can also help to point to different positions of word accents (e.g., names, French = SebasTIAN, German/English = SeBAStian). Hirschfeld (2012) lists many more exercises with respect to German and English pronunciation training.

It has to be noted that these teaching styles by Hirschfeld and Trouvain (2007) and Hirschfeld (2012) are very time consuming and focus almost completely on specific pronunciation classes. Unfortunately, most teachers will not have the time to include comprehensive pronunciation training or do not place much focus on pronunciation at all. However, the methodological procedure could be included to a smaller extent, as could (rather) unconventional exercises using the own learner's body.

Lastly, my own experience with L2 learning classes is that teachers do not spend much time on pronunciation training. When starting my first French class at university, only half an hour was devoted to phonetics. That excursion was accompanied with the statement "Listen carefully, we will not talk about this again". Of course, feedback was given throughout the class when an error was made, but irregularly and mostly in the form of recasts. The teacher did not always know how to impart specific phonetic knowledge, which might have been helpful. In another class, a man with a Spanish background had problems to perceive and produce the French vowel /y/. Although the teacher corrected him in form of explicit corrections, recasts, tried some kind of metalinguistic feedback, and produced [y], the student was not able to perceive the difference to his wrong production [u]. This resulted in a rather long, back and forth encounter that left both the learner and the teacher unsatisfied.

When teachers are not properly trained in teaching of L2 pronunciation and/or do not have the time to focus on learning and training the correct pronunciation, another approach is to make use of computer-assisted pronunciation training (CAPT) systems. These systems give the learner the possibility to train as much or as little as they want and when they want. A number of CAPT systems were evaluated in the following subsection.

An Assessment of Feedback in Computer-Assisted Pronunciation Training (CAPT) Systems

Martin (2004) brought up the question why it is important to have computer-assisted pronunciation training. He argued that these systems can help students to take the step from passive learning to active learning. Students that are self-conscious in classroom interaction are able to actively engage with the system without feeling judged by classmates or the teacher (cf. 'embarrassment and the teacher's responsibility', Ur, 2012). Martin (2004) refers to three pedagogical statements of learning:

(1) I hear and I forget.

(2) I see and I remember.

(3) I do and I understand.

Computer-assisted pronunciation training systems are especially valuable regarding the second and third statement. Learners are encouraged to produce non-native utterances ('do') in order to understand how to tackle certain problems ('understand') and feedback methods can help them to 'see' what was done wrong and to 'remember' how to avoid the mistake made. However, not all feedback methods implemented in CAPT systems are helpful. Using feedback which is difficult to interpret runs the risk of confusing and demotivating the learner. According to Neri et al. (2002) feedback should be given based on 1) error frequency, 2) error persistence, 3) perceptual relevance, and 4) robustness of error detection (for more information see discussion 2.2.2 "Which learner errors should be corrected", without 4) robustness of error detection). For this reason, several commercial and non-commercial CAPT systems, non-commercial systems for research purposes as well as research literature testing specific feedback methods that could be integrated in a CAPT system were evaluated[3] and are listed in Table 2.3. Please note that the evaluation of the listed systems was carried out with a focus on the feedback methods used. With regard to the systems listed in the table, only the first five systems and 'SpeakGreek' were available for a hands-on experience. Concerning the rest of the systems, evaluation was done based on the named references. For this reason, it can not be ruled out that some systems include, in fact, more feedback methods than listed here.

[3]The evaluation of some of the listed CAPT systems was part of an assessment by colleagues working in the IFCASL project. The detailed assessment was published in Trouvain et al. (2016c), which not only evaluated feedback with regard to pronunciation training. It also included an evaluation of available instructions, the quality and type of the exercises, scoring mechanisms as well as the pedagogical structure of the systems.

Table 2.3: List of commercial and non-commercial CAPT systems as well as research literature testing feedback methods that could be integrated in a CAPT system environment.

CAPT System	Reference	Language Pair
Cool Speech	Cauldwell (2012)	any language → English
English Central	EnglishCentral (2016)	any language → English
EyeSpeak	EyeSpeak (2016)	many language pairs
Pronunciation Coach	Rose Medical Solutions (2016)	any language → English
Rosetta Stone	Rosetta Stone (2016)	many language pairs
AzAR	Jokisch et al. (2005)	Slavic languages → German
BetterAccent Tutor	Komissarchik and Komissarchik (2000a,b)	any language → English
Dutch-CAPT	Neri et al. (2008a)	any language → Dutch
Euronounce	Demenko et al. (2009)	Slavic languages → German
FLUENCY	Eskenazi and Hansma (1998); Eskenazi et al. (2000)	any language → English
Fresh Talk	Precoda et al. (2000)	English → Spanish
PARLING	Neri et al. (2008b)	Italian → English
PLASER	Mak et al. (2003)	Chinese → English
SpeakGreek	Nicolaidis et al. (2015)	any language → Greek
Ville	Wik and Hjalmarsson (2009)	any language → Norwegian
WebGrader	Neumeyer et al. (1998)	many language pairs
WinPitch LTL II	Martin (2004, 2010)	any language pairs
	Bissiri and Pfitzinger (2009)	Italian → German
	Hirose et al. (2003)	any language → Japanese
	Meron and Hirose (1996)	any language → Japanese/ English

The last three research papers listed in Table 2.3 are not CAPT systems per se but investigated feedback methods that could be included in a CAPT system environment. They are summarized shortly.

Bissiri and Pfitzinger (2009) showed that resynthesizing the voice of Italian learners of German had a beneficial and motivating effect on learning lexical stress in German. They tested a manipulation group that was trained on resynthesized utterances, and a native speaker group that was trained on utterances recorded by a German native speaker. Participants from the manipulation group trained either on utterances from the pre-test with or without emphasis on the stressed syllable of the wrongly pronounced word. Both conditions were also incorporated in the training condition for the native speaker group. For this, the native speaker recorded the material in a normal speaking style, and with an emphasis on the stressed syllable of the wrongly pronounced word from the pre-test of the learners. According to the authors, the auditive feedback had a motivating effect for participants of the manipulation group whereas subjects of the native speaker group showed no interest in the training. This result is, however, based on evaluations by the participants. No acoustic analysis was performed comparing pre- and post-test recordings or comparing the two groups with each other. A perception experiment with German native listeners rating the correctness of word stress showed that both groups received the highest scores for post-test performances after training with emphasized utterances. Both groups performed equally well. Regarding post-test performances without emphasis on the stressed syllable, participants of the manipulation group received higher scores than participants of the native speaker group.

Meron and Hirose (1996) focused on manipulation of the learner's own voice in the context of Japanese or English as a foreign language. They integrated a manipulation procedure that resynthesized the speaker's speech in a way that the wrongly pronounced syllable was manipulated based on fundamental frequency, duration, and intensity of a reference speaker. They argued a teacher would exaggerate wrongly pronounced syllables in a classroom situation to draw focus to the correction. For this reason, the system did not only transfer the prosodic features of a reference speaker, but also emphasized the syllable that contained an error.

Lastly, Hirose et al. (2003) developed a system for learners of Japanese. The focus of the system was to teach the learners the correct pronunciation of lexical accents. Both auditory and visual corrective feedback were included. Concerning the audio feedback, the learner's voice was resynthesized with the prosodic features of the reference speaker. Additionally, visual feedback in form of stylized pitch contours and arrows indicating falling and rising pitch was displayed for the original and the modified utterance in order to highlight the differences.

With regard to the listed CAPT systems and research investigations, a list of utilized feedback method was compiled which shows how many of these systems used a specific feedback method (see Table 2.4).

To indicate if sounds, words, phrases, or sometimes complete utterances are produced correctly, many systems make use of color scales (i.e. green, yellow, red) (Precoda et al., 2000; Mak et al., 2003; Jokisch et al., 2005; Demenko et al., 2009; Neri et al., 2008a; Rosetta Stone, 2016; EnglishCentral, 2016; EyeSpeak, 2016; Rose Medical Solutions, 2016). However, if the pronunciation is not correct, most systems do not give any indication what the error is about. This is especially problematic if the color scale refers to a complete word or even larger unit. The learner is left alone to interpret the feedback, which might not be useful at all because it does not indicate the actual problem. In contrast, feedback on a sound level is only so helpful – it does indicate which sound was produced incorrectly, but without any information on what needs to be changed with respect to the articulation of the particular sound, no profound change can be made. The same goes for different types of visual corrective feedback, such as smileys (smiling/being upset), written feedback (e.g., 'correct'/'incorrect'), or tick marks (e.g., Neri et al., 2008a; Wik and Hjalmarsson, 2009; Rose Medical Solutions, 2016). When the systems indicate an erroneous production, the user does not receive any additional information and can only guess what to change in the word or utterance. Similarly, overall articulation scores (e.g., Neumeyer et al., 1998; Precoda et al., 2000; Rose Medical Solutions, 2016; EnglishCentral, 2016) do give a general overview on how well the learner is doing but fail to give detailed information about why the user is not performing at 100 percent.

All but one system (Nicolaidis et al., 2015) give the possibility to listen to a native reference speaker and the own recording. But only four systems allow the user to focus on the pre-recorded utterances in different speaking rates (Precoda et al., 2000; Martin, 2004; Wik and Hjalmarsson, 2009; Martin, 2010; EyeSpeak, 2016). This feature might be helpful for the learner to listen to particular difficult sounds and sound clusters. Also, four systems give the opportunity to look at a reference video recorded by native speakers producing individual sounds or words (Mak et al., 2003; Martin, 2004; Jokisch et al., 2005; Rose Medical Solutions, 2016). However, it is unclear how helpful this kind of feedback is. Since users are only able to see movements of the lip and, for some sounds, the tongue, they might not be able to retrieve much valuable information from this, although it might, of course, be useful for rounded and unrounded sounds. A more helpful type of visual feedback can be found in (animated) vocal tracts (e.g., Mak et al., 2003; Jokisch et al., 2005; Rose Medical Solutions, 2016) in which the learner can see the different positions of the tongue throughout the production of a sound, word, or utterance. The Pronunciation Coach (Rose Medical Solutions, 2016) gives the possibility to choose between a continuous animated illustration of a word to look at each phoneme of a word individually.

With regard to stress and pitch in general, many systems display pitch contours of the reference speaker and the learner (Komissarchik and Komissarchik, 2000a,b; Martin, 2004, 2010; Demenko et al., 2009). Some, but not all systems give information on what is expected from the learner and how to interpret the pitch contour (see Komissarchik and Komissarchik, 2000a,b). Hirose et al. (2003) chose to go without the display of a complex pitch contour and made use of a stylized pitch contour instead, indicating rising and falling movements with simple arrows. With regard to stress, Komissarchik and Komissarchik (2000a,b) applied an interesting visual feedback method: The system displays a stair-like structure in which each syllable is represented by a step. The longer the step, the longer the syllable, and the higher the step, the more intensity is used to produce the syllable. Again, the training window includes exact information on how to interpret this feedback.

Table 2.4: Visual and auditory feedback methods used in 20 CAPT systems and research investigations, including the amount of systems offering these particular features.

visual		auditory	
Color scales	9	Listening to own recording and native reference speaker	19
Explicit feedback about correctness, e.g., smiley, 'correct/incorrect', tick mark	8	Resynthesis of voice	4
Display of pitch curve	8	Different speech rates	4
Overall articulation score	8	Reference video	4
Oscillogram	7		
(Animated) vocal tract	6		
Information about pitch movement/ pitch accents/stress	5		
Feedback on duration	4		
Reference video	4		
Spectrogram	3		
Formant graph	3		
Feedback on phonation	1		
Talking Heads	1		

Other types of displays concerning the production of the correct phone or syllable duration are arrows with additional written information (e.g., 'too long'/'short') (Eskenazi et al., 2000) and percentages of the duration (with regard to the reference speaker) in combination with a color code (red/green) (Wik and Hjalmarsson, 2009). SpeakGreek (Nicolaidis et al., 2015) offers a game-like experience to train duration. However, there is no real connection to Greek or any sound contrasts that might benefit from this exercise. The user is encouraged to produce any vowel or fricative sound over a period of a couple of seconds. But no sound of the world's languages extends over this period of time. One possible application area for this type of exercise might be speech therapy – training the continuous articulation of specific speech sounds.

Another way of teaching pitch and stress is using resynthesis procedures to modify the learner's speech on the basis of suprasegmental features (f0, duration, intensity) of a reference speaker. Only a couple of systems include this feature or investigated the effect on pronunciation learning (Meron and Hirose, 1996; Hirose et al., 2003; Martin, 2004, 2010; Bissiri and Pfitzinger, 2009). This is not surprising, since resynthesizing the learner's utterance might result in artificially sounding speech and thus should be treated with caution when implementing it as a feature in a CAPT system.

Another visual feedback method that is still used frequently, is displaying the wave form of the reference speaker and of the learner (Hirose et al., 2003; Jokisch et al., 2005; Demenko et al., 2009; Martin, 2004, 2010; Eye-Speak, 2016; Rose Medical Solutions, 2016; Rosetta Stone, 2016). It is easy to display but very difficult to interpret. First, the learner ideally needs knowledge of phonetics in order to interpret the oscillogram. Second, even with a phonetic background a wave form is not helpful in most cases. It gives the impression that the wave forms of the native speaker and the learner need to look the same in order to be correct. However, it most likely will not even look the same when the reference speaker produces the same utterance twice. The learner's utterance can still be perfectly fine, even though the oscillogram looks different. By trying to match both wave forms, the learner might produce a word or utterance using random strategies that might lead to a bad or even worse pronunciation than before. What is even more difficult to interpret is the display of the spectrogram, which is incorporated in some systems, even commercial ones (Jokisch et al., 2005; Wik and Hjalmarsson, 2009; Rosetta Stone, 2016).

Talking of feedback methods which are difficult to interpret, displaying formant graphs might only be helpful for learners with a phonetic background, who know what formants are and how they correlate to vowel production. Three systems include this type of feedback (Jokisch et al., 2005; Demenko et al., 2009; Nicolaidis et al., 2015). However, it is not argued that feedback on formants is not useful for the learner in general. But an effort has to be made to incorporate it in a way that any learner can understand and interpret. For example SpeakGreek (Nicolaidis et al., 2015) developed a game-like environment. The vowel targets are displayed as fields of flowers and the learner's production controls the position of a butterfly. By producing the correct vowel, the butterfly will move to the correct field of flowers and a larger flower will bloom up. However, the graphical user interface still includes information such as F1/F2 which might be confusing. Another attempt to teach vowels without any indication of formants is made with the vowel training software Hüpf that is introduced in chapter 3.

SpeakGreek (Nicolaidis et al., 2015) implemented another approachable game-like application in order to train learners with respect to phonation. Two children sitting on a seesaw will either move up or down depending on the production of voiced or voiceless stops. This might be in particular useful for phonation contrasts of fricatives and, especially, for use in speech therapy. None of the other systems included particular feedback on phonation (with the exception of a different display of vocal cord behavior in the (animated) vocal tracts).

Lastly, one system (Ville) included a talking head as a teacher surrogate (Wik and Hjalmarsson, 2009) which gives feedback to the learners. The talking head itself can actually not be considered as feedback, but it was still included to show the possibilities a learner can have when training with a CAPT system. An evaluation of the system shows that learners training with Ville seemed to have enjoyed the talking head feature.

The following chapter discusses a novel approach to vowel training in German. Learners often show problems in the correct pronunciation of German vowels, which are characterized by vowel length differences, distinction in tenseness, and vowel quality. For this reason, a game-like learning platform was developed to help learners improve their pronunciation with respect to German vowels.

Chapter 3

French Vowel Trainer

There is evidence that French learners of German show difficulties in the correct production of German vowels. Zimmerer and Trouvain (2015b) investigated the perception of French speakers' production of German vowels by German native listeners. They performed an identification experiment using German minimal pairs. Results indicated that learners showed problems producing German vowels correctly. Although advanced learners' productions were identified more often correctly than productions of beginners, both groups showed lengthening as well as shortening errors. Interestingly, rounded vowels seemed to cause more difficulties than unrounded vowels. In addition, Jouvet et al. (2015) created phone confusion matrices allowing for a comparison of the manually corrected annotation of the produced sounds in the IFCASL corpus (see chapter 4) with the automatic alignment of the expected sounds. An analysis of these confusion matrices revealed that French learners of German showed complex interferences with vowel contrasts for length and quality.

Within the scope of a software engineering project at the Department of Informatics at Saarland University, which is obligatory for informatics students, employees of the university can propose a software they would like to have developed. My colleague Frank Zimmerer and I suggested a vowel learning environment for learners of German, which was developed by seven informatics students within one semester[4].

The vowel trainer called *Hüpf* is a prototype which was not yet tested for its effectiveness in pronunciation training. The purpose of the software is to help the user to learn the correct pronunciation of vowels and difficult vowel

[4] A very special thanks to Christian Becker, Elizabeth Pich, Florin Foss, Kevin Müller, Lukas Stemmler, Ngoufack Yol Daniel Hilaire, Sven Ziegler

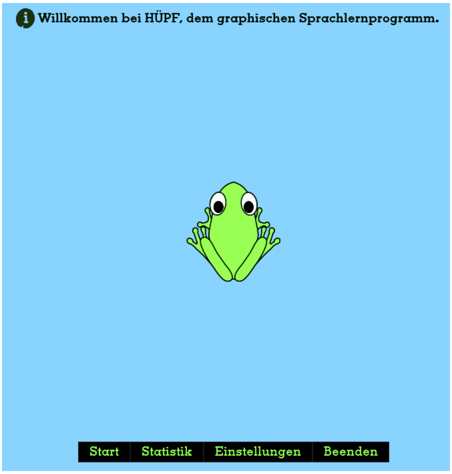


Figure 3.1: Welcome screen of the *Hüpf* software. From this screen the user can start the training, look at the overall statistics to see which vowels have been trained so far and how good the pronunciation is, and change a few settings

contrasts in German, but could be extended to other languages and sets of vowels. The user receives automatic visual feedback from the software on how well the vowel was produced in terms of formant values and duration. The speaker also has the possibility to play back a reference recording produced by a German native speaker and the users own recorded utterance to compare the productions of the words.

The name of the software *Hüpf* is based on the graphical user interface (GUI) of a frog that jumps to specific locations of water lilies in a pond. The training software was intendendly created as a game. This way, both adults and children would hopefully like to use the software as it is both helpful and entertaining. Simultaneously, feedback can be presented in an intuitive way so users without any background in phonetics can use the system equally well as users with a background in phonetics. Each water lily represents mean values of F1/F2 coordinates which were fed into the system. Figure 3.1 shows the welcome screen of the software. From this screen the user can start the training, look at the overall statistics to see which vowels have been trained so far and how good the pronunciation is, and change a few settings.

Concerning the settings (see Figure 3.2), the user can change the volume, choose a different language (currently, only German and English are avail-

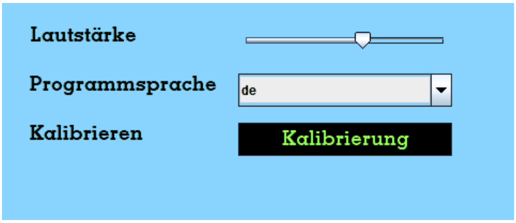


Figure 3.2: Settings of the *Hüpf* software (loudness, GUI language, new calibration).

able. This setting only refers to the language of the GUI, not the language that should be trained.), and redo the calibration. Calibration is automatically done before starting the very first learning session and consists of recordings of /a i u/. These vowels were used because they exist in many languages and define a large vowel space. The determined formant values are used in order to normalize for the reference values which are saved in the system.

After the calibration, the user gets to the training screen where up to two vowels can be chosen and trained. The user is not restricted in the choice of the vowel contrast. Each vowel is displayed as a water lily and the position of each water lily is based on F1/F2 reference values. Users have the possibility to listen to random examples for each vowel by either a male or female German native speaker which might be useful if the learner is not familiar with the displayed phonetic symbols.

By clicking on 'Lektion starten' (start session), the user enters the training session (see Figure 3.4) which consists of two training units per vowel so far. However, the number of training units can be changed. Again, the location of the vowels or water lilies depends on their F1/F2 values. The user can listen to a recording of the word by a male or female German native speaker and can also record the word to receive feedback by clicking on the microphone symbol. If the frog 'lands' on the correct water lily, the production of the vowel was correct. If, however, the frog 'lands' on a different water lily or in the water, the production was incorrect (see Figure 3.5). The position of the frog represents the normalized F1/F2 values for the vowel of the recorded word. Currently, vowel recognition is not very advanced and no automatic speech recognition is integrated. The vowel of each word is detected with the help of the intensity contour of the complete word. It is assumed that the vowel of the stressed syllable shows the highest intensity. Based on the intensity peak of the vowel, vowel boundaries are estimated.

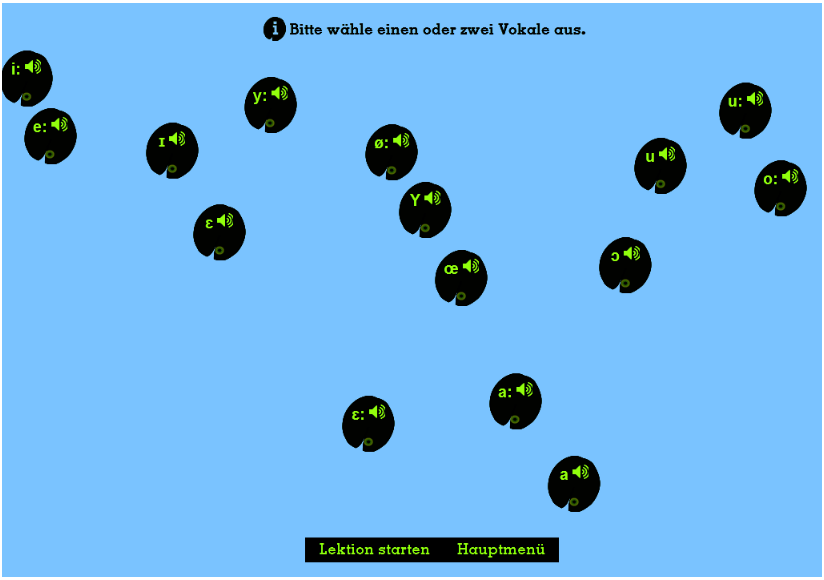


Figure 3.3: Vowel selection screen of the *Hüpf* software. Up to two vowels can be chosen and trained. The location of the vowels, displayed as water lilies, depends on their F1/F2 values.

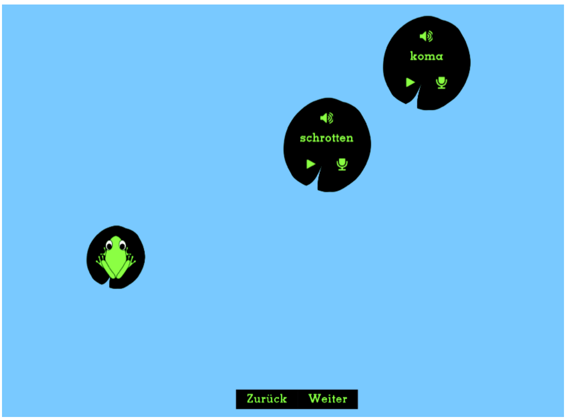


Figure 3.4: Training screen of the *Hüpf* software for the vowel pair /o: ɔ/.

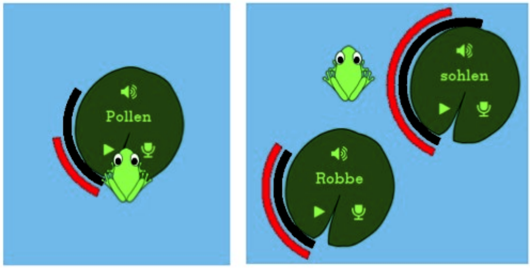


Figure 3.5: Training screen of the *Hüpf* software. Left: The frog landed on the correct water lily, the production of the vowel formants was correct. Right: The frog landed in the pond, the production of the vowel formants was incorrect.

Regarding vowel duration, two bars are displayed around each water lily. The black bar illustrates the reference length that was fed into the system and the red bar illustrates the length of the produced vowel.

After the training is completed, users will see a summary of their duration and pronunciation (F1/F2) accuracy as well as information about how many exercises they have completed for each vowel (see Figure 3.6). For an overview of all vowels, the button *Statistik* (statistics) on the home screen can be clicked. The statistics take into account how good the pronunciation was, how accurate the duration matched the reference value, and how often the vowel was practiced. The more the user practiced and the better the formant values and duration matched the reference values, the bigger the flower gets (see Figure 3.7).

After fixing some bugs in the software it would be interesting to see how well learners of German would be able to work with the software, if they find it intuitive and entertaining, and whether they will be able to improve their pronunciation of German vowels with regard to formant values and duration. Using a game-like pronunciation training system, users can learn intuitively. This way, no provision of phonetic information about the production of German vowels (and in this case formants), the German vowel system, and differences to the native language of the learner, is necessary. It is not argued that this information is not helpful, but it might overwhelm a learner without any phonetic background.

Other problems French native speakers face when learning German, and when German native speakers learn French, are discussed in the following chapter.

Figure 3.6: Individual result screen of the *Hüpf* software for the vowel /o:/ after training. Results are displayed for accuracy in length, accuracy in pronunciation (F1/F2), and number of trained items.

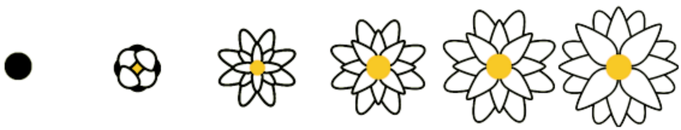


Figure 3.7: Flower states in the overall statistics screen. The better the pronunciation of the vowel, the bigger the flower.

Chapter 4

IFCASL Corpus

Within the IFCASL project (Individualized Feedback in Computer-Assisted Second Language Learning, funded by DGF and ANR) we developed a learner corpus of native and non-native speech for the language pair French-German[5] (Trouvain et al., 2013; Fauth et al., 2014; Trouvain et al., 2016a). This corpus was devoted to an in-depth analysis of both segmental and prosodic aspects of the non-native production of these languages. Whereas most studies or corpora have focused on English as a target language (cf. website on 'Learner corpora around the world', Goossens and Granger, 2016), only two existing spoken language corpora are available for the French-German language pair: the HABLA Corpus (Hamburg Adult Bilingual LAanguage) which consists of recordings of early French and German bilinguals (Kupisch et al., 2012), and the German part of IPFC-allemand (Interphonologie du Français Contemporain) which comprises recordings of advanced German learners of French (Chervinski and Pustka, 2010). The aims of designing this corpus were:

(I) Providing two native and two nonnative corpora for the French-German language pair. In addition, there are not many corpora who include recordings for a language pair in both directions, i.e. in this case French learners of German and German learners of French.

[5]This chapter is based on three papers describing on the corpus that was developed and recorded within the IFCASL project (Fauth et al., 2014; Trouvain et al., 2013, 2016a). The development of the corpus was collaborative work of the colleagues of LORIA, Nancy, France and Saarland University, Saarbrücken, Germany. I was mostly involved with the recording of the corpus data and the training and supervision of the student assistants that manually corrected the automatically aligned corpus. Additional examples and references were included.

(II) Performing analyses for phonetic and phonological research with respect to the prediction of the types of errors made by French and German learners.

(III) Utilizing the data for exercises in computer-assisted pronunciation training (CAPT) with a focus on feedback methods for the individual learner.

(IV) Supplying training and test data to improve automatic recognition of non-native speech which is known to be difficult (e.g., Goronzy et al., 2001; van Doremalen et al., 2009; Bouselmi et al., 2012).

First, a pilot corpus was recorded to test hypotheses about specific interferences between the native and non-native language of the speakers. The aim was to test the recording software and the automatic aligner regarding the performance of non-native speech alignment. The corpus was automatically aligned and only a small part was hand-corrected. Based on preliminary analyses and results, the second corpus was created with slight changes in structure and content. The complete corpus was then automatically aligned and more than 50% of the corpus was hand-corrected by French native speakers for the French part and by German native speakers for the German part. A special focus was placed on highlighting non-native pronunciation variants by the non-native speakers.

4.1 German-French Phonetic and Phonological Interferences

On a suprasegmental level non-native speech is generally characterized by reduced pitch range (e.g., Mennen, 1998; Ullakonoja, 2007; Hincks and Edlund, 2009; Busà and Urbani, 2011; Busà and Stella, 2012; Zimmerer et al., 2014), slower speaking/articulation rate (e.g. Raupach, 1980; Munro and Derwing, 1995b; Guion et al., 2000; Gut, 2009; Trouvain and Möbius, 2014b; Baese-Berk and Morrill, 2015; Jügler et al., 2016) and an increased number of pauses and disfluencies (e.g., Trouvain et al., 2016b; Trouvain and Möbius, 2014a; De Jong et al., 2015). With regard to segmental difficulties, a list of expected phonetic and phonological interferences is shown in Table 4.1 for which the IFCASL corpus can provide substantial empirical evidence. Although both systems are similar on a phonetic and phonological level (Delattre, 1964) there are small articulatory and acoustic differences between both languages (e.g., Strange et al., 2007; Pustka, 2011). Apart from the absence of certain phonemes in German and French, respectively, which can

Table 4.1: List of expected phonetic and phonological interferences in the French-German language pair.

French learners of German	German learners of French
Realization of /h/ and glottal stop /ʔ/	Liaison and enchaînement consonantique
Aspiration for /p t k/	Reduction of aspiration for /p t k/
Voiceless production of /b d g/	Voiced production of /b d g/
Realization of final devoicing	Realization of final voicing
Consonant clusters and affricates	
Realization of /ç, x/	
Postvocalic /r/ as lowered [ɐ]	Postvocalic /r/ as [ʀ]
Oral vowel + nasal consonant	Nasal vowels
Location of word stress	
Vowel quantity	
Vowel quality	
Realization and location of pitch accents	
Location of contrastive accents	
Reductions, elision, assimilations	
Mistakes induced by orthography	
Mistakes induced by cognates	

lead to pronunciation problems (e.g., /h, ç, x/ for French learners of German), they also share sets of phonemes that differ on a phonetic level which can lead to a foreign accented pronunciation. Some of the most frequent pronunciation problems are discussed in the following subsections.

4.1.1 Vowels

Figures 4.1 and 4.2 show the German and French vowel system, respectively. However, slight differences can be observed comparing French vowel systems by different researchers (cf. Tranel, 1987; Hammarström, 1998; Walker, 2001; Fagyal et al., 2006; Meisenburg and Selig, 2006; Pustka, 2011; Geckeler and Dietrich, 2012). Especially the open back vowel /ɑ/ and the nasal vowel /œ̃/ do not conform to the norm anymore (cf. Fagyal et al., 2006; Meisenburg and Selig, 2006; Pustka, 2011; Geckeler and Dietrich, 2012). In most varieties of French, /a/ and /ɑ/ do not create a contrast anymore and merged to /a/ (e.g., patte (paw) /pat/ vs. pâte (pasta) /pɑt/ are both pronounced as [pat]) . Similarly, /ɛ̃/ and /œ̃/ merged to the unrounded nasal vowel (e.g., brin (sprig) /bʀɛ̃/ vs. brun (brown) /bʀœ̃/ are both pronounced as [bʀɛ̃]) (Fagyal et al., 2006: 31pp.).

Regarding the German vowel system some inconsistencies were observed as well (e.g., Pustka, 2011 who does not distinguish between tense and lax

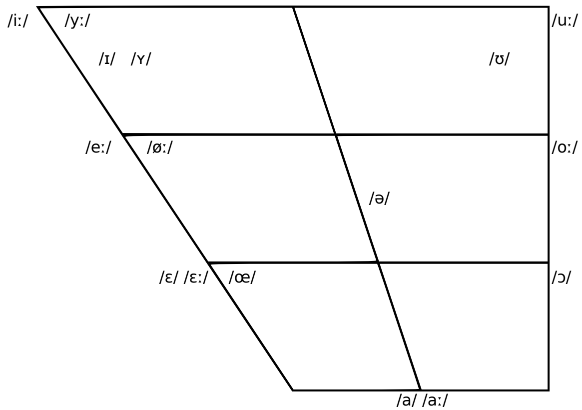


Figure 4.1: German vowel system (adapted from Wiese, 1996).

vowels on a phonological level). Most debated are the phonemic states of /ɛː/, /ə/ and /ɐ/. The long unrounded front open-mid vowel /ɛː/ does not exist in many German varieties and is often merged with /eː/ (e.g., Bären (bears) /bɛːʁən/ vs. Beeren (berries) /beːʁən/ are both pronounced [beːʁən]).

Due to these differences in the vowel systems of French and German, French learners of German have problems with the correct production of vowels. Vowel length is not distinctive in French (Walker, 2001; Meisenburg and Selig, 2006 but see Fagyal et al., 2006) but it is substantial in German, alongside tenseness (e.g., Meisenburg and Selig, 2006; Wiese, 1996). The examples in 1 and 2 illustrate this interference nicely. Possible additional phonetic problems were ignored in these examples. There are several occurrences in the corpus in which <Pollen> (pollen), which is produced with a short, lax [ɔ], was produced with a long, tense [oː] as <Polen> (polish people). This interference is crucial because it changes the meaning of the word. Example 2 shows that this problem exists in both directions in which <Polen> is produced as <Pollen>.

(1) Im Frühling fliegen Pollen durch die Luft.
(In spring pollen hurtle through the air.)

*[ɪm fʁyːlɪŋ fliːgən poːlən duɐ̯ç diː lʊft]

(2) Ich wüsste nicht, wie der schnellste Weg nach Polen ist.
(I wouldn't know the fastest way to Poland.)

*[ɪç vʏstə nɪçt viː deːɐ̯ ʃnɛlstə veːg nax pɔlən ɪst]

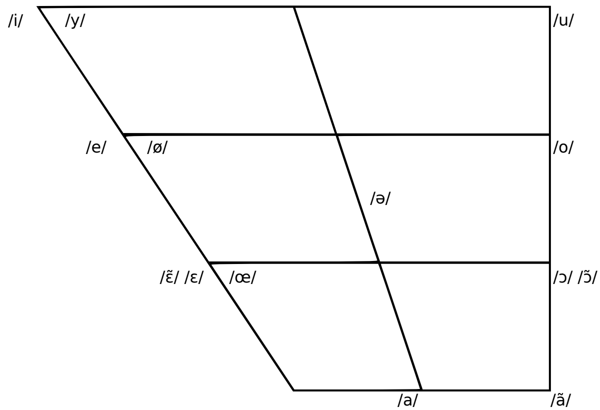


Figure 4.2: French vowel system (adapted from Pustka, 2011).

In contrast, German learners of French show problems in the correct pronunciation of nasal vowels. Also, high vowels /i y/ are produced even higher or more closed in French than in German. Whereas the open-mid vowels /ɛ, ɔ/ are produced more open which might lead to a reinforcement of the perceived foreign accent. Since French does not distinguish between tense and lax vowels, German native speakers might show difficulties in producing only tense vowels for certain syllable structures (e.g., muscle (muscle) *[mʏskl] instead of [myskl] (cf. Pustka, 2011: 99)).

Zimmerer and Trouvain (2015b) investigated the perception of French speakers' German vowels by German native listeners. They performed an identification experiment using minimal pairs. Results indicate that learners show problems producing German vowels correctly. Although advanced learners' productions are identified more often correctly than productions of beginners, both groups show lengthening as well as shortening errors. Interestingly, rounded vowels seem to cause more difficulties than unrounded vowels. Jouvet et al. (2015) created phone confusion matrices allowing for a comparison of the manually corrected annotation of the realized sounds in the IFCASL corpus with the automatic alignment of the expected sounds (for more information on annotation processes see section 4.3). An analysis of these confusion matrices revealed that French learners of German showed complex interferences with vowel contrasts for length and quality.

4.1.2 Stops and Final Devoicing

Both learners of German and French, respectively, show problems with the production of voiced and voiceless stops. As they share the same set of stops /b p d t g k/ which are realized differently on a phonetic level, these are particular difficult to produce correctly by both learner groups. French speakers differentiate between fully voiced stops and voiceless unaspirated ones with a short Voice Onset Time (VOT). In contrast, German shows a distinction in voiceless unaspirated stops with a short VOT and voiceless aspirated ones with a long VOT (e.g., Beyer, 1908; Künzel, 1977; Hammarström, 1998; Pustka, 2011).

This can induce communication problems as French learners of German show problems with the production of a long VOT when producing voiceless stops in German. Since French voiceless stops roughly coincide with the production of German voiced stops, the meaning of a word and therefore a whole sentence might change. In example 3 the word <Kasse> (register) might sound like <Gasse> (alley) to German native speakers. In contrast, the French word <gages> (wages) might sound like <cages> (cages) when produced by a German learner of French. Evidence for these interferences is shown in chapter 6 and 8.

(3) In jeder Bank gibt es eine Kasse.
(There is a register in every bank.)

*[ɪn jeːdɐ baŋk gɪpt əs a͡ɪnə gasə]

(4) Les gages sont payés à la fin du mois.[6]
(The wages will be payed at the end of the month.)

*[le kaʃ sɔ̃ peje a la fɛ̃ dy mwa]

The example of <cages> includes another problematic pronunciation for German native speakers. In German, voiced obstruents are produced voiceless at the end of a word or syllable (e.g., Wiese, 1996; Hammarström, 1998; Pustka, 2011). This phenomenon is called final devoicing which does not exist in French. For this reason, a word like <gages> will most likely be produced with the final voiceless fricative [ʃ] instead of [ʒ]. Looking at it the other way around, French native speakers usually produce a voiced obstruent at the end of a syllable. This error will not be as problematic in German – as there are no minimal pairs – as is might be in French (e.g., bague (ring) /bag/ vs.

[6]This example is not part of the corpus but an example from experiment II. It was used here to demonstrate the problematic pronunciation of stops as well as likely occuring final devoicing of /ʒ/ to [ʃ].

bac (case) /bak/). However, it will strongly contribute to a perceived foreign accent.

Bonneau (2015) and Bonneau and Cadot (2015) investigated the pronunciation of voiced /z ʒ/ in final position by German learners of French and French native speakers occurring in the French productions of the IFCASL corpus. Results showed that especially beginning learners of French produced voiceless fricatives. Furthermore, confusion matrices generated by Jouvet et al. (2015) to compare the manually corrected annotation of the realized sounds in the IFCASL corpus with the automatic alignment of the expected sounds showed that German learners of French show severe problems with obstruents in word-final position.

4.1.3 The sounds /h/, /ʔ/, /ç/

In German, /h/ belongs to the consonantal phoneme inventory (e.g., Kohler, 1995; Wiese, 1996) but does not exist in the French system (e.g., Tranel, 1987; Walker, 2001; Meisenburg and Selig, 2006; Pustka, 2011; Geckeler and Dietrich, 2012). The stereotypical concept that French speakers omit /h/ when speaking German is only partially true (e.g., Kamiyama et al., 2011; Neuhauser, 2012; Zimmerer and Trouvain, 2015a,c). Zimmerer and Trouvain (2015a,c) showed that beginners sometimes omit /h/ but tend to realize it as a glottal stop or other forms of glottalization. Advanced learners, however, realize /h/ more often and on a native-like level. Neuhauser (2012) showed that in 35 out of 44 cases, French native speakers produced a glottal fricative in German. Similarly, Kamiyama et al. (2011) found that out of 185 occurrences, 147 glottal fricatives were produced. Only in 31 cases a "null phonetic realization" (p. 1011) was observed. Interestingly, speakers show a tendency of overcompensation by inserting /h/ at various additional occasions (Zimmerer and Trouvain, 2015a,c). Example 5 shows that /h/ is sometimes inserted before words that start with a vowel. But at the same time, /h/ is omitted for <Haus> (house) for which it is obligatory. It is also noticeable that some speakers produce /h/ in <gehen> *[geːhən] (go/walk) which is not the standard German pronunciation ([geːn]) (Kleiner and Knöbel, 2015). This also constitutes a good example for an orthography induced pronunciation error.

Example 5 includes another expected pronunciation error by French learners of German concerning the realization of /ç/. This sound does not exist in the French phoneme inventory and will most likely be produced as [ʃ], which is the phonetically closest sound to [ç] which exists in the French phoneme inventory. However, this type of pronunciation mistake will not be considered as severely as others since speakers of some dialectal regions of Germany tend to pronounce /ç/ as [ʃ] (e.g., Saarland).

(5) Die riesigen Risse am Haus gehen nicht von alleine weg.
(The large cracks in the house won't go away on their own.)

*[diː riːsɪɡən rɪsə (h)am (h)a͡ʊs geː(h)ən nɪʃt fɔn (h)ala͡ɪnə vɛk]

4.1.4 Liaison and Enchaînement Consonantique

Another frequently made pronunciation error by German native speakers when speaking French is liaison which describes the realization of an otherwise silent consonant in final position before a word that starts with a vowel or approximant (e.g., Tranel, 1987; Hammarström, 1998; Walker, 2001; Meisenburg and Selig, 2006; Pustka, 2011). Example 6 illustrates this problematic. Usually, the word <mon> (my) is produced with a nasal vowel, the grapheme <n> is not pronounced. However, <mon> is followed by a word starting with a vowel and therefore has to be pronounced with the nasal stop (which becomes the onset of the following syllable) as [mɔ̃ nami].

(6) Aimez-vous mon ami?
(Do you like my friend?)

*[eme vu mɔ̃ ami]

Similar to liaison, enchaînement consonantique describes a process of resyllabification of the final consonant of a word becoming the onset of the following word if it starts with a vowel. The difference to liaison, however, is that no additional sound is realized (e.g., Tranel, 1987; Hammarström, 1998; Walker, 2001; Meisenburg and Selig, 2006; Pustka, 2011). Pustka (2011) described the difference between liaison and enchaînement consonantique with the examples shown in 7 and 8. Although resyllabification processes take place during both liaison and enchaînement consonantique, no additional sound is produced during the latter. Since neither liaison nor enchaînement consonantique exist in German, German learners of French have difficulty to apply these resyllabification processes.

(7) petite vs. petite amie

[pətit] vs. [pəti tami]

(8) petit vs. petit ami

[pəti] vs [pəti tami]

4.1.5 Cognates

Further sources of errors are cognates which are words with a similar etymological origin which often show the same spelling but have a different pronunciation (Ringbom, 2007). Two examples are shown in 9 and 10. Sometimes speakers tend to transfer the native production of these cognates to the foreign language.

(9) Le garçon a pris le car à <u>Berlin</u>.
(The boy has taken the bus to/in Berlin.)

*[le gaʀçɔ̃ a pʀi le kaʀ a <u>bɛɐ̯liːn</u>]

(10) In <u>Berlin</u> zahlt man wenig Miete.
(In Berlin you pay little rent.)

*[ɪn <u>bɛʀlɛ̃</u> t͡saːlt man veːnɪç miːtə]

4.1.6 Lexical Stress

In addition to numerous pronunciation errors on the segmental level, mistakes on the suprasegmental level include the location of the correct lexical stress in German which is not fixed (e.g., Kohler, 1995; Wiese, 1996; Féry, 1998). In contrast, French does not have lexical stress and accents are regularly placed on the last syllable of a phrase with a full vowel (e.g., Di Cristo, 1998; Féry, 2001; Jun and Fougeron, 2000; Pustka, 2011; Féry, 2014). Zimmerer et al. (2016) investigated the influence of L1 prominence on L2 production for beginning German learners of French and French learners of German. They used trisyllabic cognates with different word stress positions (first, second, and third syllable in German, see examples 11-13).

(11) Albatros (albatross)

[ʻalbatʁɔs] vs. [albatʀos]

(12) Embargo (embargo)

[ɛmʻbaɐ̯go] vs. [ɑ̃baʀgo]

(13) Labyrinth/labyrinthe (labyrinth)

[labyʻʁɪnt] vs. [labiʀɛ̃t]

Table 4.2: Number of German and French native speakers pooled across L1, L2, level of proficiency and age (Fauth et al., 2014).

# subjects	L1	L2	level	age
20	French	German	beginners	18-30 years
20			advanced	
10			beginners	15-16 years
20	German	French	beginners	18-30 years
20			advanced	
10			beginners	15-16 years

Results of a judgment task for native and non-native productions of both groups by the respective native listeners suggests that the correct suprasegmental structures in the respective L2 were not correctly acquired. Both groups were judged significantly worse regarding the correct placement of prominence in their non-native productions in comparison to native speakers.

4.2 Description of the Corpus

4.2.1 Speakers

Overall, about 100 persons were recorded: about 50 native French speakers and about 50 native German speakers. All speakers were recorded in their native (L1) and respective non-native (L2) language. To be able to distinguish between pronunciation errors that are discarded at an advanced language level and errors that are persistent even at a high level, beginners (A1-B1) and advanced (B2-C2) learners were recorded. For beginners, 20 adult speakers were recorded as well as 10 teenagers (15-16 years) (see Table 4.2 for an overview). Each group was balanced for gender. Most adult speakers were students or employees of the universities in Saarbrücken and Nancy. For recruiting the teenage learners, direct contact to secondary schools was made.

4.2.2 Questionnaire

Before recording the corpus material, each speaker was asked to complete a questionnaire which included information about the native language, age, highest educational degree and which second languages were learned (including duration). For each language, participants had to state 1) how often they had visited a country, in which this language was spoken 2) and for how long, 3) whether they speak this language with a friend or family member and 4) whether they had acquired a language certificate. In addition, they gave a 5) self-assessment of their language skills (listening, speaking, grammar, vocabulary) and stated 6) what motivation they have/had learning this language.

They were also asked 1) why they are/were learning French, 2) self-assess how talented they thought they were with regard to pronunciation, 3) which differences they thought exist between French and German regarding pronunciation and 4) if they thought computers could be helpful to learn a foreign language. An evaluation of the questionnaire showed a great variety of origin of the speakers and that, for most of them, English was the first L2 they learned.

Each participant signed an agreement that their spoken data can be used for scientific purposes. For teenagers, the agreement was signed by their parents.

4.2.3 Recordings

In contrast to other existing language learning corpora, this corpus includes speech recorded in the respective native and non-native language by each speaker. The advantage of such a corpus is that this design allows for within-subject and cross-language comparisons.

The recordings took place in quiet offices in Saarbrücken (Germany) and Nancy (France). It was decided against recordings made in a recording studio because it would not match the surroundings a learner is in when training with a computer-assisted pronunciation training (CAPT) system. Each recording session lasted about 60 minutes including completing the questionnaire.

The recordings were carried out using a head-mounted microphone (16 kHz, 16 bit) in an M-AUDIO Fast Track USB device. Recordings were saved on a Windows laptop using JCorpusRecorder (Colotte, 2013), a custom-made software developed at LORIA. Before starting the recordings, additional information had to be filled in to assign the correct file names (see Figure 4.3) and a test sentence had to be produced to be able to calibrate the microphone. For each recording, the speaker saw a single sentence presented on the screen. The recording was initiated when pressing the 'record' button and terminated when pressing the 'stop' bottom. The participants were able to record each sentence as often as they liked and listen to the recorded utterance. When they were satisfied with the pronunciation, clicking the 'next' button showed the next sentence. If the speaker spoke too loudly or too

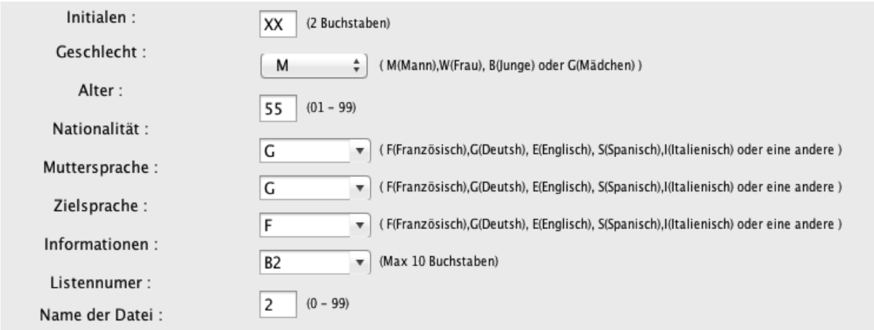


Figure 4.3: Information screen of the JCorpusRecorder (Colotte, 2013) before recording. These information had to be filled in for each speaker for each recording condition to create the correct file names.

softly, an error message indicated the problem and the recording had to be repeated.

The material of the IFCASL corpus was designed to cover many linguistic phenomena for the French-German language pair: 1) Coverage of all phonemes of both languages, 2) Coverage of the most important phonetic and phonological phenomena of German and french as a foreign language (see Table 4.1), 3) Minimal pairs, cognates (e.g., 'restaurant'), numbers and abbreviations.

The main content of the corpus consists of read speech. However, one sub-corpus also includes semi-spontaneous speech. The corpus includes four different recording conditions for each language: 1) SR (Sentence Read), 2) SH (Sentence Heard), 3) FC (Focus Condition), and 4) CT (ConTe/storytelling).

The 'Sentence Read' condition consists of 31 sentences which had to be read aloud. The 'Sentence Heard' condition is similar to the 'Sentence Read' condition as learners had to read individual sentences aloud. But the difference was that they had to listen to a recording of the sentence by a native speaker prior to recording the sentence themselves. The 'Sentence Heard' condition consists of 29 sentences. The purpose of the task was to exclude or minimize orthography induced errors. The 'Focus Condition' includes two sentences that vary in stress with respect to the word in focus (see example 14). Before recording the sentence, the learner had to listen to a question produced by a native speaker and then produce the answer with a specific focus. The focused word was indicated with bold letters. This condition was included in the corpus because languages can realize accents in different ways (see subsection 4.1.6 Lexical Stress).

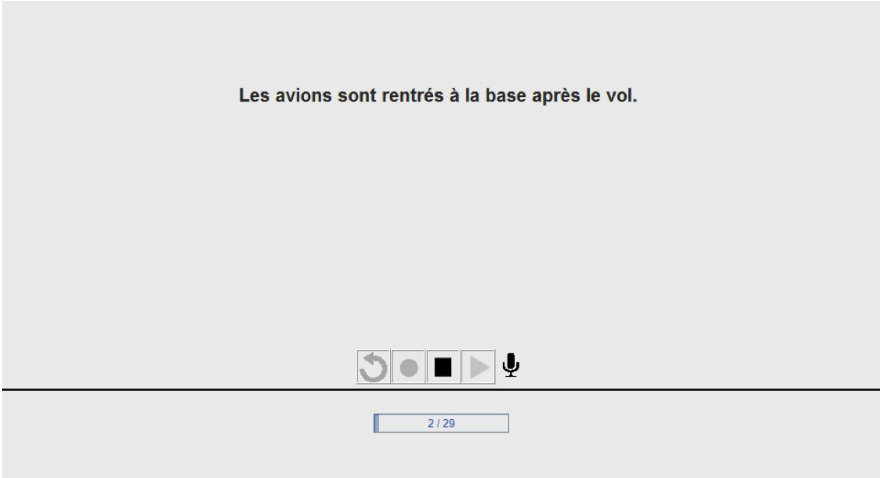


Figure 4.4: Recording window in the JCorpusRecorder (Colotte, 2013).

(14) Yvonne amène un ami. (Yvonne brings a friend.)

Yvonne amène un ami.

Yvonne **amène** un ami.

Yvonne amène **un ami**.

The story condition included the fairy tale 'The tree little pigs'. A short version of about 200 words for each language was created. Although the complete text was displayed in the recording window, a printout version was given to the participants for easier reading. This condition was included to investigate prosodic phenomena such as prosodic phrasing and prosodic phenomena on a larger scale.

4.2.4 Pilot Corpus

Before recording the main corpus, a preliminary pilot corpus was recorded. The corpus contained speech from 14 subjects (five adults and two teenagers for each language, not balanced for gender). The pilot corpus was recorded to test the technical performance of the recording software, the designed speech material, the usability of the questionnaire, and the duration of the recording session (see Fauth et al., 2014). To reduce the duration of the recording procedure for the main corpus, a second story and some focus sentences were excluded.

4.3 Annotation and Segmentation of the Corpus

The annotation and segmentation of the corpus was carried out as a two-step process. First, the speech material was forced-aligned using a speech-text alignment tool developed at LORIA, Nancy (Jouvet et al., 2011; Fohr and Mella, 2012) which uses a two-step approach for automatic phone segmentation. In the first step, the best phone sequence representation of the learner's utterance is determined by force-aligning the utterance with a model including several pronunciation variants. For better results, both native and non-native variants were considered. In the second step, phone boundaries were determined by applying forced-alignment. The hidden Markov models (HMMs) used within these processes were trained using speech of native and non-native speakers (for a more detailed description see Fauth et al. (2014)). After the force-alignment process was completed, trained student annotators manually corrected the majority of the corpus. Annotators only worked on their own native language (French or German). They marked insertions, deletions, and substitutions as well as (de)voicing processes on a phone level and re-set phone and word boundaries if necessary. Overall, more than half of the corpus was manually corrected. The non-native parts of the corpus were manually re-annotated to 80% and the native parts were corrected to 60% percent for French recordings and 25% for German recordings.

Figure 4.5 shows an example from the corpus with all six annotation tiers. The TextGrids include information on a phone, word, and sentence level.

(1) <u>Real</u>

Manually corrected annotation of the individual phones including shifting of phone boundaries of the realized phones. Deviations (e.g., phone insertions, deletions, and substitutions) from the forced-aligned (expected) phones were marked using a set of diacritics adapted from the Kiel corpus (IPDS, 1994). Before corrections are applied this tier is an exact copy of the second tier *Align*.

(2) <u>Align</u>

Phones with the best match taken from the canonical form (including pronunciation variants). Word boundaries were forced-aligned using HMMs.

(3) <u>Word</u>

Orthographic transcription of the individual words of the utterance. Word boundaries were corrected manually if necessary.

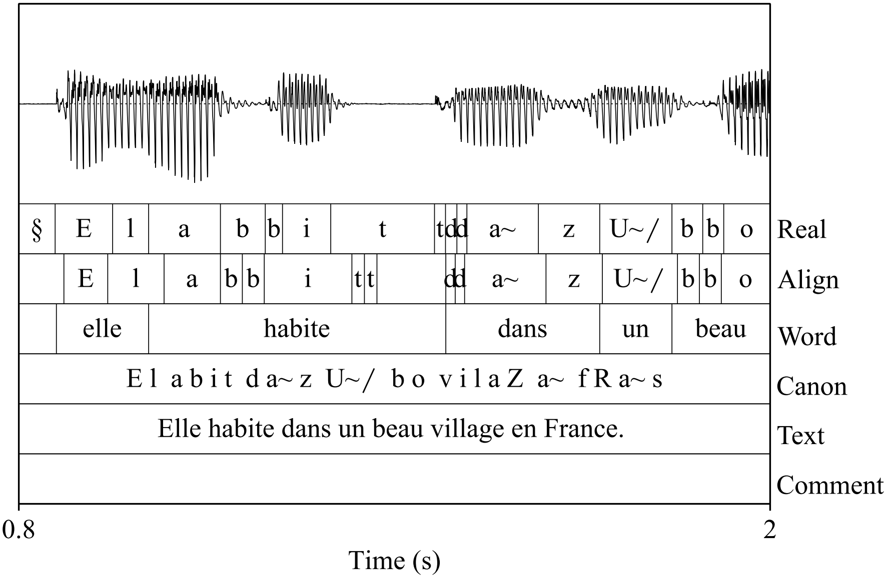


Figure 4.5: Example of a French recording by a German native speaker with the six annotation tiers after correction.

(4) Canon

Canonical form (including pronunciation variants) of the utterance based on French and German versions of SAMPA. Foundation for the forced-alignment process.

(5) Text

Orthographic transcription of the utterance.

(6) Comment

This tier was used to include additional information about the pronunciation that was not appropriate to include in the first tier *Real*. Unusual sounds, noises, or prosody related comments (e.g., incorrect lexical stress) were marked on this tier.

Chapter 5

Experiment I: Prosody Transplantation

This chapter[7] discusses an approach of prosody transplantation from recordings of German native speakers on utterances by French learners of German adapted from Boula de Mareüil and Vieru-Dimulescu (2006). The conducted perception experiment investigated the effect that this manipulation process has on the perceived foreign accentedness judged by native German listeners.

In general, evidence suggests that L1 speech is produced faster than L2 speech (e.g., Munro and Derwing, 1995b; Guion et al., 2000 regarding mean utterance duration and Raupach, 1980; Gut, 2009; Baese-Berk and Morrill, 2015; Trouvain and Möbius, 2014b regarding speaking/articulation rate). Munro and Derwing (1995b) were able to show that Mandarin learners of English produced significantly longer sentences than native English speakers. However, Flege (1979) found that native Arabic speakers produced English utterances at an equal length as American English native speakers.

Guion et al. (2000) investigated the relation between the age of learning and utterance duration. They tested native Italian speakers who came to Canada at different ages and found that sentence durations were significantly longer for speakers who arrived at a later age than for speakers who arrived at an early age who produced shorter durations which were comparable to utterances by native English speakers. These results could be transferred to native Korean speakers which suggests a generalizable effect of age. In the

[7] This chapter was first published in its full form as a proceedings article by Jügler et al. (2016). It was extended by an acoustic analysis of speaking rate and pitch range for French native speakers as well as a more detailed analysis of the perception experiment.

context of speech rate Baese-Berk and Morrill (2015) found that non-native speakers showed a slower speaking rate than native speakers but were also highly variable in their speaking rate from utterance to utterance. However, they point out that inconsistencies found in their study might be a consequence of the two tested native languages Korean and Mandarin Chinese. Munro and Derwing (1994) investigated the difference of read and extemporaneous speech between Mandarin learners of English and native English speakers. They demonstrated that native Mandarin speakers showed a slower speaking rate than English native speakers in both conditions. Trouvain and Möbius (2014b) found that for both French learners of German and German learners of French, L2 speech was produced with a lower articulation rate than speech by native speakers. However, they report that individual habits in articulation rate in the L1 were only partially transferred to L2 speech.

Riggenbach (1991) conducted an experiment to extract features that are tied to the perception of fluent non-native speech in comparison to non-fluent speech. After a fluency rating and an analysis of recorded dialogues it was concluded that speech rate and frequency of unfilled pauses are connected to judgements of nonfluency.

Regarding pitch, evidence suggests that languages show a characteristic use of pitch range and the alignment of pitch accents (e.g., Dolson, 1994; Mennen et al., 2012; Andreeva et al., 2014a,b, 2015, see Féry et al., 2011 for a comparison of German and French). Concerning pitch in an L2, a number of studies showed that learners have difficulties concerning (global) long term distributional pitch profiles reflected by differences in pitch range and the correct alignment of pitch accents. It was shown that producing the correct pitch range is hard for L2 learners (e.g., Mennen, 1998; Ullakonoja, 2007; Hincks and Edlund, 2009; Busà and Urbani, 2011; Busà and Stella, 2012; Zimmerer et al., 2014). For example, Finnish learners of Russian have been found to realize smaller pitch ranges in comparison to native speakers (Ullakonoja, 2007). Similar results could be replicated for Dutch learners of modern Greek (Mennen, 1998) and French learners of German as well as German learners of French (Zimmerer et al., 2014). However, Zimmerer et al. (2015) showed in a follow-up investigation of German learners of French and French learners of German that no pitch range differences occurred in comparing native and non-native speech for both speaker groups. In this experiment only short sentences were analyzed whereas in Zimmerer et al. (2014) the biggest differences in pitch range were shown for short stories. Also, the number of analyzed speakers was different (14 in Zimmerer et al. (2014) and 84 in Zimmerer et al. (2015)).

Overall, the role of prosody in what is perceived as a foreign accent has rarely been studied. Boula de Mareüil and Vieru-Dimulescu (2006) applied a prosody transplantation paradigm on Spanish and Italian native utterances which transfers phoneme duration and pitch contours of one language to another. A perception experiment was conducted to understand what is perceived by native listeners when combining the segmental specification of a synthesized utterance with suprasegmental features of a different language. They found evidence that listeners were more influenced by prosody than by phonemic features in assessing a foreign accent in the case of synthesized speech. Regarding modified natural speech, listeners are equally influenced by segmental and suprasegmental features. However, prosody transplantation was only applied on recordings of native speech, i.e. applying prosodic cues by a native speaker from one language to segmental information by a native speaker from another language.

The question arises whether manipulation of suprasegmental features is helpful to reduce the perceived accentedness of non-native speech by L2 learners. This question was addressed by recent studies. Ulbrich and Mennen (2015) investigated Belfast English native speakers and German learners of English with and without previous exposure to the Belfast English (BE) accent. They found that manipulating German accented utterances with the prosodic features of BE speakers lead to a reduced foreign accent rating for both German groups. They also found that transplanting L2 prosody on BE segmental information led to an increase of perceived accentedness.

Winters and Grantham O'Brien (2013) investigated prosodic transplantation on accentedness and intelligibility. They recorded German and English sentences of English native speakers with a high proficiency in German and German natives with a high proficiency in English. Manipulation was carried out for duration only and duration in combination with F0. In general, they found that manipulated sentences received higher foreign accent ratings and lower ratings of intelligibility. However, when applying native English prosody (duration and F0) to non-native productions, perceived accentedness ratings decreased. And when applying native German prosody to non-native utterances also decreased accent ratings but to a weaker degree.

Similar to the previous study, Rognoni and Busà (2013) examined speech by Italian learners of English and English native speakers. They transplanted native prosody on non-native segments and non-native prosody on native segments manipulating duration and pitch individually and in combination. They showed that manipulating both parameters yielded the strongest reduction in accent rating.

Furthermore, Jilka (2000) claims that the most important prosodic factor in the perception of foreign accent is intonation. He showed that listeners were able to successfully distinguish between American English and German low-pass filtered stimuli. However, listeners were significantly worse judging low-pass filtered stimuli with monotonous intonation. Also, foreign accent ratings of resynthesized non-native utterances with native intonation showed that the manipulated versions received lower foreign accent ratings.

Most of these presented studies focus on English as a target language. The study discussed here investigated the perceptual effect of L1 prosody transplantation on L2 speech in the case of French accented German.

5.1 Experiment

A perception experiment was conducted to test how strongly German native listeners perceive the accentedness of German utterances produced by French learners of German with a basic knowledge (A1-A2 level according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)). As a control condition, utterances produced by German native speakers were included. There were two versions of the sentences produced by the French learners: 1) the original sentences, 2) the same sentences manipulated for syllable duration and pitch based on one male and one female German native speaker. The research question of the perception experiment was whether the manipulated utterances received lower accentedness ratings than the original French accented utterances.

5.1.1 Material

The material for the perception experiment was taken from the IFCASL bilingual learner corpus (see chapter 4 and Table A.4 and A.8 in appendix A). Recorded sentences of the story of 'The three little pigs' of both French learners of German and German native speakers were used in this perception experiment. Ten French learners of German, five male and five female speakers with a basic knowledge of German, as well as ten German native speakers, five male and five female speakers, were selected. The story contained 13 sentences in total which differed in lengths (14-38 syllables, $\bar{x}=$ 23.5 syllables) and difficulty of words (e.g., the word *Schornstein* (chimney) is considered to be a difficult word for French learners of German).

As a first step, all recordings were labeled for disfluencies (e.g., hesitations, repetitions) as well as pauses that appeared after disfluencies. With the help of a Praat (Boersma and Weenink, 2013) script, the hesitation parts of these disfluencies and pauses were automatically removed to allow for a correct prosody transplantation process and in case they might interfere with the accentedness rating.

In order to apply the prosody transplantation, i.e. manipulation of syllable duration and pitch, a Praat script was used that was originally applied in a variety of studies by Boula de Mareüil and colleagues (e.g., Boula de Mareüil and Vieru-Dimulescu, 2006; Kaglik and Boula de Mareüil, 2010). This technique extracts and transplants phoneme by phoneme duration and pitch with the help of the PSOLA (Pitch Synchronous Overlap and Add) algorithm (Moulines and Charpentier, 1990). It was decided to manipulate syllables instead of phonemes because the French learners of German often deleted or inserted phonemes. Due to the incremental procedure of the script it could not be ensured that duration and pitch of the same phonemes were matched correctly. However, a correct syllable matching was much easier to obtain by checking for the same number of syllables before applying the technique. However, a few manipulated sentences sounded obviously odd. Unfortunately, this could not be resolved even after checking for correct syllable boundaries and syllable matching.

The manipulation of duration and pitch (see Boula de Mareüil and Vieru-Dimulescu, 2006) was narrowed down to the following four-step procedure:

(1) Calculating duration coefficients for each syllable or pause of a speaker with respect to the male or female model speaker.

(2) Replacing the original syllable durations for each syllable and pause.

(3) Calculating F0 coefficients for each syllable of a speaker with respect to the male or female model speaker.

(4) Replacing the F0 values for each syllable.

One male and one female German native speaker was chosen from the set of ten speakers to manipulate the French accented utterances. Male French speakers were manipulated on the basis of the male German speaker and female French speakers on the basis of the female German speaker. To decide which German speakers to use for the manipulation, mean speaking rate was calculated for each German speaker and the speaker with the median value was chosen.

To decrease the length of the experiment, only the first six sentences of the story were used. Overall, the experiment consisted of 180 trials: 60 French accented utterances without disfluencies, 60 manipulated utterances, and 60 German native utterances.

5.1.2 Acoustic Analysis of the Material

Before conducting the perception experiment, speaking rate and pitch range were extracted automatically from the utterances by the German native speakers and French learners of German using different Praat scripts to examine the difference between native and non-native utterances. In order to interpret the results for pitch range and speaking rate correctly, recordings of the French learners of German when speaking their native language were also analyzed. To calculate pitch range, minimum and maximum values were extracted using the recommended Praat pitch range settings of a floor of 75 Hz and ceiling of 300 Hz for male voices and 100 Hz and 500 Hz for female voices. To allow for cross-gender comparisons, Hz values were normalized by converting them to semi tones (st). The conversion was performed with the following formula (cf. Reetz and Jongman, 2009):

(1) $\text{Range} = 12 \times log_2(\text{max}f0/\text{min}f0)$

Data analysis was performed using JMP 12 (JMP, 2014) for all tests. For each of the parameters SPEAKING RATE and PITCH RANGE were entered as a dependent variable, SPEAKER and SENTENCE were included as random factors, and L1/L2 (German native, French native, French learners of German) and GENDER as fixed factors as well as their interaction.

Speaking Rate

The statistical analysis for speaking rate (including all sentence-internal pauses) showed that only L1/L2 showed a main effect ($F(2,37.71)=1003.86$, $p>0.0001$). Post-hoc tests showed that all contrasts were significantly different. Figure 5.1 shows that French native speakers speak faster than German native speakers when speaking in their respective native language (6 vs. 5 syllables/second, respectively). However, when they speak a second language, their speaking rate drops quite drastically to 3 syllables/second. This means French natives speak two times slower when speaking German.

Pitch Range

The statistical analysis for pitch range showed that all conditions and their interaction are significantly different (see Table 5.2). GENDER showed a main

Table 5.1: *Statistical information of the Linear Mixed Model for speaking rate.*

source	**between-groups df**	**within-groups df**	**F**	**p**
GENDER	1	15.78	0.00118	0.9147
L1/L2	2	37.71	1003.857	<0.0001*
GENDER×L1/L2	2	37.71	1.9456	0.1566

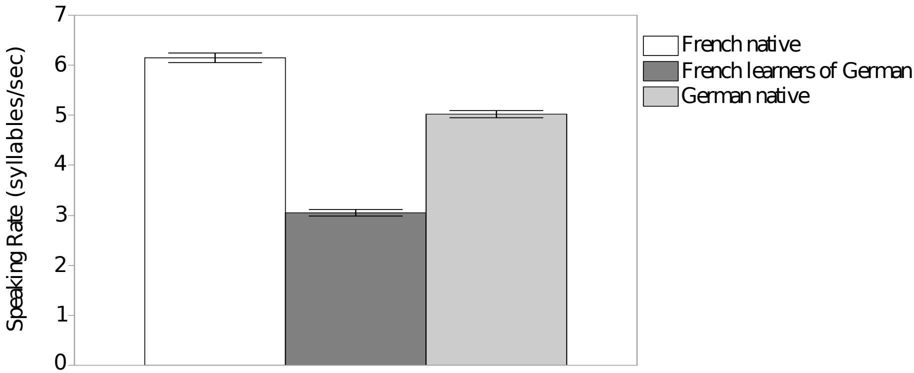


Figure 5.1: Mean values for speaking rate (in syllables per second) for male and female native German speakers, native French speakers, and French learners of German.

Table 5.2: *Statistical information of the Linear Mixed Model for pitch range.*

source	**between-groups df**	**within-groups df**	**F**	**p**
Gender	1	15.4	19.215	0.0005*
L1/L2	2	40.69	12.5204	<0.0001*
Gender×L1/L2	2	40.69	11.8983	<0.0001*

effect (F(1,15.4)=19.2, p<0.001) highlighting the difference in pitch range between male and female speakers. Women show a larger pitch range (19.9 st) than male speakers (14.6 st).

Futhermore, L1/L2 is significantly different (F(2,40.69)=12.52, p<0.0001) and post-hoc tests showed that pitch range is not different for German native speakers and French native speakers as well as German native speakers and French learners of German. However, as can be seen in Figure 5.2, French learners change their behavior in terms of pitch range by increasing it, which results in a significant difference in pitch range for French speakers speaking either in the native or non-native language.

Lastly, the interaction Gender×L1/L2 shows a main effect (F(2,40.69)= 11.9, p<0.0001) (see Figure 5.3). Post-hoc tests show that regarding female speakers, pitch range differs for German natives and French natives as well as French learners of German and French natives. This is evidence that French learners of German adjust to the pitch range differences between the lan-

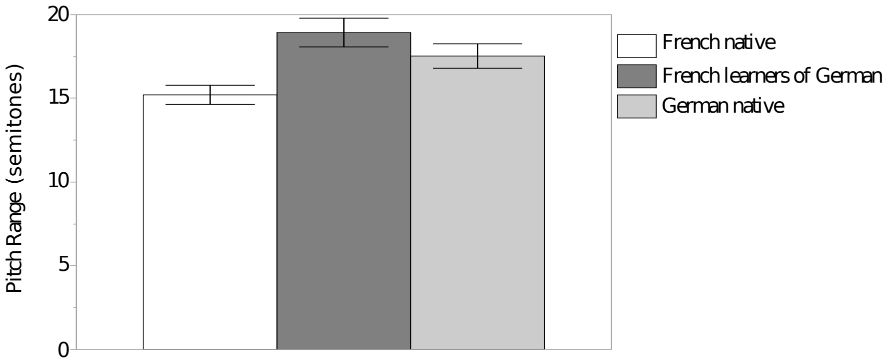


Figure 5.2: Mean pitch range (in semitones) for German native speakers, French native speakers, and French learners of German.

guages when speaking German. No difference was found between German natives and learners.

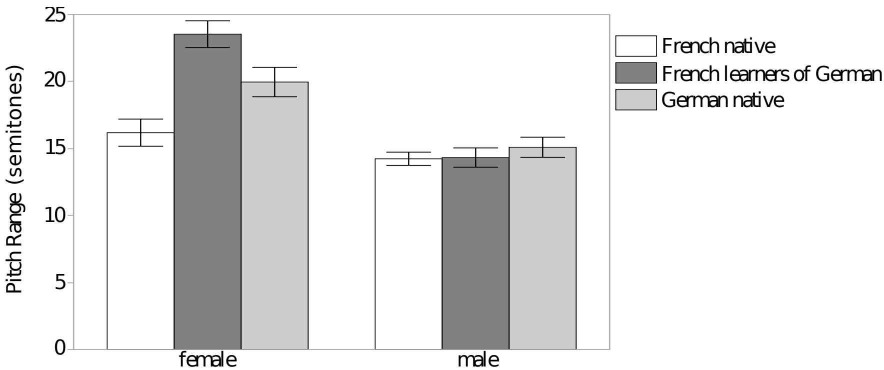


Figure 5.3: Mean pitch range (in semitones) for male and female German native speakers, French native speakers, and French learners of German.

Furthermore, French male and female speakers do not differ in regard to pitch range. However, there is a difference for German male and female speaker, with a higher pitch range for female speakers (20 st vs. 15.1 st). This difference can also be demonstrated for male and female French learners of German (female: 23.54 st, male: 14.3 st).

5.1.3 Procedure

The perception experiment was carried out as a PraatMFC experiment by ten German native listeners (four male, six female listeners from which six had a phonetic background). Participants were instructed to listen to each utterance with headphones and decide how accented each recording was (1 = not accented, 7 = heavily accented) and whether the recording sounded natural or artificial. Before starting the experiment, they received the information that some sentences were manipulated and might sound artificial. They were asked to ignore artificiality when rating accentedness. Apart from this, they did not receive any information about the language background of the speakers.

5.2 Hypotheses

Regarding foreign accentedness and naturalness ratings the following hypotheses are made:

(1) French accented recordings will receive higher foreign accentedness ratings (at the upper end of the scale) than recordings by German native speakers (lower end of the scale).

(2) Manipulated recordings will receive lower foreign accentedness ratings than the learner utterances due to prosody transplantation.

(3) Manipulated recordings will be rated more often as artificial due to occasional incorrect transfers of duration and pitch.

5.3 Results

5.3.1 Foreign Accentedness

A first analysis was carried out to test the extent to which the transplantation had an effect on perceived foreign accent. The responses were entered as a continuous variable into a linear mixed model in JMP 12 (JMP, 2014) with RATING as dependent variable, PARTICIPANT and ITEM as random effects, and SENTENCE (1-6), GROUP (native, learner, manipulated), SPEAKER GENDER, and all two-way interactions were entered as fixed factors. The analysis showed that GROUP ($F(2,154)=2534.86$, $p<0.0001$) and SPEAKER GENDER ($F(1,154)=13.67$, $p<0.001$) were significant factors. No other factor or interaction was significant (see Table 5.3).

Table 5.3: *Statistical information of the Linear Mixed Model for accentedness.*

source	**between-groups df**	**within-groups df**	**F**	**p**
SENTENCE	5	154	0.8196	0.5374
GROUP	2	154	2534.86	<0.0001*
SPEAKER GENDER	1	154	13.6785	0.0003*
SENTENCE×GROUP	10	154	0.6759	0.7456
SENTENCE×SPEAKER GENDER	5	154	1.3703	0.2384
GROUP×SPEAKER GENDER	2	154	1.8946	0.1539

Post-hoc tests for the effect GROUP showed that all three categories differ significantly from each other (see Figure 5.4). The original French accented utterances show a mean accentedness value of 5.78. It is noticeable that the manipulated versions were rated as significantly less accented than the original versions (4.98) but still received a considerably high score. However, there is still a large difference to the native sentences which were rated with a mean value of 1.24. It was also shown that male speakers received a significantly higher accentedness rating scores (4.01) than female speakers (3.86) (see Figure 5.5).

Although not included in the model, Figure 5.6 illustrates mean accentedness ratings for the individual speakers and that speakers were rated differently, e.g., speaker 511 received a considerably low mean accentedness score of 5.18 in comparison to speaker 503 who received a mean score of 6.25. Furthermore it can be seen that all French learners of German received lower scores after prosody transplantation was applied. There is also diversity in accentedness scoring for German native speakers. Taking a look at the individual behavior of the German listeners a large variability can be noticed as well (see Figure 5.7). For example, listener 02 tended to give, in general, lower ratings than listener 03.

5.3.2 Naturalness

The results of the perception experiment with regard to naturalness were entered into a Generalized Linear Mixed Model using JMP 12 (JMP, 2014) and the model was performed with a binomial distribution and Logit link function. NATURALNESS (0 or 1) was entered as dependent factor and LISTENER and STIMULUS were entered as random factors. The effects SENTENCE, GROUP, SPEAKER GENDER, and all two-way interactions were en-

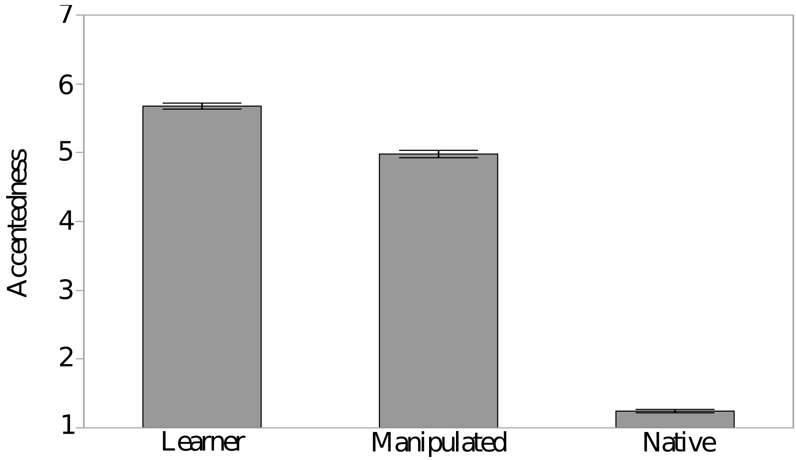


Figure 5.4: Mean accentedness ratings for native, non-native (learner) and manipulated utterances.

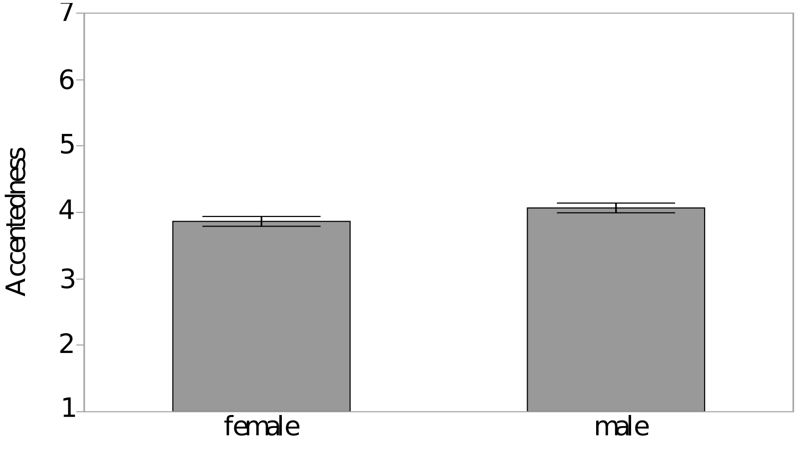


Figure 5.5: Mean accentedness ratings for male and female speakers.

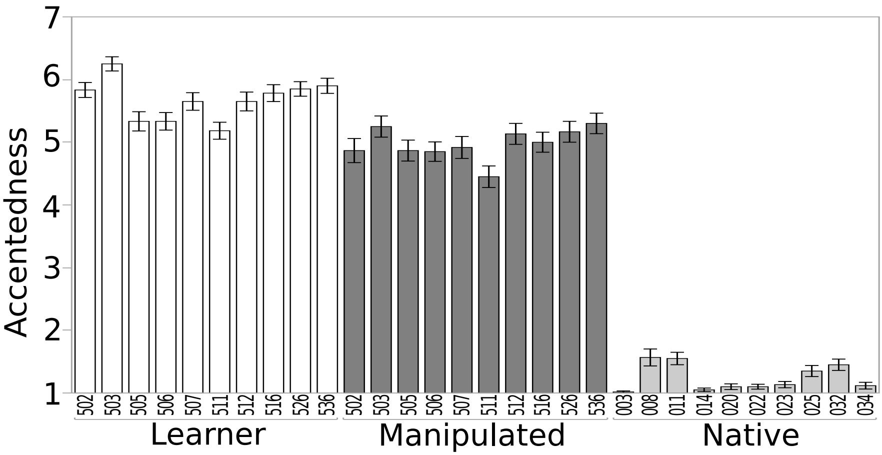


Figure 5.6: Mean accentedness ratings for the individual speakers of the three conditions.

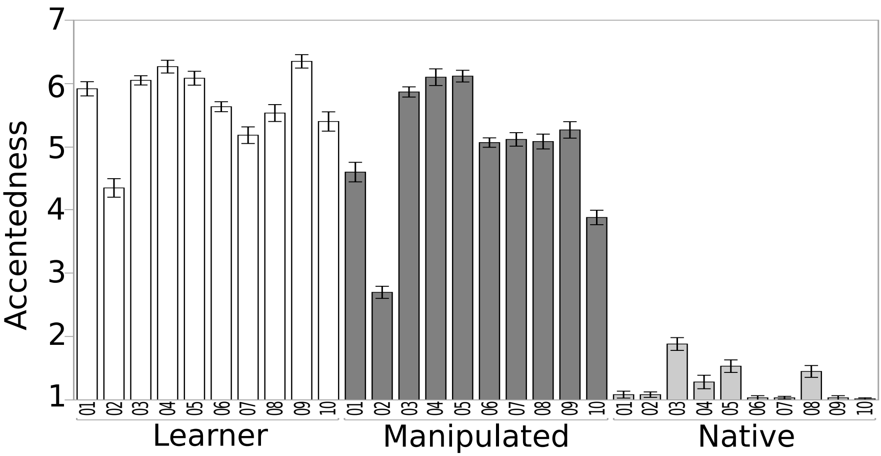


Figure 5.7: Mean accentedness ratings by the individual listeners for the three conditions.

Table 5.4: *Statistical results of the Generalized Linear Mixed Model with regard to naturalness.*

source	**df**	χ^2	**p**
SENTENCE	5	5.7057049	0.3359
GROUP	2	1110.0451	<0.0001*
SPEAKER GENDER	1	12.746291	0.0004*
SENTENCE×GROUP	10	14.246259	0.1621
SENTENCE×SPEAKER GENDER	5	5.8244338	0.3237
GROUP×SPEAKER GENDER	2	1.6198825	0.4449

tered as fixed factors. An overview of the statistical results can be found in Table 5.4.

GROUP shows a main effect ($\chi^2(2)=1110.05$, $p<0.0001$). Post-hoc tests showed that all three categories differed significantly from each other as can be seen in Figure 5.8. German native utterances were rated natural in 96.3% and French accented utterances in 91.7% of the cases. The lower naturalness rating for the accented sentences might be attributed to the fact that these were also manipulated in the sense that hesitations and pauses were removed. This might have influenced the perception of these utterances to a certain point. However, the manipulated versions were rated natural only 20.2% of the time which shows a drastic difference and that prosody transplantation needs to be improved in terms of a correct manipulation of duration and fundamental frequency. Furthermore, SPEAKER GENDER shows a main effect ($\chi^2(2)=12.75$, $p<0.001$). Figure 5.9 illustrates this result. Utterances by female speakers were rated natural in 73% and by male speakers in 65.8% of the cases

Figure 5.10 shows the mean naturalness ratings in percent for the individual speakers. It can be seen that for the French accented recordings, recordings by speaker 507 were often rated as artificial in comparison to the rest of the speakers. It might be the case that the recordings of this speaker sounded a bit off after removing hesitations and pauses. Furthermore, manipulated utterances by speaker 511 seem to sound less artificial than utterances by e.g., speaker 503. A large variability across speakers can be noticed. This however is an indication that automatic prosody transplantation can produce natural sounding utterances for some speakers.

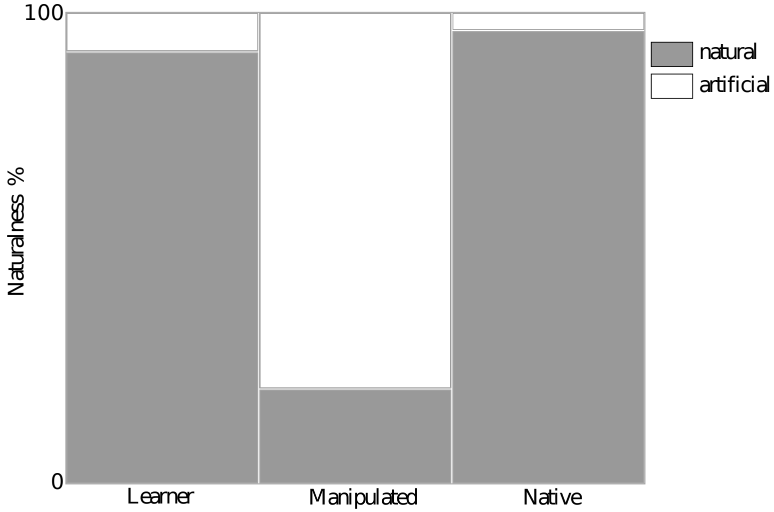


Figure 5.8: Mean naturalness ratings (in %) for native, non-native (learner) and manipulated utterances.

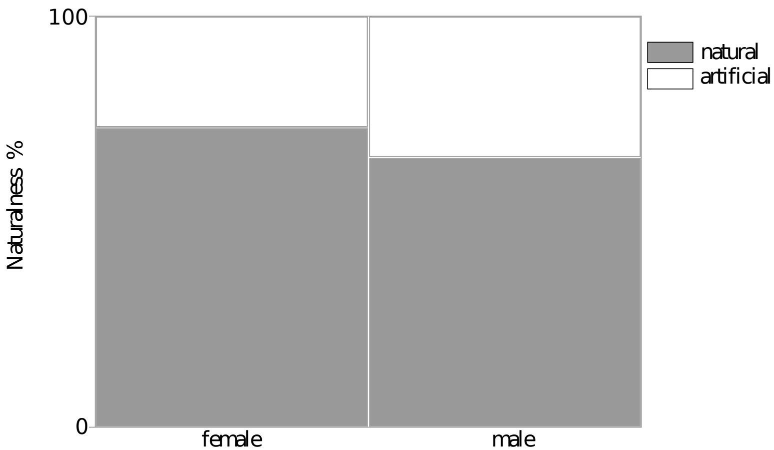


Figure 5.9: Mean naturalness ratings (in %) for male and female speakers.

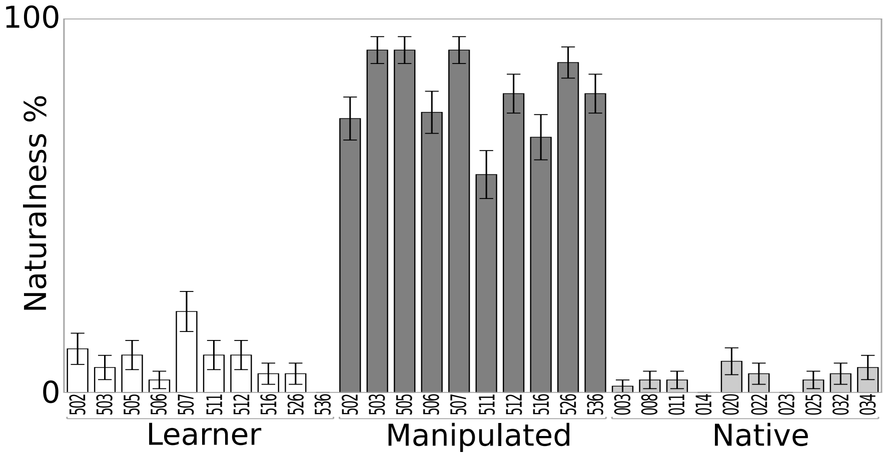


Figure 5.10: Mean naturalness ratings (in %) for the different speakers of the three conditions.

There is again a large discrepancy for German listeners who judged the recordings for naturalness. In particular listener 02 judged most of the manipulated recordings as natural sounding, in contrast to listener 08 who also rated many of the original French accented utterances unnatural.

5.4 Discussion

The material which was used for this experiment replicated earlier findings confirming that language learners produced L2 speech with a lower speaking rate than native speakers and that there is a difference in terms of pitch range between languages (female French speakers show a lower range than female German speakers). However, it was shown that female French learners of German were able to adjust their pitch range to the level of German native speakers when speaking German. Because pitch was extracted automatically, errors in the extraction of minimum and maximum pitch values can not be ruled out.

These differences led to an analysis of the degree of benefit the learners' productions can receive when native speakers' prosodic features are transplanted onto the learners' productions. By applying the technique of prosody transplantation by Boula de Mareüil and Vieru-Dimulescu (2006), syllable duration and pitch contours of a German male and female model speaker were transferred onto the non-native utterances of French learners of Ger-

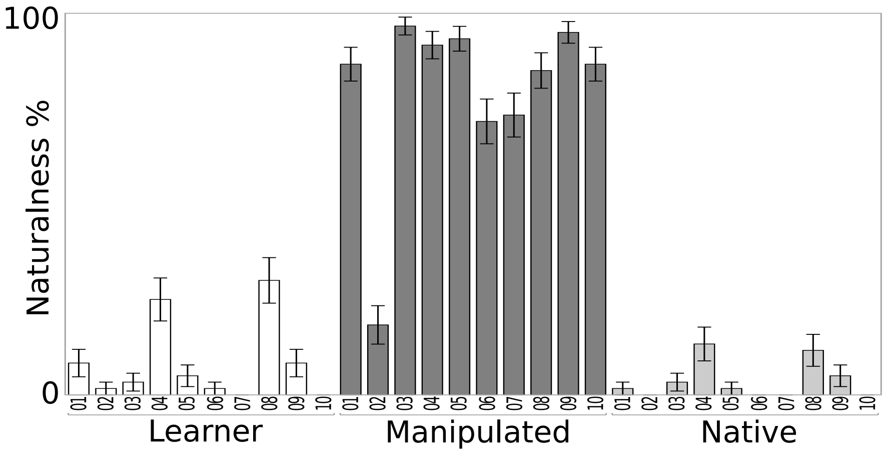


Figure 5.11: Mean naturalness ratings (in %) by the different listeners of the three conditions.

man. Results of a perception experiment suggest that the manipulation of pitch and syllable duration reduces the perceived foreign accent. However, listeners of the experiment rated the manipulated stimuli more often as artificial than unmanipulated stimuli. Nevertheless, it was shown that listeners still judged the items consistently as having less accent compared to original learner productions.

Other studies investigating the effect of prosody transplantation argued that although manipulation of (phoneme or syllable) duration and pitch contour has a beneficial effect on the perceived accentedness of non-native utterances, segmental information still has a strong influence (e.g., Winters and Grantham O'Brien, 2013; Rognoni and Busà, 2013; Ulbrich and Mennen, 2015). This perception experiment can be interpreted in a similar way. Although foreign accent rating decreased significantly for original productions of learners for manipulated utterances, the rating for manipulated utterances is still considerably high. This might be an impact of non-native segmental cues which were not manipulated.

Another explanation might be the influence of the manipulation procedure itself. Winters and Grantham O'Brien (2013) noted the decline in quality of the produced stimuli after applying the PSOLA synthesis algorithm which is also noticeable for this experiment. Manipulated productions were rated significantly more often as unnatural compared to native and unmanipulated items. Only 121 items (out of 600) were rated as natural. As a matter of fact, for some utterances the prosody transplantation created a strong artificial outcome, including odd pitch behavior as well as syllables with either too long or too short duration. But even if the prosodic cues were transplanted correctly, most of the manipulated utterances remained somewhat artificial.

Despite perceiving manipulated stimuli as less natural, utterances were consistently rated as having less accent compared to original learner productions. This means that transplanting syllable duration and pitch from a native utterance to a learner's non-native production does have a positive influence on the perception of non-nativeness.

Regarding feedback in second language learning, prosody transplantation might be considered to be a promising technique. Bissiri and Pfitzinger (2009) showed that resynthesizing the voice of Italian learners of German had a beneficial and motivating effect on learning lexical stress in German. Also, Henry et al. (2007) proposed a tool for analyzing, processing and visualizing a learner's speech to help acquiring the correct prosody of a foreign language. It would be interesting to see whether language learners find it helpful to hear their own voice manipulated for syllable duration and pitch contour, and whether they would be able to extract and implement useful information from the manipulation to their productions.

Meron and Hirose (1996) proposed a tool which focused on manipulation of the learner's own voice in the context of Japanese or English as a foreign language. They integrated a manipulation procedure that resyntheized the speaker's speech in a way that the wrongly pronounced syllable was manipulated based on fundamental frequency, duration, and intensity of a reference speaker. They argued a teacher would exaggerate wrongly pronounced syllables in a classroom situation to draw focus to the correction. For this reason, the system did not only transfer the prosodic features of a reference speaker, but also emphasized the syllable that contained an error.

Hirose et al. (2003) also developed a system for learners of Japanese. The focus of the system was to teach the learners the correct pronunciation of lexical accents. Both auditory and visual corrective feedback were included. Concerning the audio feedback, the learner's voice was resynthesized with the prosodic features of the reference speaker. Additionally, visual feedback in form of stylized pitch contours and arrows indicating falling and rising pitch was displayed for the original and the modified utterance in order to highlight the differences. Pronunciation training experiments showed that learners of Japanese improved the production of Japanese lexical accents after using the system.

Chapter 6

Experiment II: Exposure to a Native Speaker and Modified Voice

The previous experiment showed that manipulation of speech leads to a decrease in perceived foreign accentedness. This experiment[8] was conducted with German natives to test whether auditory feedback or training procedures are helpful for the learner to improve the pronunciation of French voiced and voiceless stops. German and French mark the distinction between voiced and voiceless stops /b d g p t k/ differently. French speakers differentiate between fully voiced plosives and voiceless unaspirated ones with a rather short Voice Onset Time (VOT). In contrast, German shows a distinction by voiceless unaspirated plosives with a short VOT and voiceless aspirated ones with a long VOT (e.g., Beyer, 1908; Künzel, 1977; Hammarström, 1998; Pustka, 2011). French learners of German and German learners of French most likely transfer phonetic knowledge of their respective L1 to the L2 production which can result in foreign accented productions and difficulties in intelligibility (cf. Flege, 1995; Best, 1994; Kingston, 2003).

Figure 6.1 illustrates this difficulty. The wave forms display the production of /b/ and /p/ by a German native speaker (top) and a French native speaker (bottom) with a similar right vowel context ([a]). Taking a closer look at the German /b/ and the French /p/ it is obvious that both sounds

[8]This experiment was first published in its full form as a proceedings article by Jügler and Möbius (2015). Within the frame of this dissertation it was extended by a more detailed acoustic analysis.

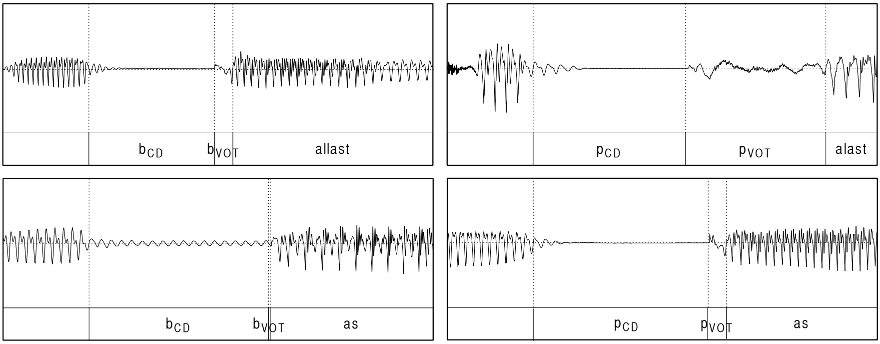


Figure 6.1: Illustration of German (top) and French (bottom) stops. Voiced stops are displayed on the left and voiceless stops on the right.

are produced similarly and show a short positive VOT which characterizes a voiceless unaspirated stop. Please note that although we are talking about a voiced stop in German on a phonological level, the phonetic realization is actually voiceless. Additionally, French phonologically voiced stops are produced with a continuous vocal fold vibration during the closure. German native speakers do produce this continuous vibration occasionally but it usually only occurs due to coarticulation effects, e.g., in the word <Laden> (store) as [la:dən] or [la:dn̩] (see also Möbius (2004) for voicing profiles of phonemically voiced stops in German). Because the phonologically voiced /d/ is enclosed by two voiced sounds, and onset and offset of vocal fold vibration is a slow process, phonation carries over to the stop sound and is therefore fully voiced. However, a conscious production of a voiced closure phase in spontaneous speech takes practice for a German native speaker.

Two different strategies were tested: manipulation of the speaker's own voice and the exposure to a French native speaker. Probst et al. (2002) argued that learners focus on several features when focusing on a native speaker's utterance. While one category can be described as linguistic features, the other is linked to individual features of the model speaker. By introducing the learner's own voice, the learner can disregard individual features and focus more on L2 features. However, regarding phonetic manipulation, manipulating only one phonetic feature might not be sufficient, i.e. a possible phonetic interaction might not be taken into account (e.g., longer vowel duration before a voiced stop in English). In this respect, listening to a native speaker might be more helpful. Additionally, a learner's production usually contains several mispronunciations which are not addressed in the manipulated version and are left uncorrected.

There are many studies which investigated manipulation of suprasegmental or segmental features and tested if manipulation is helpful to reduce the perceived accentedness of non-native speech (e.g., Boula de Mareüil and Vieru-Dimulescu, 2006; Felps et al., 2009; Zhao et al., 2012; Winters and Grantham O'Brien, 2013; Rognoni and Busà, 2013; Ulbrich and Mennen, 2015; Jügler et al., 2016). Most studies focused on manipulation of prosodic features due to the less complex modification process. Furthermore, manipulation was mostly carried out on recorded utterances which were subsequently rated for perceived accentedness. The training procedure proposed in this experiment includes an additional step: The manipulated version is presented to the learner and tested if useful information can be extracted from it.

Recent studies investigated the effect of manipulation of a learner's voice as a feedback method to improve pronunciation. Bissiri and Pfitzinger (2009) showed that resynthesizing the voice of German learners of Italian had a beneficial and motivating effect on learning lexical stress in German. They stressed that L2 prosody is often taught with the help of visual illustration of intonation. However, this approach might not be sufficient to teach lexical stress because it comprises a combination of acoustic features (F0, duration, and intensity). They also highlight that recasts by a native speaker might also not be adequate since the speech of a native speaker and a learner shows a number of differences (e.g., voice quality, segmental differences). These differences might be confused with cues that are crucial for the production of the correct stress pattern (see also Probst et al., 2002). In order to manipulate the speech of Italian learners of German, who often show problems with the correct stress placement in morphologically complex words, prosodic features (speech rate, intonation, an intensity) of a German native speaker were transferred to the same utterance produced by an Italian speaker. They tested two groups: a manipulation group that was trained on the resynthesized utterances and a native speaker group that was trained on utterances recorded by a German native speaker. The learners were trained on utterances with and without emphasis on the stressed syllable of the wrongly pronounced word of the pre-test. According to the authors, the auditive feedback had a motivating effect for participants of the manipulation group whereas subjects of the native speaker group showed no interest in the training. Unfortunately, this statement is only based on the learners' opinions about the training. No acoustic analysis was performed comparing pre- and post-test recordings or comparing the two groups with each other. However, the authors performed a perception experiment judging the correctness of the produced lexical stress with German native listeners for the pre- and post-test recordings of the learners. The test showed that both groups received the highest scores for post-test performances after training with emphasis, and that both groups performed equally well. Regarding post-test performances after training without emphasis, participants of the manipulation group received higher scores than participants of the native speaker group.

Nagano and Ozawa (1990) conducted an experiment that compared the impact of the learner's manipulated speech and speech of a native speaker on learning the correct lexical stress placement in English. They tested two groups of Japanese speakers on the basis of individual English words. Before starting the training, they recorded forty samples. Each training session consisted of the same ten English words, had a duration of about 30 minutes, and was carried out twice a day for two days. The learners first listened either to the manipulated version or the recordings of a native speaker. The learner then recorded the word in question and received feedback. Unfortunately, no information about the type of feedback is given. Subsequently, they listened to and compared two speech utterances. After each training session, all participants were recorded (post-test condition). To investigate the impact of the two training methods, recordings of the pre- and post-test were evaluated. Two male native American English speakers evaluated the pronunciation of the individual words on a 7-point scale. Results suggest that learners who trained on their own manipulated speech received higher scores (better lexical stress placement) after training than learners who trained on recordings of a native speaker.

De Meo et al. (2016) also successfully applied a prosody transplantation approach for Chinese learners of Italian. They chose two Italian sentences that can have different meanings depending on the used pitch contour (request vs. order, and granting vs. threat). In a first perception experiment, native Italian listeners assessed a prerecorded short text for the degree of foreign accent and only speakers who were judged as strongly foreign-accented were included in the experiment. Chinese learners of Italian recorded a number of sentences and for participants of the manipulation group, recordings were subsequently manipulated for pitch contour, duration, and intensity based on Italian native speakers. Each participant was asked to listen to the recorded utterances of a native speaker or to the own manipulated speech and to imitate the presented utterances. After a training of only five minutes, the subjects recorded the sentences again for later comparisons. The recordings of the pre- and post-test were assessed by Italian native listeners on the basis of perceived foreign accentedness, intelligibility, correct speech act identification, and communication effectiveness. In general, both training conditions improved the prosodic performance of the learners. Self-imitation seemed to be especially beneficial for the correct identification of order and request.

Furthermore, communication effectiveness increased slightly for both groups, but more so for self-imitation performances. In regard to intelligibility, no difference was found for the two training conditions. However, for perceived accentedness, self-imitation helped to receive more native accent assessments.

Manipulation of the learner's voice also found its way into CAPT systems. The software Win Pitch LTL II (Martin, 2004) gives the opportunity to modify four prosodic parameters: fundamental frequency, intensity, syllable duration, and pauses, either for the model speaker's or the learner's utterance, using graphic commands. Unfortunately, there is no known research that tests the actual effect on the improvement of pronunciation.

Hirose et al. (2003) developed a tool to teach lexical accents to learners of Japanese which generates visual and audio feedback. The learner's speech is manipulated in terms of its prosodic features based on the features of a Japanese native speaker. They conducted an experiment with 8 learners of Japanese to test how helpful the audio and visual (lines/arrows indicating pitch contour) feedback is. Unfortunately, again, no acoustic analysis was carried out. They compared the number of trials for each word/sentence of a baseline tool without the feedback types vs. the new system that included the feedback methods. They only found a difference in number of trials for sentences which was 3.8 trials for the new system, and 4.8 for the baseline system. An additional questionnaire indicated that learners think that both audio and visual feedback was useful.

This short overview of studies investigating the effect of exposure to native speakers and the manipulated learner's speech showed that both types of auditory presentations help the student to improve the pronunciation. Many studies found that training on manipulated own speech of the learner is more beneficial than training on recordings of a native speaker. However, most studies did not include acoustic analyses but are based on evaluations by native speakers. Also, many of the discussed studies focused on suprasegmental problems only. The following study investigated whether training with recordiings of a native speaker or the learner's own manipulated voice is helpful to improve the production of French voiced and voiceless stops by German learners of French.

6.1 Experiment

The experiment described here tested three groups: a Control (CG), a Manipulation (MG), and a Native Speaker Group (NG). Each group was tested for a set of French and German sentences containing minimal pairs contrasting in word-initial stops. The CG subjects did not receive any feedback whereas the MG subjects were presented with their own manipulated speech throughout the experiment. In contrast to most of the presented studies, this experiment concentrates on the manipulation of segmental features. The NG subjects listened to recordings of a French native speaker (female, 28, Strasbourg) who also served as the model speaker for later comparisons.

6.1.1 Subjects

Each group consisted of five female and five male native German speakers (19-38 years, M: 23.7 years, SD: 3.9 years). The participants were all students from Saarland University with basic knowledge of French (A1-A2 level according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)) based on self-assessment. It cannot be ruled out that some participants had a higher level than declared since the language level was not verified. Each participant had to fill out a questionnaire and state since when they spoke French and it was made sure that they had not spent a longer time (> 1 month) in a French speaking country. Learning French at school until higher education entrance qualification (Abitur) was still considered as a beginner's level. Subjects were randomly assigned to one of the three experimental groups.

6.1.2 Material

For each sound contrast (/b-p/, /d-t/, /g-k/), seven French minimal pairs differing in syllable-initial position were embedded in short sentences (overall 42 sentences, from here on called experimental targets). Embedding the target words in sentences is preferred to presenting them in isolation because suprasegmental entities, like intonation or speaking rate, can be controlled more thoroughly (Künzel, 1977). All experimental targets were nouns preceded by the vowel [e] (e.g., <les>) to ensure a consistent left sound context. Additionally, stops were directly followed by any vowel (see examples 1 and 2 for the contrast /g-k/ in French). Words with a stop-following consonant were not included because coarticulation processes might affect the production in a way that, e.g., aspiration is not fully produced but rather released into the following consonant (e.g., <place>, [plas]).

(1) Les <u>c</u>ages à oiseaux sont très petites.
(The cages for birds are really small.)

(2) Les <u>g</u>ages sont payés à la fin du mois.
(The wages will be payed at the end of the month.)

Table 6.1: Number of German and French sentences included in the study.

	experimental	training	∑
French	42	24	66
German	42	24	66

To allow for a cross-language comparison, 42 additional experimental items were selected for German and embedded in appropriate sentences. Each word was preceded by the vowel [ə] (e.g., <weiße> (white), see examples 3 and 4 for /g-k/ in German). Due to a limited number of available pairs in German, two quasi-minimal pairs were included (<Teller> (plate) [tɛlɐ] vs. <Delle> (dent) [dɛlə], <Tochter> (daughter) [tɔxtɐ] vs. <Docht> (wick) [dɔxt]).

(3) Der weiße Guss auf dem Kuchen ist lecker.
(The white icing on the cake is delicious.)

(4) Der erste Kuss ist etwas Besonderes.
(The first kiss is something special.)

In addition to the experimental targets, four training targets for each stop in both languages were included (overall 6×4 words). The training targets were nouns with an initial stop followed by any vowel but they were not minimal pairs. The complete set of French and German sentences (Table 6.1) was recorded by all three groups. A complete list of the used sentences can be found in Table B.1 and Table B.2 in appendix B.

6.1.3 Manipulation

Manipulation was carried out manually and affected only Voice Onset Time (VOT). Since German speakers differentiate between voiceless unaspirated stops with a short VOT and voiceless aspirated stops with a long VOT, subjects most likely show too long VOT for French plosives. Therefore, VOT had to be shortened based on values of a French reference speaker (female, age 28, Strasbourg). Manipulation was carried out for the French training targets and was only applied if the aspiration was longer than the golden speaker's.

If VOT had to be shortened, the end of the aspiration was deleted. In hindsight it would have been better to shorten VOT starting from the center of the aspiration to keep formant transitions more or less intact. However, the manipulated words were not perceived as strange or unnatural. It was also made sure that cutting was always done at zero-crossing to prevent the generation of jumps in the sound wave which would result in audible clicks.

It could be argued that not only length of VOT is crucial for the production and perception of voiced and voiceless stops since in contrast to German, French voiced stops are characterized by a fully voiced closure. It was decided against a manipulation of phonation of the closure because it is not straightforward to modify the signal in order to become voiced. A modification of the phonation is technically challenging and often yields unsatisfactory perceptual results. Preserving the natural stimulus quality was deemed to be of higher importance.

Künzel (1977) also calls attention to the fact that not only the duration of voice onset time plays a part in contributing to the perception of voiced and voiceless stops but also the overall stop duration and duration of the preceding vowel. However, to keep the manipulation method as simple as possible, it was only focused on VOT only.

6.1.4 Procedure

Recordings took place on two subsequent days. They were carried out in quiet office rooms using a head-mounted microphone (16 kHz, 16 bit) in an M-AUDIO Fast Track USB device. Recordings were saved on a Windows laptop using a custom-made software that was developed at LORIA ('Corpus Recorder', Colotte 2013). The sentences were presented to each participant in a randomized order. The speakers were able to read and record the sentences as often as they liked but did not get any information on how to pronounce them.

On the first day, German and French sentences were recorded as a baseline for later comparisons. The second recording session on the subsequent day differed for the three groups (see Figure 6.2). Subjects of the Control Group were asked to read the same set of French sentences from the previous day once again without receiving any additional information or feedback on their pronunciation. Participants of the Manipulation and Native Group performed recordings in two parts: a training and a transfer block. Since feedback is only useful if sufficient information is provided (Butler and Winne, 1995) the focus on stops was pointed out to the subjects before they started the recording session. They were informed about the fact that there is a difference between the pronunciation of French and German but no phonetic details were given. The sound in question was underlined so participants were able to concentrate on the crucial word and sound. Participants only

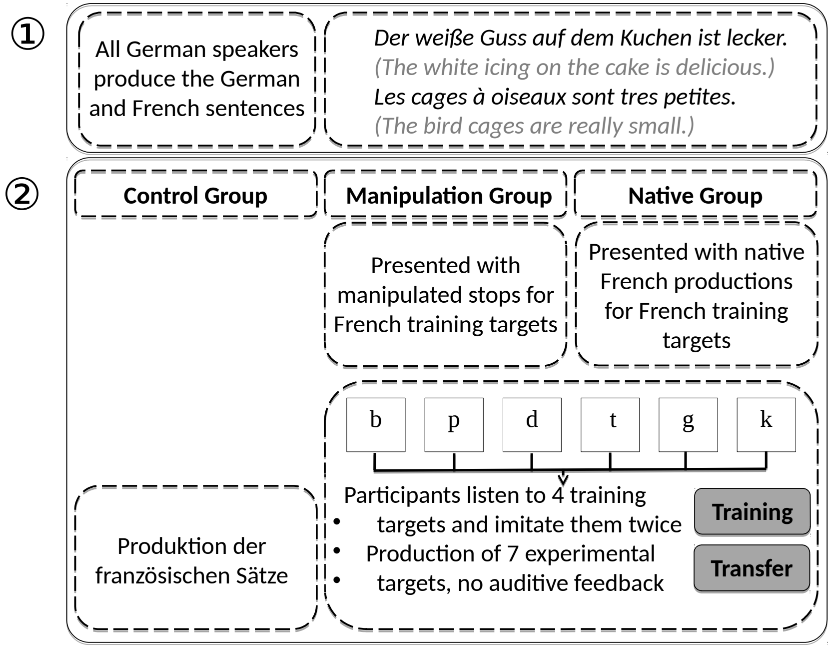


Figure 6.2: Overview of the experiment procedure of Experiment II. The structure for the second recording session differs for the three experimental groups.

worked on one stop at a time to allow them to concentrate on it and to develop a strategy.

Subjects were able to listen to either the manipulated version of their own recordings (MG) from the pre-test or recordings by a native French speaker (NG) in the training phase. Additionally, participants of the MG were told that their speech had been modified. The task was to record the sentences again and to focus on the stop in question, keeping in mind that German and French stops are produced differently. They were instructed to listen to the stop carefully and to try to imitate the manipulated or native version they heard. Participants were asked to record each sentence at least twice, in the hope that more repetitions or imitations might have a greater learning effect.

After the training block for each stop, subjects moved on to the transfer block. To ensure that they were not only able to imitate the manipulated or native productions from the training block but were also able to transfer the newly gained knowledge onto untrained words, a block of seven experimental target sentences was provided. At this point, they were not able to listen to either their manipulated voice or recordings of the native speaker. Stops in question were, again, highlighted.

6.2 Hypotheses

This experiment concentrated on acoustic analyses of stops in German and French utterances by German native speakers. The focus was placed on the duration of VOT. For the analysis, VOT was labeled for German utterances as well as the French recordings of the pre- and post-test. The following predictions were made:

(1) German speakers transfer their native phonetic knowledge to the French utterances (Flege, 1995; Kuhl and Iverson, 1995). VOT values will therefore be significantly different from stops of the French reference speaker. VOT values are expected to be longer than native French stops.

$$\text{VOT French}_{L2} > \text{VOT French}_{L1}$$

(2) There is no significant difference between the first and second recording for the Control Group. Improvements from repetition only are expected to be small due to the lack of training.

$$\text{VOT CG}_{Rec1} = \text{VOT CG}_{Rec2}$$

(3) VOT values of the second recording will be significantly shorter than for the first recording for the Manipulation and Native Speaker Group due to an improvement induced by auditory feedback and training.

$$\text{VOT MG,NG}_{Rec1} > \text{VOT MG,NG}_{Rec2}$$

(4) VOT of the second recording reduces more for the Manipulation Group than for the Native Speaker Group because speakers can focus on the manipulated features of the recording(cf. Probst et al., 2002).

$$\text{VOT MG}_{Rec2} < \text{VOT NG}_{Rec2}$$

(5) Participants of the Manipulation Group (and maybe of the Native Speaker Group) improve to the level of the French reference speaker

$$\text{VOT MG(,NG)}_{Rec2} = \text{VOT Golden Speaker}$$

Many researchers argued that if the necessary manipulation techniques can be applied, the best golden speaker for a learner is the learner himself (e.g., Felps et al., 2009; Nagano and Ozawa, 1990; Peabody and Seneff, 2006). By modifying a specific aspect of the non-native speech, all other parameters are kept the same. Following Bissiri and Pfitzinger (2009) and Nagano and Ozawa (1990) who examined both the influence of a native speaker and modified learner speech, the speaker's manipulated voice was more beneficial than a native model speaker. Learners were able to concentrate on (supra-)segmental differences only and not speaker specific differences. In the case of this investigation learners could concentrate only on the difference of VOT for stops. However, Probst et al. (2002) argue that when learners have to focus on several features when imitating a native speaker, learning might be facilitated by choosing a native speaker with similar individual features as the learner. The authors proposed three features that are easily controlled for: gender, F0, and articulation rate. They found that learners training with the Fluency system (Eskenazi and Hansma, 1998), who chose a native speaker similar to their own voice, improved more than learners who chose a mismatch.

Wang and Lu (2008) point out that even if some CAPT systems provide a set of different speakers, there is still a lack of knowledge in determining the best reference speaker for a language learner. They investigated which voice features are preferred by language learners for imitation. In contrast to Probst et al. (2002), Wang and Lu (2008) do not use several model speakers but resynthesize the recordings of one teacher in regard to different voice features: speech rate and pitch-formants. In order to do so, they integrated the resynthesis procedure in the CAPT system CASTLE (Computer-Assisted Stress Pattern Teaching and Learning Environment) (Lu et al., 2010). They investigated how preferable the resynthesized versions were rated by fifteen learners of English with different backgrounds and proficiencies. They showed that a model speaker with the same gender and similar speed to a learner's voice was not always the preferred speaker by a learner (which is similar to Probst et al. (2002) who found that participants did not always choose model speakers similar to their own voices). However, they did not examine whether a model speaker with similar voice features had a positive effect on pronunciation.

In this experiment, learners were not compared to the golden speaker on the basis of individual voice features. This was not suitable for this experiment due to the small set of learners. Also, only one golden speaker was used instead of several speakers which would have otherwise increased the complexity of the experiment. Since it was not the aim of this experiment to test the impact of different native speakers on the production of French stops but rather to see whether training with a native speaker has a beneficial effect in general, only one native speaker was tested.

6.3 Results

Duration of VOT was labeled using Praat (Boersma and Weenink, 2013). The traditional approach to analyze voiced stops is to report negative VOT values which represent the voiced closure. However, the durations of these negative values would have a strong effect on the calculation of (positive) mean VOT values which does not coincide with the research question of this experiment (the golden speaker used for this experiment produced fully voiced closures between -60 ms and -105 ms). Since this experiment focused on the length of aspiration (which is not captured by a negative VOT) only positive values were measured. Therefore, the VOT of fully voiced stops was represented by 1 ms for convenience and later treated as 0 ms in the analysis. This method coincides with the proposed measure After Closure Time (ACT) by Mikuteit and Reetz (2007) for research on VOT.

VOT values were analyzed using JMP (JMP, 2014). Overall, 119 values were excluded due to hesitations while producing the stop or wrongly pronounced plosives. Subsequently, values were entered into a linear mixed model with VOT as the dependent factor, SPEAKER and ITEM as random factors, GENDER, MATERIAL (training/transfer targets), FOLLOWING SOUND, and the combination of GROUP (Control/Manipulation/ Native), TEST (first/ second recording), LANGUAGE (French vs. German), and PHONATION (voiced/ voiceless) as independent factors. Due to the structure of this experiment, it was not possible to enter these factors individually because of empty cells and dependencies (e.g., German recordings were only made for the first recording session.).

The results of the statistical analysis indicated no main effect of GENDER ($F(1,26.13)= 0.6567$, $p=0.425$), which confirms the overall expectation that the articulation of plosives is not gender specific. MATERIAL showed no effect ($F(1,333.6)= 1.0349$, $p=.3098$), which indicates that the targets of the training (imitation) phase did not differ from the targets of the transfer phase. Therefore, if participants were able to imitate the manipulated or native stops, respectively, they were also able to transfer this production strategy to the experimental targets. FOLLOWING SOUND showed a significant influ-

Table 6.2: Statistical information of the Linear Mixed Model.

source	between-groups df	within-groups df	F	p
FOLLOWING SOUND	13	93.03	3.1662	<0.0006*
GENDER	1	26.13	0.6567	0.425
MATERIAL	1	333.6	1.0349	0.3098
GROUP_TEST_LANGUAGE_PHONATION	19	541.6	48.9333	<0.0001*

ence on VOT (F(13,93.03)=3.1662, p<0.001) which was expected. Furthermore, the factor combination showed a main effect (F(19,541.6)=48.93333, p<0.0001). Post-hoc tests were carried out to take a closer look at the specific contrasts with regard to the combination.

6.3.1 Comparison of German and French Recordings

Due to the structure of the experiment, adding all factors into the model would generate empty cells and dependencies. To take a closer look at the influence of the factor LANGUAGE, the combination of GROUP, TEST, LANGUAGE, and PHONATION was further analyzed by running post-hoc tests for specific contrasts.

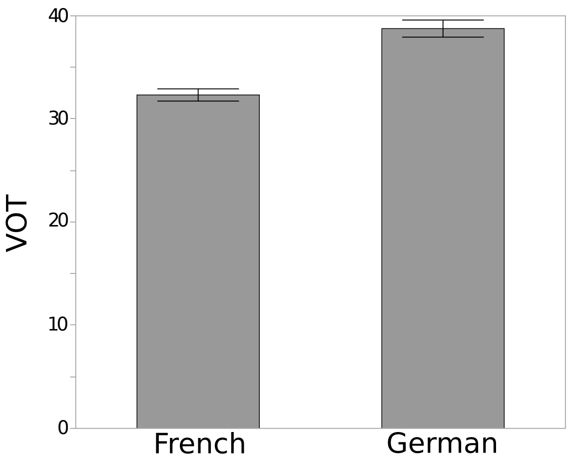


Figure 6.3: Mean VOT values (ms) for German and French stops of the first recording across groups by German learners of French.

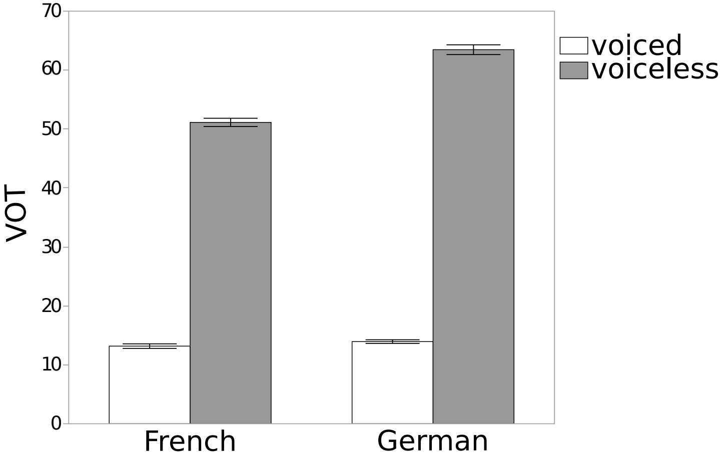


Figure 6.4: Mean VOT values (ms) for voiced and voiceless German and French stops of the first recording across groups by German learners of French.

A comparison of the French and German recordings of the pre-test shows a difference between the two languages in terms of VOT ($t(11)=-4.032$, $p<0.001$). In general, German has a mean VOT of about 39 ms and French of about 32 ms (see Figure 6.3). Going into more detail and taking a closer look at voiced and voiceless stops in both languages (see Figure 6.4) it is noticeable that the mean durations for voiced stops are similar for German and French (14 ms and 13 ms, respectively). No significant difference was found. However, there is a larger difference for voi3eless stops (64 ms for German and 51 ms for French) which was significantly different ($t(5)=-5.375$, $p<0.01$). It is interesting to see that German learners of French seem to have an understanding that they need to reduce VOT for voiceless stops when speaking French but not for voiced stops. However, VOT values for the French recordings are still considerably high.

To exploit whether the three groups behave similarly, another contrast analysis was executed. In Figure 6.5 it can be seen that all three groups behave similarly with regard to the production of VOT of voiced and voiceless stops. Again, durations of VOT of voiced stops are similar for French and German for all groups and do not show a significant difference. However, there is a significant difference for the duration of voiceless stops for both languages, with higher values for German than for French (Control ($t(1)=-6.286$, $p<0.001$); Manipulation ($t(1)=-4.595$, $p<0.01$); Native ($t(1)=-4.365$, $p<0.05$)). No difference between groups was found.

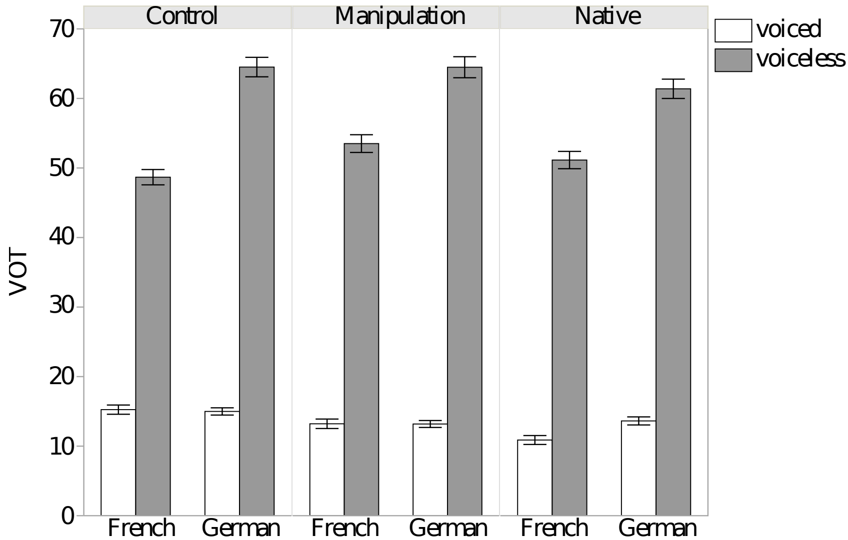


Figure 6.5: Mean VOT values (ms) for voiced and voiceless German and French stops of the first recording for the Control, Manipulation, and Native Group by German learners of French.

6.3.2 Comparison of the First and Second French Recordings

No significant differences were found for the first recordings between the three experimental groups for either voiced or voiceless stops, which allows for a comparison across groups. Comparing the performance of the speakers for the first and second French recording, a significant reduction for VOT was found only for voiceless stops for all three groups (Control ($t(1)=2.2858$, $p<0.05$); Manipulation ($t(1)=7.1187$, $p<0.0001$); Native ($t(1)=13.168$, $p<0.01$), see Figure 6.6). Mean VOT values shown in Table 6.3 indicate that the reduction for CG is only 3 ms, whereas the difference for MG and NG is 7 and 13 ms, respectively. Since the analysis is carried out for VOT, changes will always be in a small millisecond range. However, it is doubtful that a reduction of 3 ms for the CG is sufficient for a noticeable perceptual effect.

Although MG showed a significant improvement for voiceless stops in the second recording, it is not significantly different to the CG. Both the CG and MG are different to the NG who improved the most ($t(1)=2.472$, $p<0.05$) and $t(1)$-2.549, $p<0.05$, respectively). A comparison of voiced stops of the second recording between groups showed no differences.

Table 6.3: Mean VOT values (ms) of the experimental groups for the first and second French recording by German learners of French and reference values of the French reference speaker.

		Rec 1	**Rec 2**
Control	*voiced*	15	15
	voiceless	49	46
Manipulation	*voiced*	13	11
	voiceless	54	47
Native Speaker	*voiced*	11	12
	voiceless	51	38
Golden Speaker	*voiced*	3	
	voiceless	30	

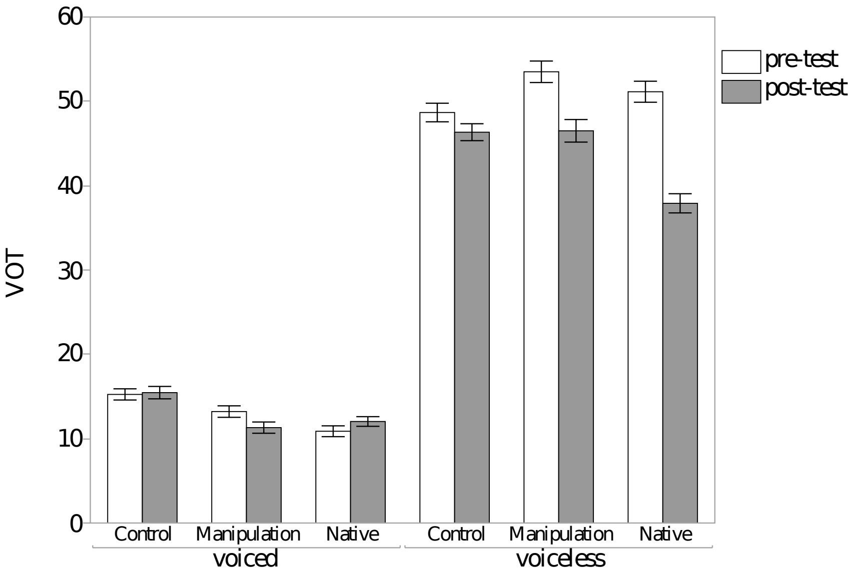


Figure 6.6: VOT values (ms) of voiced and voiceless stops for the first and second French recordings by German learners of French.

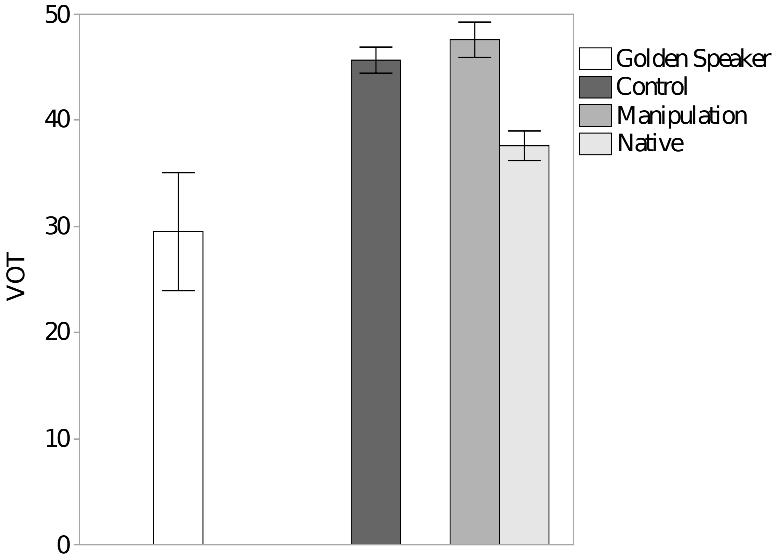


Figure 6.7: VOT values (ms) for voiceless stops of the second French recording by German learners of French in comparison to the golden speaker.

6.3.3 Comparison to the Golden Speaker

A comparison of voiceless stops of the second recording showed that only speakers from the NG managed to reduce their VOT to a level that is not significantly different from the golden speaker (t(1)=1.0963, p=0.2799, see Figure 6.7). Therefore, the exposure to a native speaker did have a beneficial impact in reducing the VOT for voiceless stops.

As for voiced stops, all three groups are not significantly different to the golden speaker (see Figure 6.8), although there is a clear difference in duration.

6.3.4 Comparison of Places of Articulation (PoA)

To analyze the influence of place of articulation, an additional model was performed. Because this model was applied to a subset of the data (French productions of experimental targets for the three groups) a full factorial linear mixed model with VOT as dependent factor was performed. SPEAKER and ITEM were treated as random factors and FOLLOWING SOUND, GENDER, TEST, GROUP, PHONATION, and PLACE OF ARTICULATION as indepen-

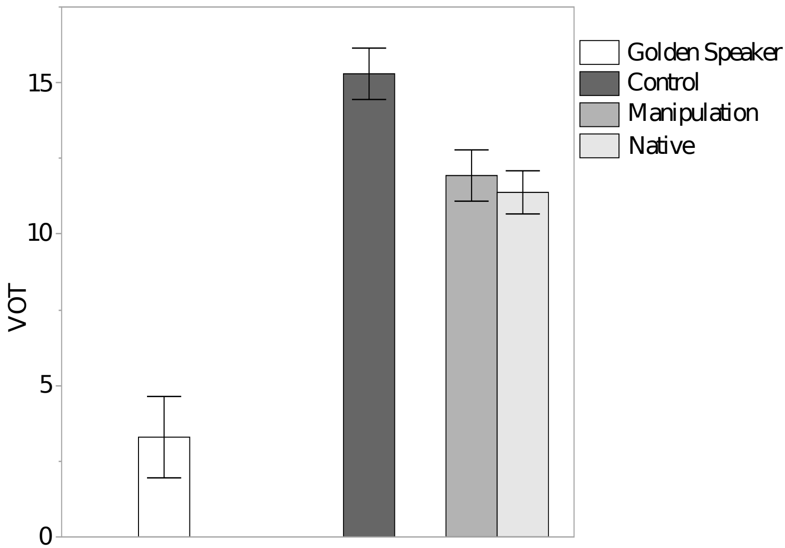


Figure 6.8: VOT values (ms) for voiced stops of the second French recording by German learners of French in comparison to the golden speaker.

dent factors. Additonally, all interactions of TEST, GROUP, PHONATION, and PLACE OF ARTICULATION were added as well (see Table 6.4).

The results of the statistical analysis indicate an effect of TEST ($F(1, 2342)= 55.5896$, $p<0.0001$) with a reduction of VOT for the second recording (32 ms vs. 29 ms). Also, TEST×GROUP shows a significant difference ($F(2, 2342)=45.7867$, $p<0.0001$). Post-hoc tests indicate a significant improvement from the first to second recording for all three groups over all stops. Also, PHONATION shows an effect ($F(1, 27.95)=85.5511$, $p<0.0001$) with longer VOT values for voiceless than for voiced stops (47 ms and 13 ms, respectively) which is to be expected. Both TEST×PHONATION and GROUP×PHONATION interactions are significant ($F(1, 2342)=4.1721$, $p<0.05$ and $F(2, 2342)=7.4294$, $p<0.0001$, respectively). Since these aspects have been captured in the previous sub-chapters, they will not be discussed here in detail but are mentioned as part of the complete statistical model. With regard to PLACE OF ARTICULATION only the main factor, but no interaction, reached significance ($F(2, 27.91)=51.9274$, $p<0.0001$). Post-hoc tests show that all three places of articulation differ significantly from each other (bilabial: 24 ms, alveolar: 33 ms, velar: 35 ms).

Table 6.4: Statistical information of the Linear Mixed Model with focus on place of articulation (PoA).

source	**between-groups df**	**within-groups df**	**F**	**p**
GENDER	1	25.76	0.011	0.9174
FOLLOWING SOUND	9	33.46	2.0161	0.0685
TEST	1	2342	55.5896	<0.0001*
GROUP	2	29.68	0.8829	0.4242
TEST×GROUP	2	2342	45.7867	<0.0001*
PHONATION	1	27.95	85.5511	<0.0001*
TEST×PHONATION	1	2342	4.1821	0.041*
GROUP×PHONATION	2	2342	7.4294	0.0006*
TEST×GROUP×PHONATION	2	2342	0.9551	0.3849
POA	2	27.91	51.9274	<0.0001*
TEST×POA	2	2342	2.1807	0.1132
GROUP×POA	4	2342	2.2952	0.0571
TEST×GROUP×POA	4	2342	0.6092	0.656
PHONATION×POA	2	27.94	1.1769	0.323
TEST×PHONATION×POA	2	2342	1.7084	0.1814
GROUP×PHONATION×POA	4	2342	0.4444	0.7766

6.3.5 Number of Produced Fully Voiced Stops

Although this experiment does not focus on the production of fully voiced closure durations of voiced stops it might still be interesting to see how many fully voiced stops were produced by the participants in the pre- and post-test condition. It was hypothesized that subjects of the CG and MG would not improve significantly in the number of fully voiced stops. Since subjects of the MG were trained on their own voice which was manipulated only for the length of VOT, a significant improvement is not to be expected. However, participants of the NG might have had the chance to extract useful information from the recordings of the native speaker regarding the production of fully voiced stops. But because voicing is not a feature of German phonology with regard to the production of stops, participants might not have paid attention to it. or had problems producing it correctly. Since a fully voiced stop was treated with a VOT of 0 ms, the effect of an increased number of fully voiced stops was not observable in the previously discussed data.

To verify whether the number of produced fully voiced stops differed between groups for the pre- and post-test and whether learners improved from pre- to post-test, a Generalized Linear Mixed Model was performed in JMP. The data set was extended by the variable FULLY VOICED which included information whether a fully voiced stop was produced or not (yes vs. no). Additionally, the model was carried out for a subset of the data (French recordings of the pre- and post-test of voiced stops for the three experimental groups). FULLY VOICED was entered into the model as dependent factor. SPEAKER and ITEM were treated as random effects and FOLLOWING SOUND, GENDER, TEST, GROUP, MATERIAL, and PLACE OF ARTICULATION as fixed effects. Additionally, all interactions of TEST, GROUP, MATERIAL, and PLACE OF ARTICULATION were added into the model. The model was performed with a binomial distribution and Logit link function.

Table 6.5: Statistical results of the Generalized Linear Mixed Model for number of fully voiced stops.

source	**df**	χ^2	**p**
FOLLOWING SOUND	9	21.316508	0.0113*
GENDER	1	157.07862	<0.0001*
MATERIAL	1	6.4903099	0.0108*
TEST	1	0.2326261	0.6296
MATERIAL×TEST	1	7.7614146	0.0053*
GROUP	2	8.3917708	0.0151*
MATERIAL×GROUP	2	2.8901323	0.2357
TEST×GROUP	2	26.925399	<0.0001*
MATERIAL×TEST×GROUP	2	0.9376778	0.6257
POA	2	62.152643	<0.0001*
MATERIAL×POA	2	9.485227	0.0087*
TEST×POA	2	1.4691923	0.4797
MATERIAL×TEST×POA	2	1.191512	0.5511
GROUP×POA	4	2.1276705	0.7123
MATERIAL×GROUP×POA	4	3.5375713	0.4722
TEST×GROUP×POA	4	4.1984916	0.3798
MATERIAL×TEST×GROUP×POA	4	1.5877908	0.811

Table 6.5 shows the statistical results of the Generalized Linear Mixed Model. A main effect was found for MATERIAL ($\chi^2(1)=6.49$, $p<0.05$) indicating a difference between the training and experimental items. Table 6.6 shows that percentagewise more fully voiced stops were produced for the training targets than for the experimental target.

The interaction MATERIAL×TEST was found to be significant ($\chi^2=7.76$, $p<0.01$). Post-hoc tests show that there is a difference between the first

Table 6.6: Number of produced fully voiced stops (and percentages) for the training and experimental items of the French recordings.

	fully voiced	not fully voiced
Training	235 (34)	460 (66)
Experimental	315 (26)	887 (74)

Table 6.7: Number of produced fully voiced stops (and percentage) for the experimental and training targets for the first and second French recording.

		yes	no
Training	Rec1	123 (36)	222 (64)
	Rec2	112 (32)	238 (68)
Experimental	Rec1	140 (23)	461 (77)
	Rec2	175 (29)	426 (71)

and second recording for the experiment items (χ^2=7.2, $p<0.01$) but not for the training items (χ^2=2.1, $p=0.147$). Speakers produced six percent more fully voiced stops for the experimental items in the post-test condition (see Table 6.7). Furthermore, experiment and training items show a significantly different number of produced fully voiced stops in the first recording (χ^2=14.26, $p<=0.001$) but not for the second recording (χ^2=0.002, $p=0.97$). In summary this means that training did not help to increase the number of fully voiced stops for the training items. However, speakers were able to compensate for the difference between the two sets in the second recording.

A main effect was found for GROUP ($\chi^2(2)$=8.39, $p<0.05$) indicating a difference for groups with respect to the produced number of fully voiced stops. Post-hoc tests show that significantly less fully voiced stops were produced by the Control Group (n=157) compared to the Manipulation Group (n=197) ($\chi^2(1)$=6,02 $p<0.05$) and the Native Group (n=196) ($\chi^2(1)$=6.35, $p<0.05$). No significant difference was found for the Manipulation and Native Group ($\chi^2(1)$=0.04, $p=0.85$).

Furthermore, the interaction TEST×GROUP ($\chi^2(2)$=26.93, $p<0.0001$) shows an effect (see also Figure 6.9). Post-hoc tests show a significant difference for the first and second recording for the Manipulation Group ($\chi^2(1)$= 12.08, $p<0.001$) and Native Group ($\chi^2(1)$=17.72, $p<0.001$) whereas no significant change was observed for subjects of the Control Group ($\chi^2(1)$=1.57, $p=0.21$). Taking a look at the raw data, participants of the Manipulation

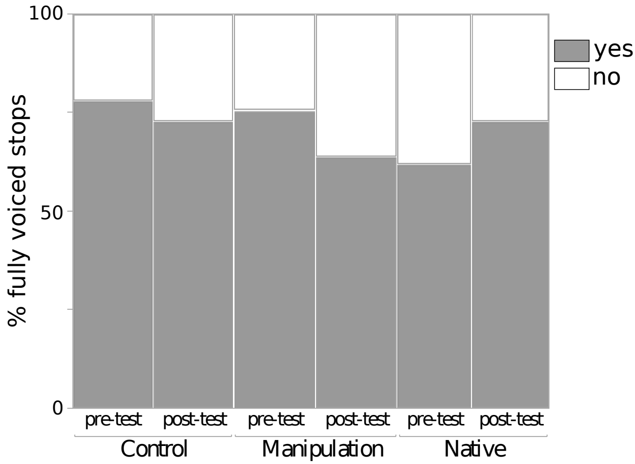


Figure 6.9: Percentages of produced fully voiced stops.

Group increased their number of fully voiced stops from 80 to 117. However, the number of produced fully voiced stops by participants of the Native Group decreased from 113 to 83. It is surprising to see that the number of fully voiced stops increased for participants of the Manipulation Group who did not receive any feedback on voicing in the sense of vocal fold vibration during closure. In contrast, participants who trained with a native speaker who produced fully voiced stops, got worse. Participants might have tried to produce more fully voiced stops after the exposure to recordings of the native speaker. Unfortunately, without any additional information what is expected from them, pronunciation might have got worse or at least less fully voiced stops were produced.

Taking a closer look at PLACE OF ARTICULATION ($\chi^2(2)$=62.15, p<0.0001), post-hoc tests show significant differences for all contrasts (bilabial vs. alveolar: $\chi^2(1)$= 59.24, p<0.0001; bilabial vs. velar: $\chi^2(1)$=19.03, p<0.0001; alveolar vs. velar: $\chi^2(1)$= 11.53, p<0.001, see Table 6.8).

The interaction MATERIAL×POA (χ^2=9.49, p<0.01) was also found significant. Post-hoc tests show that experimental and training targets are only significantly different for bilabial stops (χ^2=17.04, p<0.0001). Comparing the places of articulation to each other both with respect to experimental and training items, the same picture emerges as seen for PLACE OF ARTICULATION. Please refer to Table 6.9 for more information.

Further on, GENDER shows a main effect ($\chi^2(1)$=157.08, p<0.0001). Female speakers produced significantly less fully voiced stops than male speak-

Table 6.8: Number of produced fully voiced stops for the French recordings for the different places of articulation across groups.

	fully voiced	not fully voiced
Bilabial	144	256
Alveolar	69	337
Velar	102	294

Table 6.9: Number of produced fully voiced stops (and percentages) for the French recordings for the experimental and training words for the different places of articulation.

		fully voiced	not fully voiced
Bilabial	Experimental	144 (36)	256 (64)
	Training	124 (53)	111 (47)
Alveolar	Experimental	69 (17)	337 (83)
	Training	43 (18)	192 (82)
Velar	Experimental	102 (26)	296 (74)
	Training	68 (30)	157 (70)

ers (158 vs. 393). This result cannot be explained at this time and might be connected with individual language skills. Also, FOLLOWING SOUND was found to show a significance ($\chi^2(9)$=21.32, p<0.05). No other factors or interactions were statistically significant.

6.3.6 Individual Speaker Differences

To this point the effect SPEAKER was always treated as random effect. Although all speakers were French beginners, they do show large variability in the production of VOT for voiced and voiceless stops as well as the extent of their improvements from the first to the second French recording. A statistical analysis was carried out for each experimental group for the French recordings only. VOT was entered into the Linear Mixed Model as dependent factor, WORD as a random effect, and FOLLOWING SOUND, MATERIAL, TEST, PHONATION and SPEAKER as independent effects as well as all possible interactions of the last three variables. This time, SPEAKER was not treated as a random effect in order to investigate the impact of the individuality on VOT. GENDER was not included in this model due to dependency problems

Table 6.10: Statistical information with focus on individual differences for speakers of the Control Group.

source	between-groups df	within-groups df	F	p
FOLLOWING SOUND	10	52.66	2.8774	0.0061*
MATERIAL	1	102	0.004	0.95
TEST	1	1187	2.3621	0.1246
SPEAKER	9	1187	64.7765	< 0.0001*
TEST×SPEAKER	9	1187	2.686	0.0043*
PHONATION	1	52.3	271.0933	< 0.0001*
TEST×PHONATION	1	1187	4.1613	0.0416*
SPEAKER×PHONATION	9	1187	25.6295	< 0.0001*
TEST×SPEAKER×PHONATION	9	1187	1.2333	0.2702

(half of the speakers were female, the other half were male). Since GENDER was not significantly different in the main analysis it was not included here.

Table 6.10 shows the statistical results of the model for the Control Group. Because a similar model was conducted previously, it will be mostly focused on the factor SPEAKER and its interactions. SPEAKER shows a main effect ($F(9,1187)=64.68$, $p<0.0001$) which is not surprising due to the large inter-speaker variability.

Furthermore, post-hoc tests for the interaction TEST×SPEAKER showed that only three participants improved their VOT productions (speaker 3 ($t(1)=2.65$, $p<0.01$); speaker 8 ($t(1)=2.09$, $p< 0.05$), and speaker 9 ($t(1)= 2.48$, $p<0.05$)) which is not surprising since no participant received training. Why these participants improved cannot be explained at this point. Additional training beyond the experiment is possible but unlikely because the second recording was made only one day after the first recording. Also, the interaction SPEAKER×PHONATION shows a main effect ($F(8,1187)=25.63$, $p<0.0001$) which is not surprising because this significance is based on the difference between voiced and voiceless stops. Further on, the interaction TEST×SPEAKER×PHONATION does not show a significant effect ($F(9, 1187)= 1.23$, $p=0.2702$). However, for illustration purposes, a graph of the three-way interaction can be found in Figure 6.10 which clearly shows the speaker-specific differences and behaviors. For example, speaker 1 got worse for both voiced and voiceless stops when comparing the first and second recording while others improved slightly (e.g., speaker 8).

Taking a look at the results for the statistical model based on the data of the Manipulation Group (see Table 6.11), SPEAKER shows a main effect (F(9,

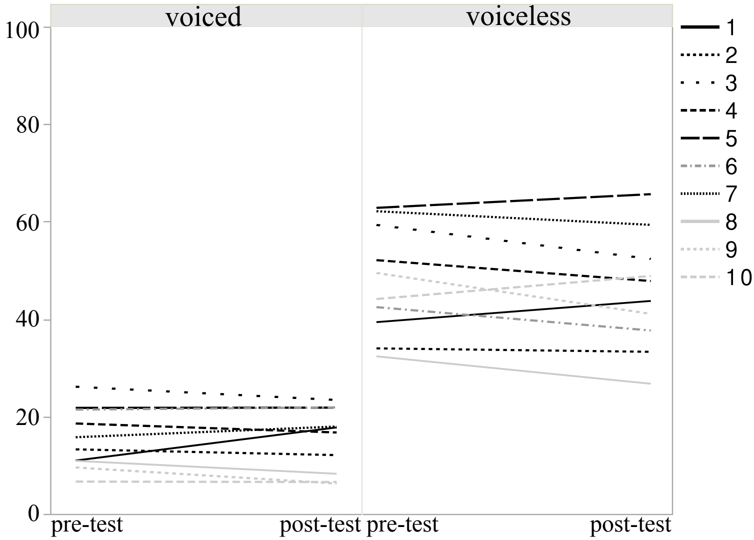


Figure 6.10: Mean VOT values (ms) for the French recordings of the individual speakers of the Control Group for voiced and voiceless stops.

Table 6.11: Statistical information with focus on individual differences for speakers of the Manipulation Group.

source	**between-groups df**	**within-groups df**	**F**	**p**
FOLLOWING SOUND	10	52.49	2.6855	0.0098*
MATERIAL	1	99.22	1.821	0.1803
TEST	1	1204	46.3525	<0.0001*
SPEAKER	9	1204	116.624	<0.0001*
TEST×SPEAKER	9	1204	5.5923	<0.0001*
PHONATION	1	52.17	341.4007	<0.0001*
TEST×PHONATION	1	1204	15.494	<0.0001*
SPEAKER×PHONATION	9	1204	34.4622	<0.0001*
TEST×SPEAKER×PHONATION	9	1204	3.5236	0.0003

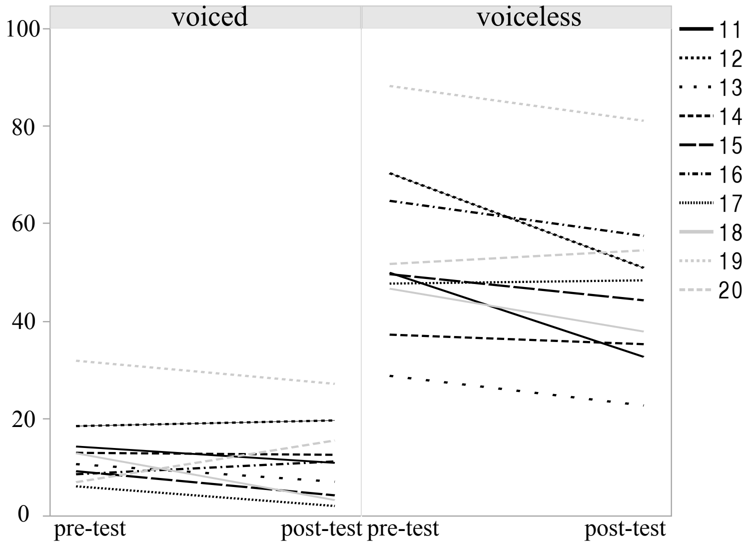


Figure 6.11: Mean VOT values (ms) for the French recordings of the individual speakers of the Manipulation Group for voiced and voiceless stops.

1204)=46.35, p>0.0001) as well as the interaction TEST×SPEAKER (F(9, 1204)=5.59, p<0.0001). Post-hoc tests for the interaction show that only three speakers did not improve significantly (speaker 14, 16, and 17). Again, SPEAKER×PHONATION shows a main effect (F(9,1204)=34.4622, p<0.0001) which is based on the articulatory difference between voiced and voiceless stops. In contrast to the statistical analysis of the Control Group, the interaction TEST × SPEAKER × PHONATION is significantly different (F(9,1204)=3.52, p<0.001). Taking a closer look at Figure 6.11 it is noticeable that speakers from the Manipulation Group show a larger variation than speakers from the Control Group. And that not all speakers reduced their VOT values in the second recording. Results for performed post-hoc tests show that six of ten speakers improved their production of voiceless stops significantly (Table 6.12). Surprisingly, two speakers show significant differences in their production of voiced stops. Taking a closer look it can be seen that while speaker 18 improved, speaker 20 got significantly worse in the second recording.

For speakers of the Native Group, SPEAKER is also significantly different (F(9, 1138)= 45.86, p<0.0001). Figure 6.12 illustrates how diverse the participants behave in respect to the production of voiced and voiceless stops. One speaker in particular stands out (speaker 29) who produced relatively short VOT values from the beginning. The interaction TEST×SPEAKER is

Table 6.12: Post-hoc tests for the interaction TEST×SPEAKER×PHONATION for speakers of the Manipulation Group. (n.s. = not signficant, * = significant, α=0.05)

speaker	**voiced**	**voiceless**
11	n.s.	*
12	n.s.	*
13	n.s.	*
14	n.s.	n.s.
15	n.s.	n.s.
16	n.s.	*
17	n.s.	n.s.
18	*	*
19	n.s.	*
20	*	n.s.

also significant (F(9,1137)=7.42, p<0.0001). Post-hoc tests show that at least 6 speakers improved. There is one speaker (speaker 26) who received a p-value of 0.05 which is not significantly different in the strict sense. Speakers 22, 23, and 29 did not change significantly. Furthermore, the interaction TEST×PHONATION shows a main effect (F(9,1138=17.91), p<0.0001). Also, the three-way interaction TEST×SPEAKER× PHONATION is significant (F(9,138)=2.97, p<0.01). Results of the post-hoc tests are shown in Table 6.14. It can be seen that no speaker improved significantly for voiced stops. For voiceless productions, only two speakers did not improve significantly (speaker 23 and 29). The results of speaker 29 might be explained by the fact that this speaker was already quite good from the beginning (22 ms in the pre-test condition and 18 ms in the post-test condition). Note that the golden speaker produced voiceless stops on average with a VOT of 30 ms (cf. Table 6.3).

6.4 Discussion

This study examined the effect of two training procedures on the pronunciation of voiced and voiceless stops by German learners of French. It was shown that the manipulation of the speaker's voice, here the reduction of VOT, had a motivating effect on the production of these sounds (Manipulation Group (MG)). However, the improvement was only effective for voiceless but not for voiced stops. The same result was found for exposure to recordings of

Table 6.13: Statistical information with focus on individual differences for speakers of the Native Group.

source	**between-groups df**	**within-groups df**	**F**	**p**
FOLLOWING SOUND	10	53.15	2.5722	0.0128*
MATERIAL	1	105.9	0.0036	0.9523
TEST	1	1141	82.4681	<0.0001*
SPEAKER	9	1138	45.8601	<0.0001*
TEST×SPEAKER	9	1137	7.4174	<0.0001*
PHONATION	1	52.67	222.15	<0.0001*
TEST×PHONATION	1	1141	108.9303	<0.0001*
SPEAKER×PHONATION	9	1138	17.9142	<0.0001*
TEST×SPEAKER×PHONATION	9	1138	2.9705	0.0017*

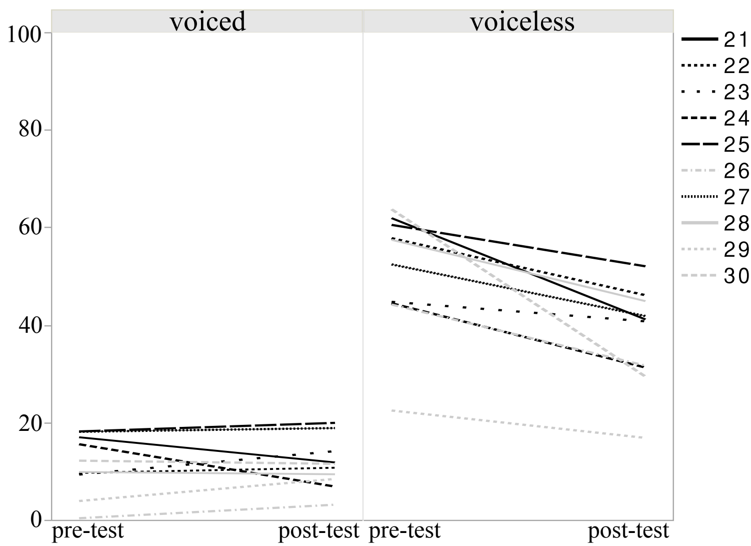


Figure 6.12: Mean VOT values (ms) for the French recordings of the individual speakers of the Native Group for voiced and voiceless stops.

Table 6.14: Post-hoc tests for the interaction TEST×SPEAKER×PHONATION for speakers of the Native Group. (n.s. = not signficant, * = significant, α=0.05)

speaker	voiced	voiceless
21	n.s.	*
22	n.s.	*
23	n.s.	n.s.
24	n.s.	*
25	n.s.	*
26	n.s.	*
27	n.s.	*
28	n.s.	*
29	n.s.	n.s.
30	n.s.	*

a native French speaker (Native Speaker Group (NG)). Again, participants were able to reduce the duration of VOT for voiceless stops. The reduction for the NG was even more distinct than for the MG. It was shown that, although MG subjects improved their pronunciation of voiceless stops, the improvement did not show a signficant difference to the productions of the Control Group (CG). Additionally, both groups differed in comparison to the NG, with higher VOT values for CG and MG. This result contrasts with findings by Bissiri and Pfitzinger (2009) and Nagano and Ozawa (1990) who found that subjects who trained on their own modified voice, achieved better pronunciation results than subjects training with recordings of a native speaker. At this point, it has to be highlighted that the studies by Bissiri and Pfitzinger (2009) and Nagano and Ozawa (1990) did not carry out any acoustic analyses. Furthermore, they investigated the effect of manipulation of suprasegmental features. As mentioned before, many studies that investigated the quality and impact of manipulation procedures, focused on modifications on a suprasegmental level. One reason might be that the transplantation or modification of pitch, duration, and intensity is achieved with less effort and resulting in an output with a higher quality.

Zhao et al. (2012) showed that segmental modification of speech reduced the quality more profoundly than modifications of suprasegmental features. Additionally, manipulation of phones is a complex procedure that needs to take various variables into account. Coarticulation processes and (language specific) phonetic dependencies have to be considered (i.e., duration differences for vowels, that precede final stops, in English (e.g., Luce and Charles-Luce, 1985)). It is true that this experiment actually did focus only on one particular parameter: Voice Onset Time. The reason was to keep manipulation processes as simple as possible. The aim of this experiment was to investigate whether the modification of a feature like VOT, which focuses on durational differences of only a few milliseconds, can be perceived by language learners. And it seems that participants of the MG managed to perceive these differences and transferred them to their post-test production. However, this improvement was not as profound as expected. It has to be kept in mind that, for both feedback methods, the training and transfer phase for each stop was quite short. Only four training and seven experimental targets were included. Even with such a small number of training targets, an improvement for voiceless stops was observed. A longer training phase and a larger training set would be necessary to see if the improvement could be reinforced.

A comparison of the NG with the golden speaker showed no significant difference. This demonstrates the benefit of being exposed to native productions, since learners were able to reduce VOT to the level of a native speaker, while speakers of the MG failed to do so. Since no other studies did compare the post-test results to the recordings of native speakers, no comparison to the literature can be drawn.

One important point is that auditory feedback only improved the pronunciation of voiceless but not voiced stops. It might be argued that manipulation of VOT is not a sufficient method to have an effect on voiced stops. This is of course true, since French voiced stops are produced with a fully voiced closure which was not modified by the manipulation procedure. One reason is that modifying the phonation of a signal is a difficult task. One approach would be to replace the voiceless closure with a fully voiced closure taken either from a recording of the learner of from a recording of a native speaker. However, many factors have to be considered (such as pitch and phonetic context (i.e. coarticulation processes)) in order to make a replacement successful on a perceptual level. One approach of manipulating voicing was recently tested by Ghosh et al. (2016). They replaced the unvoiced production of final fricatives by German learners of French with voiced fricatives by French native speakers and also changed the duration of the preceding vowel accordingly. In addition, F0 of the transferred French fricative was also modified to ensure pitch continuity. However, the method was not yet tested for efficiency.

However, exposure to native speech also failed to show an effect on improving the pronunciation of voiced stops, although it contained all relevant features. It is unclear whether learners were not able to perceive the differences or whether they are not able to apply the knowledge extracted from the native utterances. Following Flege (1995) (but see also Kuhl and Iverson, 1995; Kuhl et al., 2008; Escudero, 2005; van Leussen and Escudero, 2015), French stops are similar to German stops on a phonetic level and therefore fall into the corresponding German phonetic categories. This will result in a German accented production of the stops. Reducing the VOT of German accented voiceless stops with a long aspiration seems to be easier to realize by the learners since German voiced stops are already characterized by a short VOT. However, producing a voiced stop without any VOT (according to French voiced stops, ignoring the state of phonation of the closure) is not as straightforward. It might be advantageous for the learner to be exposed to more than one native speaker in order to be exposed to a higher variation of native speech. To address the question whether higher variability in speech is helpful for the learner to improve the pronunciation of stops, a high variability training study was conducted. This study is discussed in chapter 8.

Finally, Kartushina et al. (2015) argued that improvements in the production of a second language might not be perceived as 'sounding native' or 'more native-like' by native listeners of the respective language. To test this claim, a perception experiment was carried out which is discussed in the following chapter.

Chapter 7

Experiment III: Evaluation of Experiment II by French Native Speakers

The presented study 'Experiment II: Exposure to a Native Speaker and Modified Voice' in chapter 6 has shown that training with manipulated and native stimuli has a positive impact on the production of voiceless stops, and, in particular, that training with a native speaker is the most effective method. The next step is to investigate whether this improvement in the speech of German natives speaking French is noticeable by French native speakers. A perception experiment is presented in this chapter.

7.1 Experiment

7.1.1 Subjects

Overall, 51 participants completed the perception experiment (21-70 years, M: 32.9 years, SD: 12.0 year). They were all French native speakers and completed a short questionnaire before starting the experiment.

7.1.2 Material

In consequence of the number of recorded stimuli and the number of subjects of experiment II, it was decided to reduce the number of stimuli used

Table 7.1: Minimal pairs which were excluded from the perception experiment due following nasal vowels and same vowel contexts.

bilabial	**alveolar**	**velar**
ponts-bonds	teint-daims	camps-gants
pains-bains	temps-dents	cars-gares
pique-bique	thons-dons	cas-gars

in the perception experiment from seven minimal pairs per place of articulation to four. Mostly, minimal pairs including a nasal vowel were excluded because these sounds are difficult to produce correctly for German native speakers. Since the listeners also rated the foreign accentedness of the stimuli, additional difficult sounds (e.g., nasal vowels) could have influenced the ratings. Furthermore, the following vowel context was kept as diverse as possible. Therefore, additional minimal pairs with the same vowel context were excluded. Table 7.1 shows the minimal pairs that were excluded from the perception experiment.

For the analysis of experiment II, 119 stimuli were excluded in which the speaker showed a hesitation while producing a stop, or failed to produce the target sound. As a next step, stimuli were excluded when only the first or second recording by a speaker was available and when the minimal pair was not complete, i.e. both the voiced and voiceless variant had to be available. Overall, 1316 stimuli were used in the perception experiment. However, this number of stimuli was still too high to be assessed by each participant. For that reason, five lists were created – four lists with 264 stimuli and one list with 260 stimuli (in total 1316 stimuli). Each list included the complete minimal pair set (i.e., voiced and voiceless productions of the pre- and post-test, n=4) without any repetitions. The stimuli were randomly associated to each list in consideration of an even distribution across speakers, groups, and places of articulation.

7.1.3 Procedure

The perception experiment was conducted as an online-based study using the Percy software framework developed by employees of the phonetics institute at the university of Munich (Draxler, 2011, 2014). Participants had access to the experiment via a specific web address. Before starting the experiment subjects were asked to fill in a short questionnaire. Besides basic information about gender and age they also had to answer the following questions:

Table 7.2: A list of 22 metropolitan regions in France that the participants could choose from and number of participants from these regions.

Abbreviation	French Region	Number of Participants
FR-A	Alsace	2
FR-B	Aquitaine	1
FR-C	Auvergne	1
FR-D	Bourgogne	0
FR-E	Bretagne	0
FR-F	Centre-Val de Loire	1
FR-G	Champagne-Ardenne	0
FR-H	Corse	0
FR-I	Franche-Comté	2
FR-J	Île-de-France	8
FR-K	Languedoc-Roussillon	1
FR-L	Limousin	0
FR-M	Lorraine	30
FR-N	Midi-Pyrénées	0
FR-O	Nord-Pas-de-Calais	1
FR-P	Basse-Normandie	1
FR-Q	Haute-Normandie	0
FR-R	Pays de la Loire	0
FR-S	Picardie	1
FR-T	Poitou-Charentes	0
FR-U	Provence-Alpes-Côte d'Azur	0
FR-V	Rhône-Alpes	1

- Where did you grow up/Where have you lived for the last five years? (All French regions were listed.)
- Is French your native language?
- Do you have any additional native languages? (If yes, please state.)
- Did you learn German as a foreign language? (If yes, please state how long you have learned German.)

The questionnaire included a question about where the participants grew up or lived for more than five years because of the different dialects spoken in France. These might have an influence on the perception of the produced stops and the assessment of the foreign accent. The possibility to choose ‘other countries‘ was given, if a participant grew up in another French speaking country, like Belgium or Canada. However, this experiment focused on speakers from France since Belgian or Canadian French differs from French spoken in France (refer to Pustka (2011) for more information). A list of the available regions and the number of participants from these regions can be found in Table 7.2.

The perception experiment started directly after the questionnaire was finished. A short introduction screen gave all necessary information. Participants were asked to wear headphones throughout the entire experiment. They heard one stimuli at a time and had to decide whether they heard the word with a voiced or voiceless stop. Both options were given and could be clicked. Then they were asked to evaluate the foreign accentedness on a scale from 1 (not foreign accented/native) to 7 (strongly foreign accented). It was pointed out that the words they were presented with were cut from a continuous recording which might have affected the perception of the words, for example when the speaker had a high speaking rate, extracted short words might sound a bit weird or unnatural. They were asked to ignore this fact for the evaluation of foreign accentedness, to the extent possible.

7.2 Hypotheses

Regarding identification and foreign accentedness ratings, the following hypotheses are made:

(1) Perception of foreign accented French voiced and voiceless bilabial stops is affected by L1, resulting in a moderate error rate in the identification test following Best (1994).

(2) Items of the first recording (before training) show no difference in identification between groups.

(3) Items of the second recording show a difference between groups with better identification for voiceless stops of the Native Group in contrast to the Control Group and Manipulation Group according to the significant acoustic improvement. No difference is expected for voiced stops.

(4) Items of the first recording (before training) show no difference in foreign accentedness rating between groups.

(5) Items of the second recording show a difference with lower accentedness rating for voiceless stops for the Manipulation and Native Group according to the significant acoustic improvement. Accentedness scores are expected to be lower for the Native Group. No difference is expected for voiced stops.

7.3 Results

The results of the perception experiment for the identification test were entered into a Generalized Linear Mixed Model using JMP (JMP, 2014) and the model was performed with a binomial distribution and Logit link function. Overall, 1106 values were excluded from the statistical model which included all data points from participants who did not complete the experiment. Correctness (0 or 1) was entered as dependent factor and Speaker, Listener, and Item were entered as random factors. The effects Region, List, and Gender Speaker were included as independent factors as well as Test (pre/post-test), Group (Control/Manipulation/Native), Phonation, and poa as well as all their interactions, but without the 4-way-interaction.

The evaluations for perceived foreign accentedness (1-7) were entered into a Linear Mixed Model. Accentedness was entered as dependent factor and Speaker, Listener, and Item were entered as random factors. The effects Region, List, and Speaker Gender were included as independent factors as well as Test (pre/post-test), Group (Control/Manipulation/-Native), Phonation, and Place of Articulation as well as all their interactions, but without the 4-way-interaction

7.3.1 Correctness

Table 7.3 illustrates the results of the statistical analysis. A main effect was found for List ($\chi^2(4)$=41.1, p<0.0001) which indicates that the items in the different lists were not identified equally well. Because items were assigned randomly to each list this might suggest that certain items produced by certain speakers were identified better or worse than others. However, the reason for this result cannot be explained with certainty at this point. Figure 7.1 illustrates that the correct identification rate was relatively high for all lists, however, list 3 seems to include more items that were difficult to identify with about 14% incorrect answers.

The effects Test ($\chi^2(1)$=31.7, p<0.0001) and Group ($\chi^2(2)$= 42.3, p<0.0001) also show main effects indicating different identification behavior for pre- and post-test items as well as items produced by participants of the

Table 7.3: Statistical results of the Generalized Linear Mixed Model for correct and incorrect answers of the perception experiment.

source	**df**	χ^2	**p**
List	4	41.091875	<0.0001*
Test	1	31.740794	<0.0001*
Group	2	42.31727	<0.0001*
Test×Group	2	14.132462	0.0009*
Phonation	1	7.623927	0.0058*
Test×Phonation	1	6.0716121	0.0137*
Group×Phonation	2	69.985348	<0.0001*
Test×Group×Phonation	2	15.321513	<0.0001*
poa	2	141.41905	<0.0001*
Test×poa	2	4.9719055	0.0832
Group×poa	4	9.3100712	0.0538
Test×Group×poa	4	6.0358125	0.1965
Phonation×poa	2	28.155984	<0.0001*
Test×Phonation×poa	2	3.0139969	0.2216
Group×Phonation×poa	4	14.369973	0.062
Region	12	76.168143	<0.0001*
Speaker Gender	1	0.6142319	0.4332

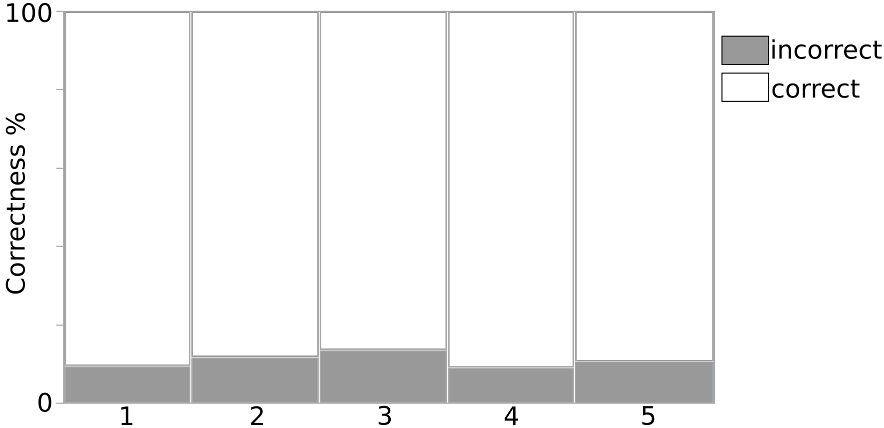


Figure 7.1: Percentage of correct and incorrect identifications for all five lists.

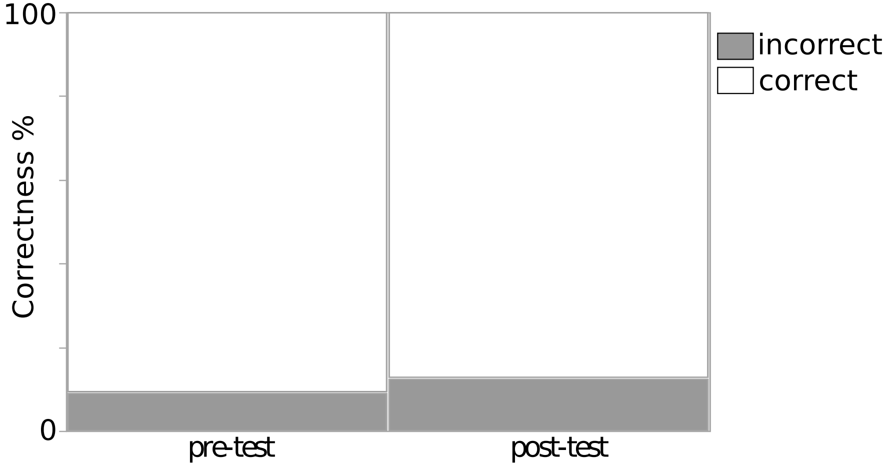


Figure 7.2: Percentage of correct and incorrect identifications for pre- and post-test conditions.

different groups. Figures 7.2 and 7.3 show the percentages of correct and incorrect identifications for each effect. Interestingly, more items were identified incorrectly in the post-test condition than in the pre-test condition. Another curious result is that productions by the Native Group were identified worse than for the Control Group ($\chi^2(1)=12.1$, $p<0.001$) and Manipulation Group ($\chi^2(1)=42.3$, $p<0.001$).

The interaction Test×Group was also significant ($\chi^2(2)=14.1$, $p<0.001$). Post-hoc tests revealed that there was no significant difference between the identification of the first and second recording for the Control and Manipulation Group ($\chi^2(1)=3.2$, $p=0.0736$ and $\chi^2(1)=2.8$, $p=0.0951$, respectively). However, significantly more items were identified incorrectly for the Native Group (see Figure 7.4).

Furthermore, Phonation shows a main effect ($\chi^2(1)=7.62$, $p<0.01$). Post-hoc tests showed that items starting with a voiceless stop were identified more often correctly than voiced stops. Although the difference is significant, the difference between voiced and voiceless stops is marginal. Voiced stops were identified incorrectly about 11.5% and voiceless stops about 10.5%. This difference is mostly consistent within the interactions Test×Phonation ($\chi^2(1)=6.07$, $p<0.05$), Group× Phonation ($\chi^2(2)=69.99$, $p<0.0001$), and Test×Group×Phonation ($\chi^2(2)= 15.32$, $p<0.0001$) with one exception: voiceless items by participants of the Native Group were identified incorrectly more often in the first and in the second recording than voiced stops.

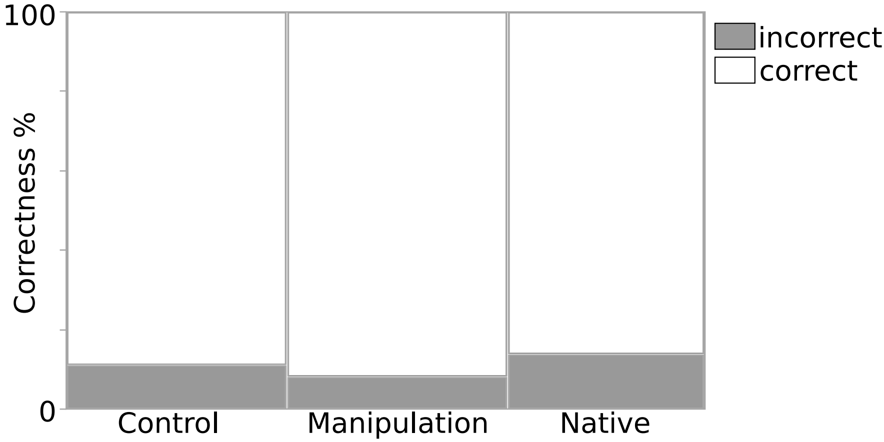


Figure 7.3: Percentage of correct and incorrect identifications for the three groups.

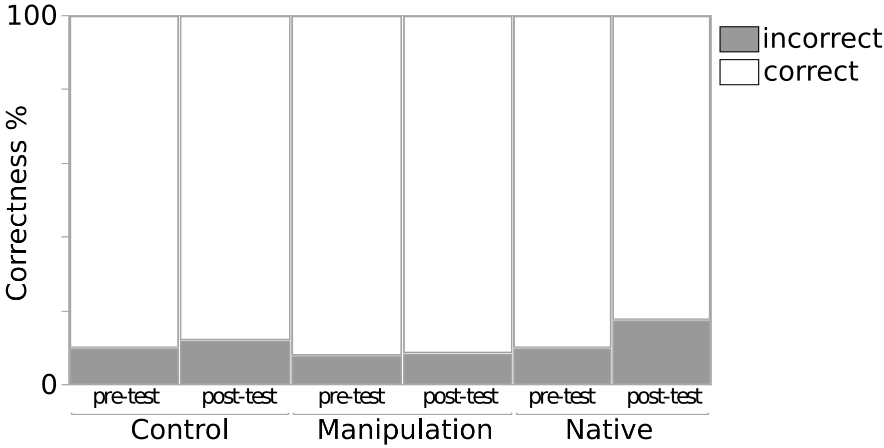


Figure 7.4: Percentage of correct and incorrect identifications for the interaction TEST×GROUP.

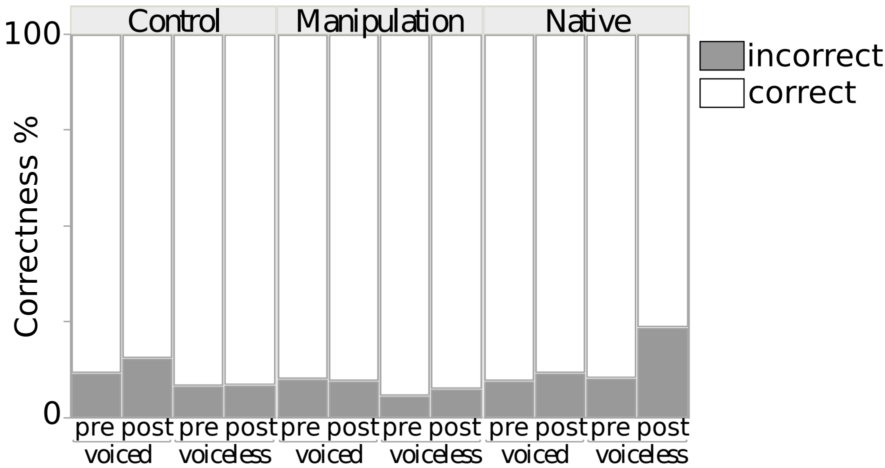


Figure 7.5: Percentage of correct and incorrect identifications for the interaction TEST×GROUP×PHONATION (pre = pre-test, post = post-test).

Interestingly, the percentage of wrongly identified items increased for voiceless stops in the second recording for the Manipulation and the Native Group as well as for voiced stops for the Control and Native Group (see Figure 7.5).

PLACE OF ARTICULATION also shows a significant effect ($\chi^2(2)$=141.42, p< 0.0001) indicating a difference for items with velar stops which were identified more often incorrectly (see Figure 7.6). With the exception of PHONATION×POA ($\chi^2(2)$=28.16 , p<0.0001) no other interaction with POA is significant. This interaction shows that within each phonation condition, all places of articulation differ from each other regarding correct identification. However, when comparing voiced and voiceless items within each place of articulation, only bilabial stops show a difference with a higher percentage of wrong identifications for the voiced plosives (see Figure 7.7).

Lastly, REGION seems to play an important role in the identification experiment ($\chi^2(12)$=76.17 , p<0.0001). Figure 7.8 illustrates the percentage of the correctly identified items. It is obvious that listeners from all regions performed relatively well ranging from 95% for speakers from Alsace (FR-A) to 83% for speakers from Franche-Comté (FR-I) and Rhône-Alpes (FR-V). However, the different French dialects might have influenced the way that listeners hear and judge the produced items by the German learners of French. However, these results can also be attributed to listener specific judgments since some regions were only represented by one or two speakers. No other effects or interactions are significant.

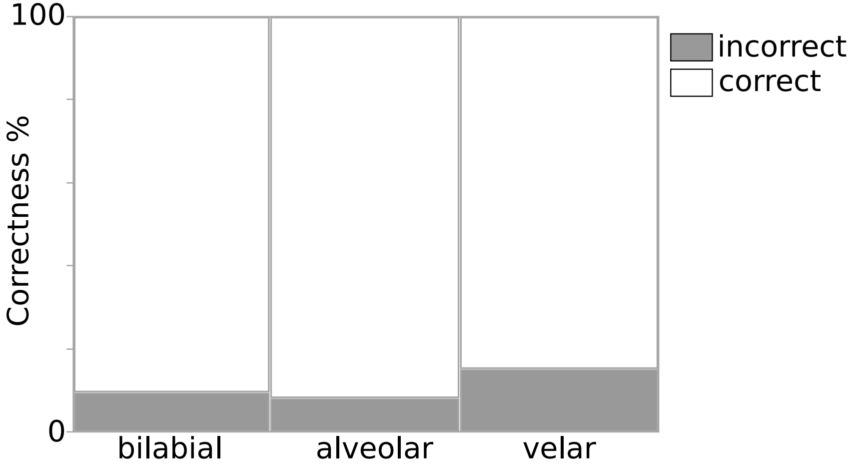


Figure 7.6: Percentage of correct and incorrect identifications for bilabial, alveolar and velar stops.

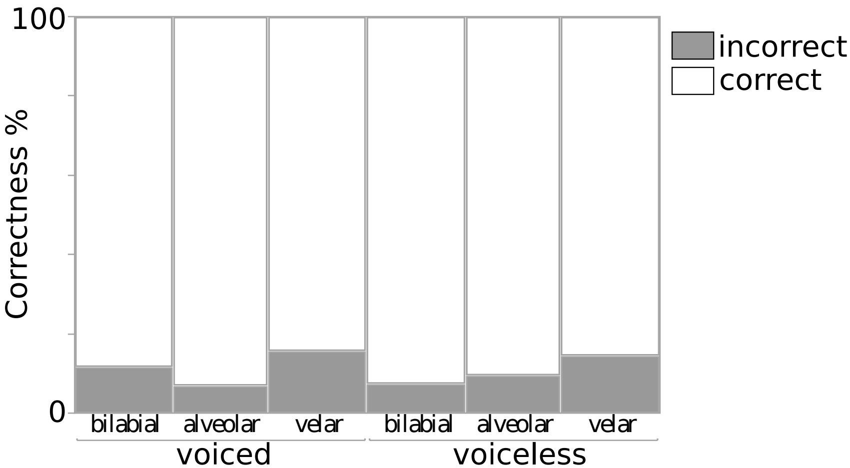


Figure 7.7: Percentage of correct and incorrect identifications for voiced and voiceless bilabial, alveolar, and velar stops.

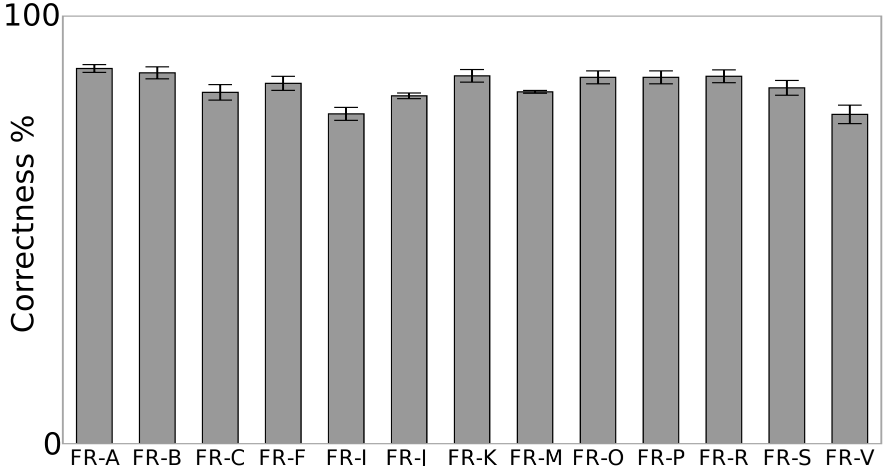


Figure 7.8: Percentage of correct identifications for listeners from the different French regions.

Although not included as an independent factor in the model, taking a look at the individual speakers shows that productions of some speakers were identified correctly almost all of the time (e.g., 96.1% correct identifications for speaker 14). However, there are also speakers whose productions were more problematic to identify (e.g., 77.8% correct identifications for speaker 29). It is also noticeable that more speakers of the Native Group received lower correctness scores than speakers of the other two groups. Possible reasons will be addressed in the discussion.

7.3.2 Foreign Accentedness

Table 7.4 lists the results of the statistical analysis for the foreign accentedness evaluation. A main effect was found for Test (F(4,34.08)=1.25, p<0.0001) indicating significantly different evaluations for the first and second recording. Figure 7.10 shows that items produced in the first recording received significantly higher accentedness ratings (3.6) than items in the second recording (3.4), although the difference seems marginal.

The interaction Test×Group is also significant (F(2,13292)=14.73, p<0.0001). Post-hoc tests showed that items of the Manipulation (t(1)=4.65, p<0.01) and the Native Group (t(1)=7.25, p<0.0001) received lower accentedness scores in the second recording (see Figure 7.11).

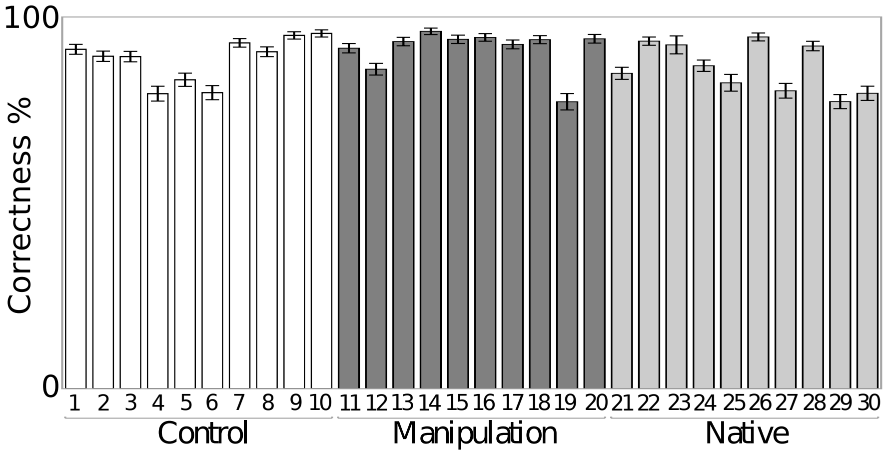


Figure 7.9: Percentage of correct identifications for the individual German learners of French.

Table 7.4: Statistical results of the Linear Mixed Model for the evaluation of foreign accentedness.

source	**between-groups df**	**within-groups df**	F	**p**
LIST	4	34.08	1.2545	0.3068
TEST	1	13292	47.8098	<0.0001*
GROUP	2	27.04	0.1952	0.8239
TEST×GROUP	2	13292	14.7302	<0.0001*
PHONATION	1	17.99	4.8369	0.0412*
TEST×PHONATION	1	13292	5.5779	0.0182*
GROUP×PHONATION	2	13292	4.3811	0.0125*
TEST×GROUP×PHONATION	2	13292	1.523	0.2181
POA	2	18	4.7637	0.0219*
TEST×POA	2	13292	2.3615	0.0943
GROUP×POA	4	13295	9.8742	<0.0001*
TEST×GROUP×POA	4	13292	1.2895	0.2715
PHONATION×POA	2	17.99	0.3527	0.7076
TEST×PHONATION×POA	2	13292	3.0844	0.0458*
GROUP×PHONATION×POA	4	13292	5.0105	0.0005*
REGION	12	34	0.9061	0.5506
GENDER SPEAKER	1	26.03	0.4422	0.5119

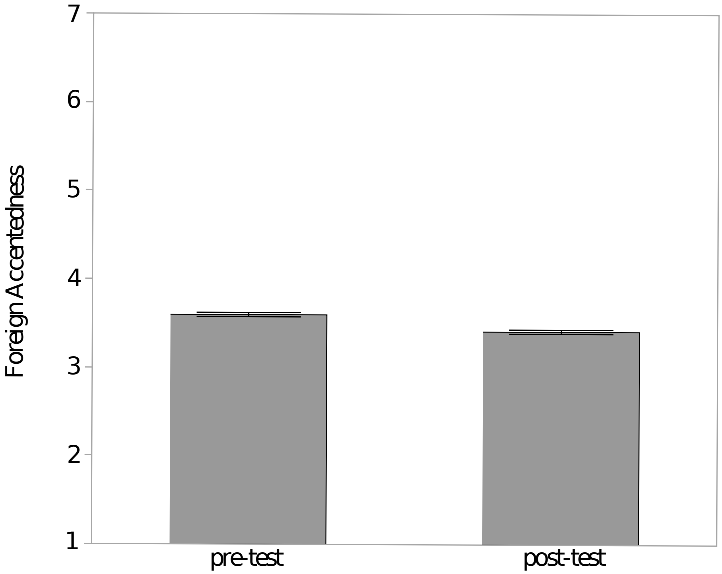


Figure 7.10: Mean accentedness evaluation across the pre- and post-test condition.

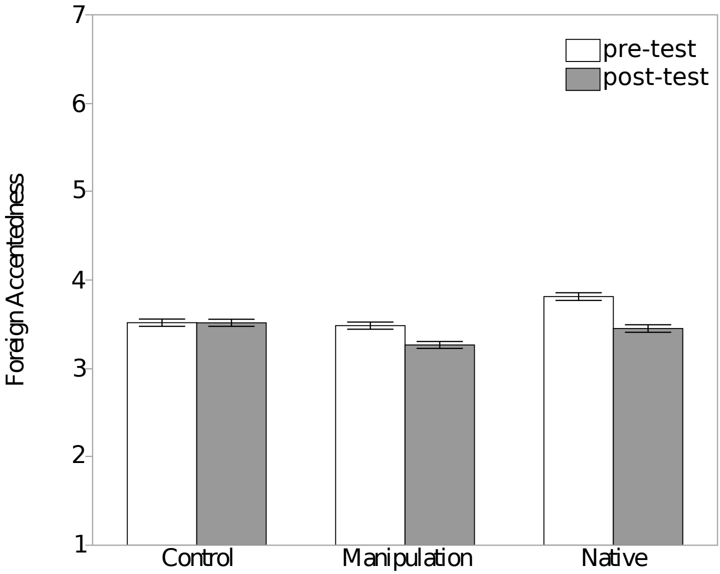


Figure 7.11: Mean foreign accentedness ratings for the pre- and post-test condition and for the different groups.

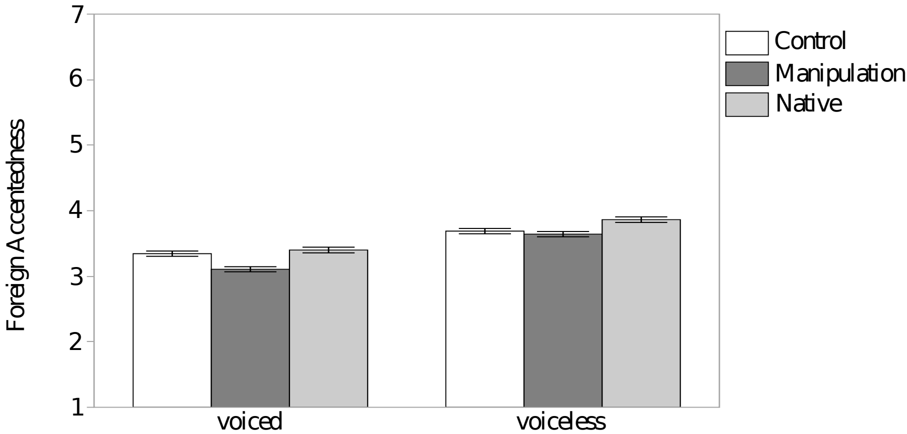


Figure 7.12: Mean foreign accentedness ratings for voiced and voiceless stops for the different groups.

The effect PHONATION (F(1, 17.99)=4.84, p<0.05) shows that voiced stops received lower accentedness scores (3.3) than voiceless stops (3.7). However, post-hoc tests for the interaction TEST×PHONATION (F(1, 13292)=5.58, p<0.05) shows that this difference disappears in the second recording. Furthermore, post-hoc tests for the interaction PHONATION×GROUP (F(2, 13292)=4.38, p<0.05) shows that there is no difference between voiced and voiceless stops between groups. Within groups, only the Manipulation (t(1)=-2.63, p<0.05) and the Native Group (t(1)= -2.18, p<0.05) differ between voiced and voiceless stops (with higher ratings for voiceless stops) as can be seen in Figure 7.12.

PLACE OF ARTICULATION also shows a main effect (F(2, 18)=4.76, p<0.05). Post-hoc tests revealed that velar stops received significantly higher accentedness ratings than bilabial (t(1)=-2.886, p<0.01) and alveolar stops (t(1)=-2.39, p<0.05) (see Figure 7.13). Furthermore, the significant interaction GROUP×POA (F(4, 13295)=9.87, p<0.0001) showed that this pattern reoccurs within all groups except for the Manipulation Group for which no contrast was significantly different (see Figure 7.14).

Taking a look at the post-hoc tests of the significant interaction TEST× PHONATION ×POA (F(2, 13292)=3.08, p<0.05) it can be seen that voiced and voiceless items for all places of articulation received significantly lower accentedness ratings in the second recording except for items with initial voiced alveolar stops. Also, voiceless items tended to get evaluated as more accented than voiced stops. Items with bilabial and alveolar stops are per-

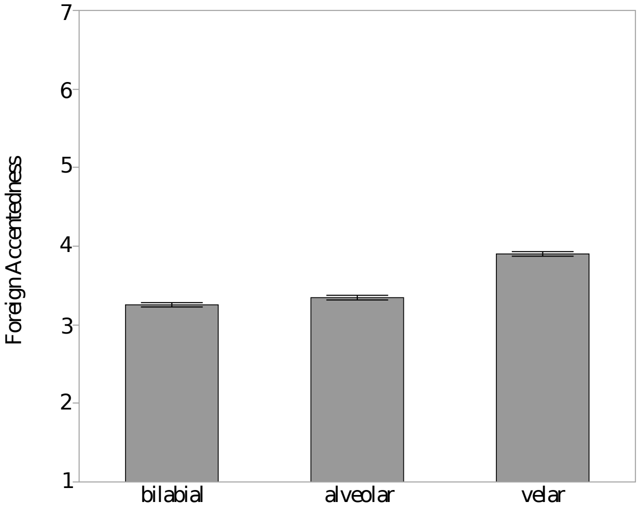


Figure 7.13: Mean foreign accentedness ratings for the different places of articulation.

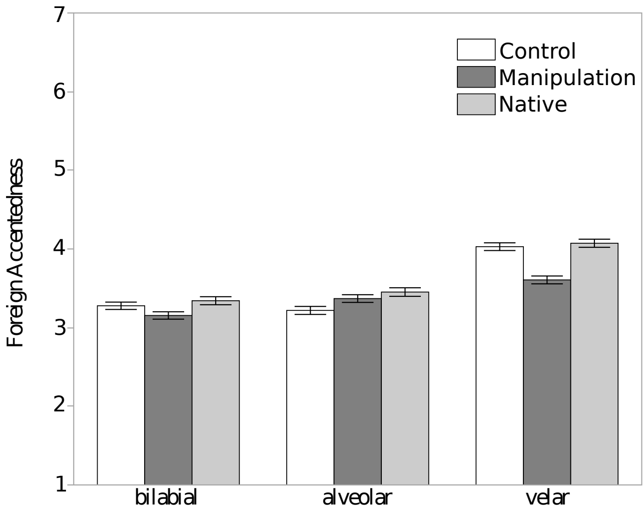


Figure 7.14: Mean foreign accentedness ratings for the different places of articulation and three groups.

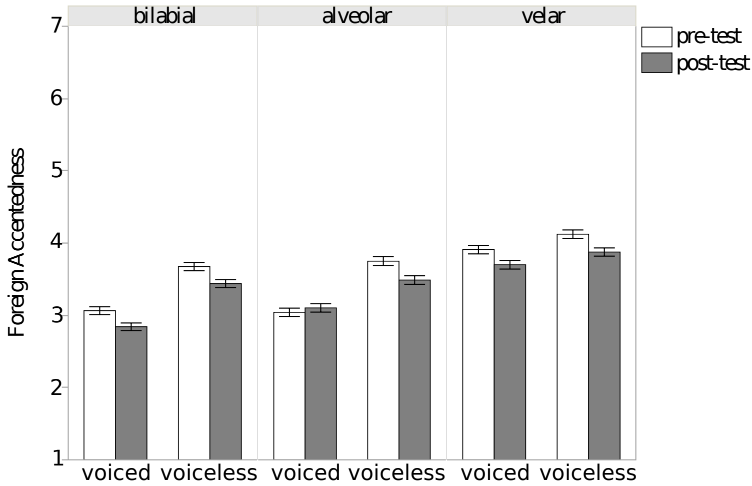


Figure 7.15: Mean foreign accentedness ratings for the interaction TEST×PHONATION×POA.

ceived as less accented than velar stops, as was already discussed. However, the largest difference in mean accentedness ratings is never larger than one rating point (on a scale from 1 to 7).

Lastly, the interaction GROUP×PHONATION×POA ($F(4, 13292)=5.01$, $p<0.001$) is significant and is illustrated in Figure 7.16. It mostly summarizes the results that were already discussed: higher ratings for voiceless stops and velar stops in general. It is noticeable that items of the MG received lower ratings for voiced bilabial and velar, and voiceless velar stops in comparison to the other groups.

Regarding the individual speakers (SPEAKER is not included in the model as a fixed factor) it can be seen that there is a large variability for speakers (see Figure 7.17) ranging from a mean foreign accentedness score of 2.4 (speaker 8 and 18) to 4.9 (speaker 19).

7.4 Discussion

This perception experiment was carried out to investigate how French native listeners evaluated the productions of voiced and voiceless stops by German learners of French from experiment II. This experiment tested two feedback methods and showed that listening to and imitating a French native speaker helped the participants to reduce their production of voiceless stops to a

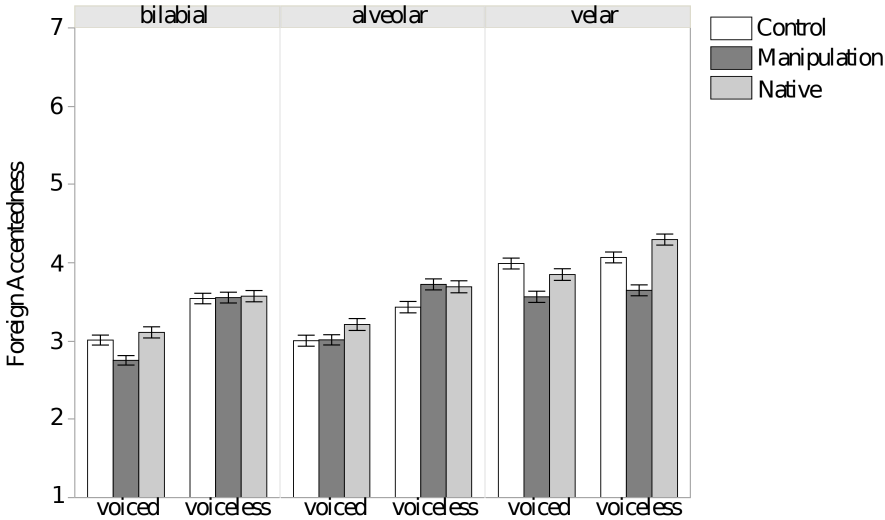


Figure 7.16: Mean foreign accentedness ratings for the interaction Group× Phonation×poa.

level of a native speaker by reducing VOT duration. Although speakers of the Manipulation Group, who listened to their modified own voice, also improved their voiceless stops, they failed to show a significant difference to the Control Group who did not receive any type of feedback or training. Overall, 51 native French speakers were asked to identify isolated words in an identification experiment, and to evaluate the stimuli for foreign accentedness on a scale from 1-7 (1 = not foreign accented/native, 7 = strongly foreign accented).

Regarding the identification scores it was shown that items of the second recording were identified more often incorrectly. This already suggests a change on an acoustic level for the presented stimuli which, however, did not improve but rather corrupt the identification rate. Furthermore, it was shown that stimuli produced by the Native Group were identified incorrectly more often than the items produced by the Control and Manipulation Group (Control: 11.2% , Manipulation: 8.3%, Native: 14% wrongly identified items). Also, significantly fewer items were identified correctly for the Native Group for the second recording in contrast to the first recording. Both recordings of the other groups did not show a significant difference. Furthermore, voiceless stops received lower correctness scores than voiced stops. One reason for this result is that French listeners knew from the beginning that the productions they rated were produced by non-native speakers and therefore adjusted their identification behavior accordingly to identify the initial stops. It was

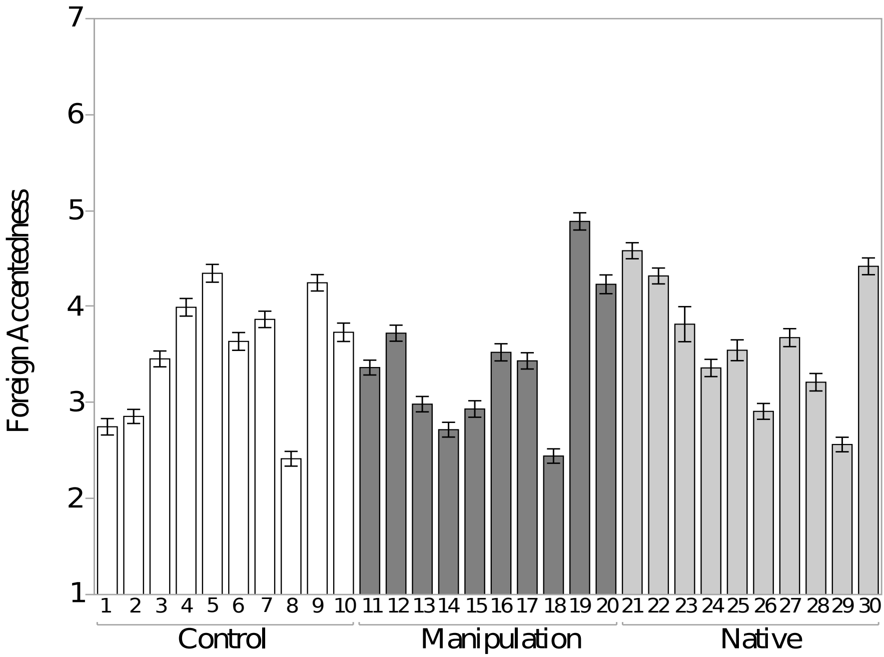


Figure 7.17: Mean foreign accentedness ratings for the individual German learners of French.

already shown that German learners of French have problems producing fully voiced stops, which are expected in native French, and therefore, a different strategy to contrast between non-native voiced and voiceless French stops had to developed by the listeners. For this reason, they were able to hear the difference between both phonetically voiceless stops with a short and long VOT.

However, French listeners seemed to have problems when it came to identifying the productions by speakers of the Native Group from the second recording. The recording in which the learners evidently changed their VOT duration by shortening it to the level of a French native speaker. Table 7.5 from the experiment II shows the mean VOT values for participants of all three groups and from the golden speaker for voiced and voiceless stops. Although there is a clear distinction between voiced and voiceless stops for participants of the Native Group, there are speakers for which this distinction is not as distinct after the training. Figure 7.18 shows the large inter-speaker variability, especially for voiceless stops. Looking at VOT values of voiced

Table 7.5: Mean VOT values (ms) for the experimental groups for the first and second French recording and reference values of the golden speaker.

		Rec 1	**Rec 2**
Control	*voiced*	16	15
	voiceless	48	46
Manipulation	*voiced*	14	12
	voiceless	53	48
Native Speaker	*voiced*	11	11
	voiceless	51	38
Golden Speaker	*voiced*	3	
	voiceless	30	

and voiceless stops, the difference can get smaller for some German learners of French. For example, voiced and voiceless items by speaker 29 might get confused with each other more easily than items by speaker 23 or 25 because the distance between the mean VOT values for voiced and voiceless stops is rather small. Therefore, the production of these stops might fall into one phonetic category.

Regarding foreign accentedness it can be summarized that items of the Manipulation and Native Group received lower accentedness scores in the second recording which coincides with the acoustic improvement after training. It was interesting to see that voiceless stops received higher accentedness scores than voiced stops (3.3 and 3.7, respectively). One might have expected the opposite result because most voiced stops were not produced with a fully voiced closure duration and therefore fall into the voiceless category of the French phoneme inventory. Since rating the foreign accentedness entails more than rating the production of the stop, the overall accentedness was captures by the evaluations. Furthermore, ratings were overall in a moderate range which might be attributed to a certain amount to the still non-native production of stops. Differences in accent rating were mostly observed in the range of one rating point (between 3 and 4). However, individual speakers were rated highly variable with scores from 2.4 (speaker 8 and 18) to 4.9 (speaker 19).

It can be summarized that French native listeners identified German accented French items quite well. However, they seemed to get confused with voiceless items from the Native Group produced after training with a significantly shorter VOT. This result indicates that the acoustic difference

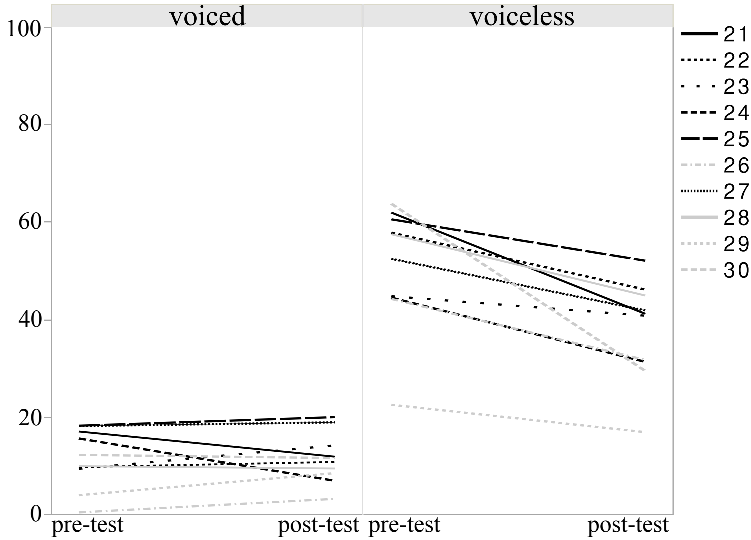


Figure 7.18: Mean VOT values (ms) for the French recordings of the individual speakers of the Native Group for voiced and voiceless stops.

found in experiment II was noticeable for French natives. Generally speaking, training the production of French stops with a native speaker seems to be a promising method which needs further exploration. Note that participants in experiment II only trained their pronunciation for a few minutes.

The following experiment investigated another training procedure exposing participants to a number of French native speakers.

Chapter 8

Experiment IV: High Variability Training

The results of experiment II showed that exposure to and imitation of a native speaker helped German learners of French to improve the pronunciation of voiceless stops. This experiment[9] investigated if the exposure to several native speakers had an even greater impact. It was shown that High Variability Training (HVT), where high-variability refers to the use of multiple model speakers producing the stimuli in identification tasks, is known to contribute to a better performance in both perception and production of difficult sound contrasts, such as the English /l/ - /ɹ/ for Japanese native learners (Logan et al., 1991; Bradlow et al., 1997, 1999). In these studies, participants completed an identification training for minimal pairs contrasting between /ɹ/ and /l/ recorded by five speakers of American English. The training phase was carried out over a period of 3-4 weeks and feedback was provided immediately for correct and incorrect answers. The overall identification accuracy showed a significant improvement after training. Perceptual evaluation of the learners' productions by American English listeners showed that the production of /ɹ/ and /l/ after training received higher rankings than before training. Furthermore, improvements could be maintained even three months after training.

The positive effect of high variability training was also shown in other studies such as Wang et al. (1999) which investigated the learning of Man-

[9]This experiment was first published in its full form as a proceedings article by Jügler et al. (2015) which discussed preliminary results with six participants per group. Within the scope of this dissertation, ten participants per group were analyzed and the acoustic analysis was carried out in more detail.

darin tones by American English native speakers. They trained the identification of the four Mandarin tones using the method applied by Logan et al. (1991) in eight sessions within the course of two weeks. It was shown that training resulted in more native-like productions and that identification rate increased by 21% from pre- to post-test and was also generalized to new, untrained stimuli as well as new Mandarin speakers and stimuli. Furthermore, this improvement was retained even after six months. They also point out that they found high variability among the trained participants and that subjects with an initial lower score improved more than subjects with a higher score in the pre-test.

Sadakataa and McQueen (2013) investigated different training conditions for Dutch learners of Japanese geminate and singleton variants of /s/. This study is of particular interest because it compares two training modalities: low variability training (limited set of words by a single speaker) with high variability training. Additionally, they tested speech and non-speech stimuli and also included a discrimination task. Results showed that HVT led to a better performance than low-variability training and was transferred to the identification of untrained stops and affricates. However, discrimination did not improve after HVT. Also, non-speech materials did not show any improvement for both training methods.

In an earlier study, Wong (2012) conducted an experiment testing the effect of high and low variability phonetic training on the production of the English vowel contrast /æ/-/e/ by Cantonese learners of English. The participants trained in ten perceptual training sessions which lasted for about 10 minutes each. Wong (2012) found that both training approaches improved the perception of the two vowels. However, similar to Sadakataa and McQueen (2013), the improvement showed a stronger effect for participants who trained with high variability training. Additionally, subjects generalized to new words and speakers and also improved their production of the vowels (18% for high variability training and 7% for low variability training). In a later study, Wong (2014) preformed a similar experiment in which she took different proficiency levels into account (besides low and high variability training). Cantonese learners of English trained on the vowel pair /e/-/æ/ in ten perceptual training sessions which lasted for about 10 minutes each. Results were similar to the earlier study. Participants in both training conditions improved but subjects training with HVT improved more. Additionally, subjects generalized to new words and speakers and also improved their production of the vowels (22% for HVT and 16% for LVT). Interestingly, learner proficiency did not have a significant effect. In another study, Wong (2015) conducted a similar study training Cantonese learners of English on the English /iː/-/ɪ/ contrast. It was shown that male learners improved their

F1, F2 and vowel duration after 20 training sessions of 10 minutes whereas female speakers only improved in terms of F1 and vowel duration. Although there are strong individual differences, Wong (2015) concludes that HVT is a good learning alternative that can be incorporated in school since it is not a complex design and not difficult to develop.

Another study testing the effect of HVT was conducted by Lambacher et al. (2005). They trained Japanese Learners of English on the five American English vowels /æ, ɑ, ʌ, ɔ, ɝ/. Participants trained over a period of six weeks using an identification task. In contrast to other studies, the identification task did not consist of a binary forced-choice task but included all five vowels. The vowels were analyzed acoustically and evaluated by native American English speakers. Results showed that the learners improved in their identification of the vowels and that their production of the vowels improved as well.

Nishi and Kewley-Port (2007) investigated the effect of HVT on learning non-native vowel contrasts by Japanese learners of English. They point out that most studies focusing on training vowels never trained more than five vowels in total. For this reason, they examined the influence of training set size for improving in vowel identification. Participants trained for nine days for 90 minutes per session either on nine American English monophthongs (full set) or on three more difficult vowel contrasts (subset). The performance of the participants was tested before and after training as well as three months after training. Results showed that both training groups improved after training, but the group training on the full set improved more than the subset group, and generalized to untrained words by new speakers. Additionally, improvements could be maintained even after three months. However, the subset group did not show any improvements for untrained vowels that were included in the full set condition. Nishi and Kewley-Port (2008) extended their previous study and included Korean learners of English. In this experiment they tested three different set conditions: full set, three days on subset and six days on full set, and six days on full set and three days on subset. The two hybrid conditions were tested to investigate whether a training condition focusing on more difficult vowel contrasts while still maintaining a full set learning condition leads to a better performance and stronger improvement. Results indicate that participants who trained with the full set improved more than the two hybrid conditions.

Although not HVT, Rochet (1995) showed that Chinese learners of French were able to improve their perception and production by training on synthetic syllables of a /bu/-/pu/ continuum. For the pre-test, participants first had to imitate words recorded by a native French speaker that contained initial and intervocalic voiced and voiceless bilabial, alveolar, and velar stops followed by /u/ and bilabial stops before /a/ and /i/. At the same time, they were asked to identify the initial consonant for each word. Secondly, they had to identify synthetic CV stimuli with a VOT varying from -60 to 130 ms for which C was either a bilabial stop followed by /a, i, u/ or an alveolar or velar stop followed by /u/. Regarding training, only bilabial stops followed by /u/ were used in seven training sets. A fading technique was used which incrementally reduced the magnitude of the contrast to help learners to focus on small acoustic details such as VOT. Participants were allowed to continue to the next training set when they received a success rate of 95%. Results of the post-test showed that the training was successful and that the boundary between voiced and voiceless stops shifted to 28 ms (in contrast to the pre-test categorical boundary of 40 ms). The effect was transferred to labial stops followed by /a, i/, alveolar and velar initial stops. However, intervocalic stops did not improve significantly. It was also shown that identification for voiceless initial stops of natural French stimuli as well as the production of voiceless stops improved.

Building new phonetic categories resulting from perceptual training with synthetic stimuli was also shown for English native speakers learning the three-way distinction between voiced, voiceless unaspirated, and voiceless aspirated stops differing in voice onset time (McClaskey et al., 1983). They created two sets of synthetic CV syllables with bilabial and alveolar stops that differed in VOT from -70 to 70 ms. Participants passed through an identification pre-test for two categories corresponding to the English categories, a familiarization phase in which they listened to a set of ordered stimuli with -70, 0, and 70 ms, a training phase with a three category identification task for bilabial stimuli only, and an identification post-test for three categories (as opposed to the two-way distinction from the pre-test). On the second day, the last two phases were repeated, again, using bilabial stimuli to test whether possible improvements are still existent. Finally, participants had to transfer their gained knowledge onto alveolar stops. They showed that the subjects were able to extract useful information from the training session and were able to reliably identify the stimuli within the three categories fully voiced, voiceless unaspirated, and voiceless aspirated. Also, subjects generalized to new stimuli (new place of articulation). Generalizability for place of articulation was also shown by Pisoni et al. (1982) who also trained American English speakers on a 3-way stop distinction using synthetic CV stimuli.

However, Kartushina et al. (2015) argued that training usually only improves the training modality, i.e. when training on production, perception will only benefit little from it, if at all, and the other way around when training on perception. In contrast to the study discussed here, Kartushina et al.

(2015) investigated the effect of production training with visual feedback on the perception and production of non-native Danish vowels. They showed that pure production training did also improve the perception of the vowels but that improvements are not balanced, i.e. they described an improvement of only 5% in perception but 17% in production.

8.1 Experiment

Since HVT seems to be a reliable approach to improve both the production and the perception of difficult L2 sound contrasts, we investigated the effect on the production and perception of French fully voiced and voiceless unaspirated plosives with a short VOT by German native speakers. As mentioned before, German stops are distinguished differently, which results in difficulties regarding the correct pronunciation and perception. Because both German and French have a binary distinction of voiced and voiceless sounds it is interesting to see how well German learners of French are able to hear the differences between voiced and voiceless stops and whether and to what extent they are able to adopt a near-native French pronunciation by training on perception only.

8.1.1 Subjects

A Control and an Experiment Group were tested in this experiment. Subjects of the Experiment Group received feedback and completed a set of training sessions whereas subjects of the Control Group did not receive any feedback nor training.

Each group consisted of three male and seven female speakers (16-37 years, M: 24.2 years, SD: 4.87 years) with basic knowledge of French (A1-A2 level according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)). All participants were students or employees at Saarland University, one participant was a high school student.

8.1.2 Material

For each sound contrast (/b-p/, /d-t/, /g-k/), French minimal pairs differing in initial, medial, and final word position were selected (see examples 1-3).

(1) **b**ain [bɛ̃] (*bath*) vs. **p**ain [pɛ̃] (*bread*)
dé**b**it [debi] (*debit*) vs. dé**p**it [depi] (*pique*)
trom**b**e [tʀɔ̃b] (*cloudburst*) vs. trom**p**e [tʀɔ̃p] (*trumpet*)

(2) **d**é [de] (*dice*) vs. **t**hé [te] (*tea*)
é**d**ile [edil] (*municipality*) vs. é**t**hyle [etil] (*ethyl*)
la**d** [lad] (*stable boy*) vs. la**tt**e [lat] (*slat*)

(3) **g**arrot [gaʀo] (*tourniquet*) vs. **c**arreau [kaʀo] (*pane*)
é**g**ard [egaʀ] (*consideration*) vs. é**c**art [ekaʀ] (*gap*)
bè**g**ue [bɛg] (*stammerer*) vs. be**c** [bɛk] (*beak*)

To create a less complex learning environment, only word initial /b/-/p/ contrasts were included in the training sessions. Word medial and final bilabial minimal pairs as well as alveolar and velar pairs in all word positions were included in generalization tests after the post-test. These tests were used to examine whether any improvements can be transferred to different word positions and different places of articulation (as shown by Logan et al., 1991; Bradlow et al., 1997, 1999; Rochet, 1995; Pisoni et al., 1982).

8.1.3 Procedure

The experiment comprised a production and a perception task. Participants were tested on different days in a quiet office room, whereas the training was performed online at home (Figure 8.1).

During the first session, subjects of both groups were asked to produce 146 French words, which included words of all places of articulation and word positions, to record a baseline for later comparisons. These words were part of the perception test of the pre- and post-test as well as of the generalization tests (see Table 8.1). Recordings were made in quiet office rooms using a head mounted microphone (16 kHz, 16 bit) on an M-AUDIO Fast Track USB device. Recordings were saved on a Windows laptop using a custom-made software developed at LORIA ('Corpus-Recorder', (Colotte, 2013)). The words were presented to each speaker in a randomized order. Then the perception test was performed by all participants (pre-test). It was set up as an online experiment using the Percy framework (Draxler, 2011, 2014). Participants of both groups were asked to listen to isolated French words spoken by a male (34, Bitche, Lorraine) and a female (28, Strasbourg, Alsace) French native speaker and were presented with two buttons displaying the voiced and voiceless orthographic variant of the auditory stimulus. They had to decide which variant was presented. Participants did not receive any feedback in this part of the experiment and were allowed to listen to the presented word only once.

The post-test production and perception experiment was performed three weeks later and were extended by three generalization tests, which included additional words differing in word position and place of articulation (see

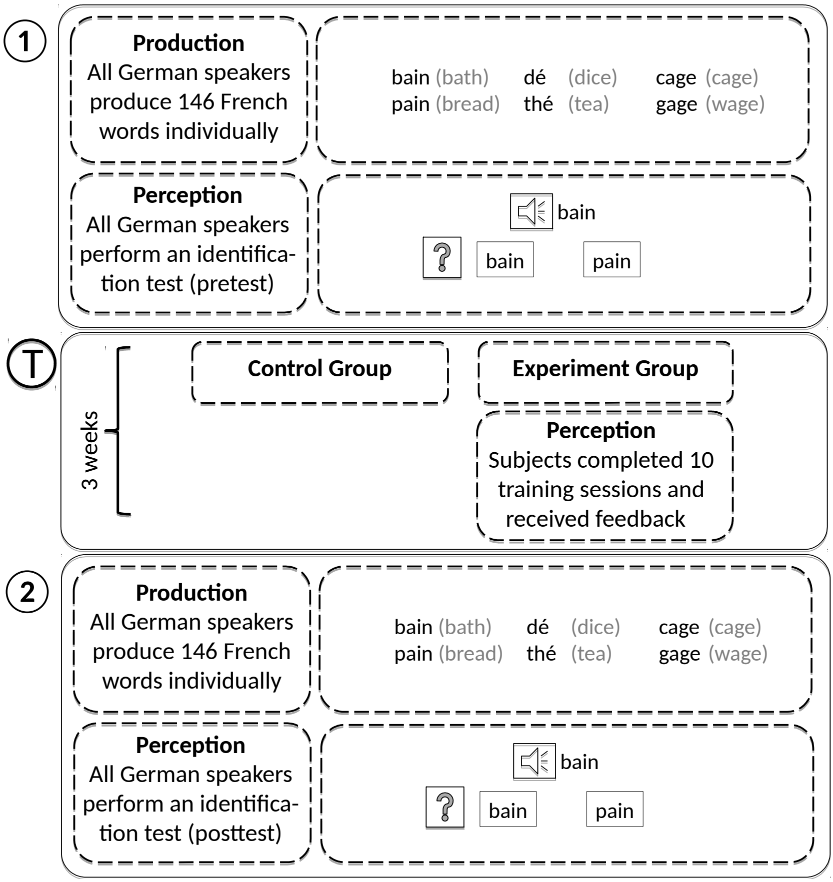


Figure 8.1: Overview of the HVT procedure.

Table 8.1: Number of French words used in the identification tests. Words differed between tests.

	/b-p/	/d-t/	/g-k/	∑
Pre-/Post-test	64 initial	8 initial	8 initial	80
Generalization	8 initial 6 medial 8 final	8 initial 6 medial 8 final	8 initial 6 medial 8 final	66
				146

Table 8.1). In the three weeks between the first and second session, subjects of the Experiment Group had to perform ten training sessions at home. They were instructed to distribute the sessions evenly and only perform one training per day. It was also suggested that they perform the training in a quiet environment and use headphones all the time. The training included 60 bilabial French words which were included in the pre- and post-test. At this point, words produced by six different French native speakers (three male, three female) were presented to the participants, including the two French speakers from the pre- and post-test. Participants of the Control Group did not perform any training sessions.

In the training sessions, feedback was given by changing the color of the pressed button: green for correct response, i.e. the word matched the audio played, red for a mismatch between audio and response. Furthermore, the feedback box displayed a corresponding green check mark or a red X. The green feedback was displayed for 500 ms, the red for 750 ms. Correct input thus led to an overall reduction in experiment duration, which may be an additional incentive for training.

8.2 Hypotheses

The following predictions were made:

(1) Interferences of the L1 result in longer positive VOT values for voiceless stops as well as unvoiced productions of phonologically voiced stops by German learners of French due to a transfer of phonetic features from the native language (see Speech Learning Model (SLM (Flege, 1995) or Natural Language Magnet Model (NLM) (Kuhl and Iverson, 1995)).

(2) Perception of French voiced and voiceless bilabial stops is affected by L1 interferences, resulting in a moderate error rate in the identification test (see Perceptual Assimilation Model (PAM) (Best, 1994) and Second Language Linguistic Perception Model (L2LP) (Escudero, 2005; van Leussen and Escudero, 2015)).

(3) High Variabililty Training improves the perception and production of voiced and voiceless bilabial stops for subjects of the Experiment Group.

(4) Improvements can be transferred to other places of articulation and word positions (generalization).

8.3 Results

8.3.1 Perception

The results of the perception experiment for the identification test were entered into a Generalized Linear Mixed Model[10] using JMP 12 (JMP, 2014) and the model was performed with a binomial distribution and Logit link function. Due to differences in the implementation of voicing between the languages, voiced and voiceless stops were analyzed separately. It was also distinguished between the identification of items that received training and untrained items of the generalization tests.

For trained items, CORRECTNESS was entered into the model as dependent factor and SPEAKER and ITEM as random factors. Since the delayed post-test was only carried out for the Experiment Group and to avoid similarities in JMP, additional subsets had to be built. The first subset was analyzed in order to see if participants of the Experiment Group improved in identification after training and if subjects of the Control Group did not improve due to the lack of training (*training subset*). The other subset, including only data by the Experiment Group, was analyzed in order to see whether potential improvements by participants of this group were sustained even after three months (*sustainability subset*). For the *training subset* TEST (first vs. second identification test) and GROUP (Control vs. Experiment) were included as independent factors as well as their interaction. For the *sustainability subset* only TEST was entered into the model.

[10] In comparison to the published proceedings article by Jügler et al. (2015) which included an analysis using a Linear Mixed Model, this chapter makes use of a Generalized Linear Mixed Model. After a discussion with colleagues it was decided that the usage of this statistical model is more appropriate.

Table 8.2: Summary of the different subsets for the perception analysis.

condition	subset	group	test
trained items	*training*	control/experiment	pre/post
	sustainability	experiment	pre/post/delayed post
untrained items	*generalization I*	control/experiment	generalization 1
	generalization II	experiment	generalization 1/2

Similar to trained items, two subsets were built in order to analyze untrained items. To analyze the first set of generalization tests of the Control and Experiment Group, GROUP, WORD POSITION (initial/medial/final) and POA were entered into the model as independent factors, as well as all possible interactions (*generalization I subset*). To compare the second set of generalization tests made after the delayed post-test recording for the Experiment Group with the first set of generalization tests, an additional model was run only for results of the Experiment Group. This model was run without the effect GROUP but to be able to make comparisons with the effect TEST (*generalization II subset*).

WORD POSITION and POA were not included in the model for the trained items because participants only trained on initial, bilabial stops.

Voiced Trained Items

Overall it can be said that participants of both groups showed considerably high identification scores from the beginning in identifying voiced initial, bilabial stops (see Figure 8.2). Table 8.3 illustrates the results of the statistical analysis of voiced trained items of the pre- and post-test for the Control and Experiment Group (*training subset*). Only GROUP showed a main effect ($\chi^2(1)=10.86$, $p<0.001$). Subjects of the Control Group identified about 97.4% items correctly whereas participants of the Experiment Group identified about 95.6% correctly. However, this small difference is still significant. No other factors showed an effect.

For the *sustainability subset*, TEST showed a main effect ($\chi^2(2)=12.64$, $p<0.01$). Post-hoc tests showed that participants of the Experiment Group got significantly worse after training (95.5% for pre-test and 94.7% for post-test, ($\chi^2(1)=4.81$, $p<0.05$), but note the small difference of only 0.8%), but that they improved again after the three months break to the level of the pre-test.

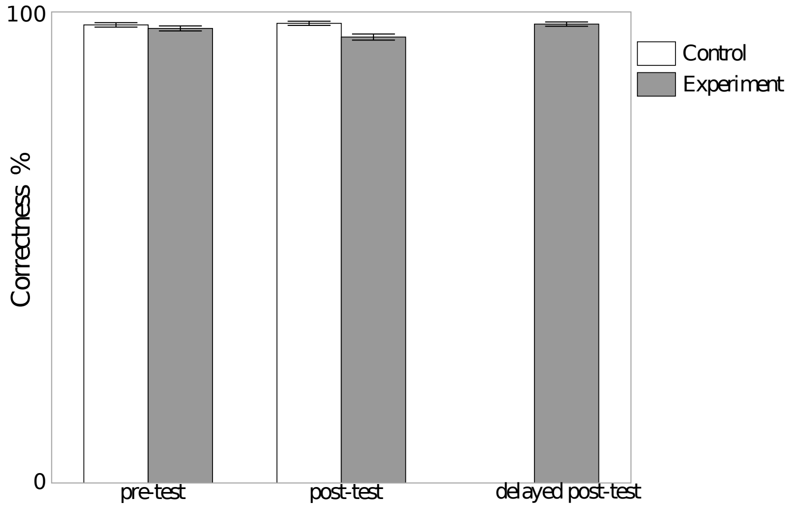


Figure 8.2: Correct identifications (in percent) for trained voiced items for the Control and Experiment Group.

Voiceless Trained Items

Although participants were not as good in identifying voiceless stops as they were for voiced stops they still showed a considerably high identification rate (see Figure 8.3). TEST showed a main effect ($\chi^2(1)=105.87$, $p<0.0001$) with a higher percentage of correct answers for the post-test condition. This indicates that the perceptual training helped participants to improve the correct identification of items. In the pre-test condition about 78% of the words were identified correctly, which was improved to about 89% in the post-test condition.

Table 8.3: Statistical results of the Generalized Linear Mixed Model for correct and incorrect identifications for trained voiced items of the pre- and post-test (*training subset*).

source	**df**	χ^2	**p**
TEST	1	0.8790945	0.3484
GROUP	1	10.862813	<0.0009*
TEST×GROUP	1	3.0682595	0.0798

Table 8.4: Statistical results of the Generalized Linear Mixed Model for correct and incorrect identifications for trained voiceless items of the pre- and post-test (*training subset*).

source	**df**	χ^2	**p**
TEST	1	105.867	<0.0001*
GROUP	1	1.0363482	0.3087
TEST×GROUP	1	45.944439	<0.0001*

Table 8.5: Correct identifications (in percent) for the pre- and post-test for participants of the Control and Experiment Group for voiceless trained stops.

		incorrect	**correct**
Control	Pre-test	18%	82%
	Post-test	14%	86%
Experiment	Pre-test	26%	74%
	Post-test	8%	92%

The interaction TEST×GROUP also showed a main effect ($\chi^2(1)=45.94$, $p< 0.0001$). Post-hoc tests showed that all contrasts were significant (see Table 8.5). This means that participants of the Control and Experiment Group differed before training and subjects of the Experiment Group were significantly worse than participants of the Control Group. However, subjects of the Experiment Group improved by 18% and although the Control Group did get better as well (by 4%), the Experiment Group outperformed the Control Group. This is strong evidence for the efficacy of the perceptual training.

Regarding the *sustainability subset*, TEST showed a main effect ($\chi^2(2)= 190.17$, $p< 0.0001$). Post-hoc tests showed that, as discussed before, participants of the Experiment Group significantly improved their correctness scores after training. This improvement was sustained even after three month as no significant difference was shown between post-test and delayed post-test ($\chi^2(1)=0.64$, $p=0.42$).

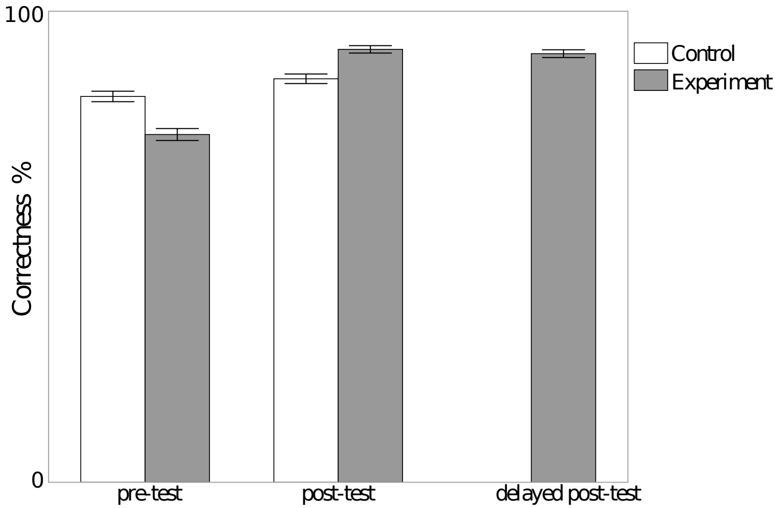
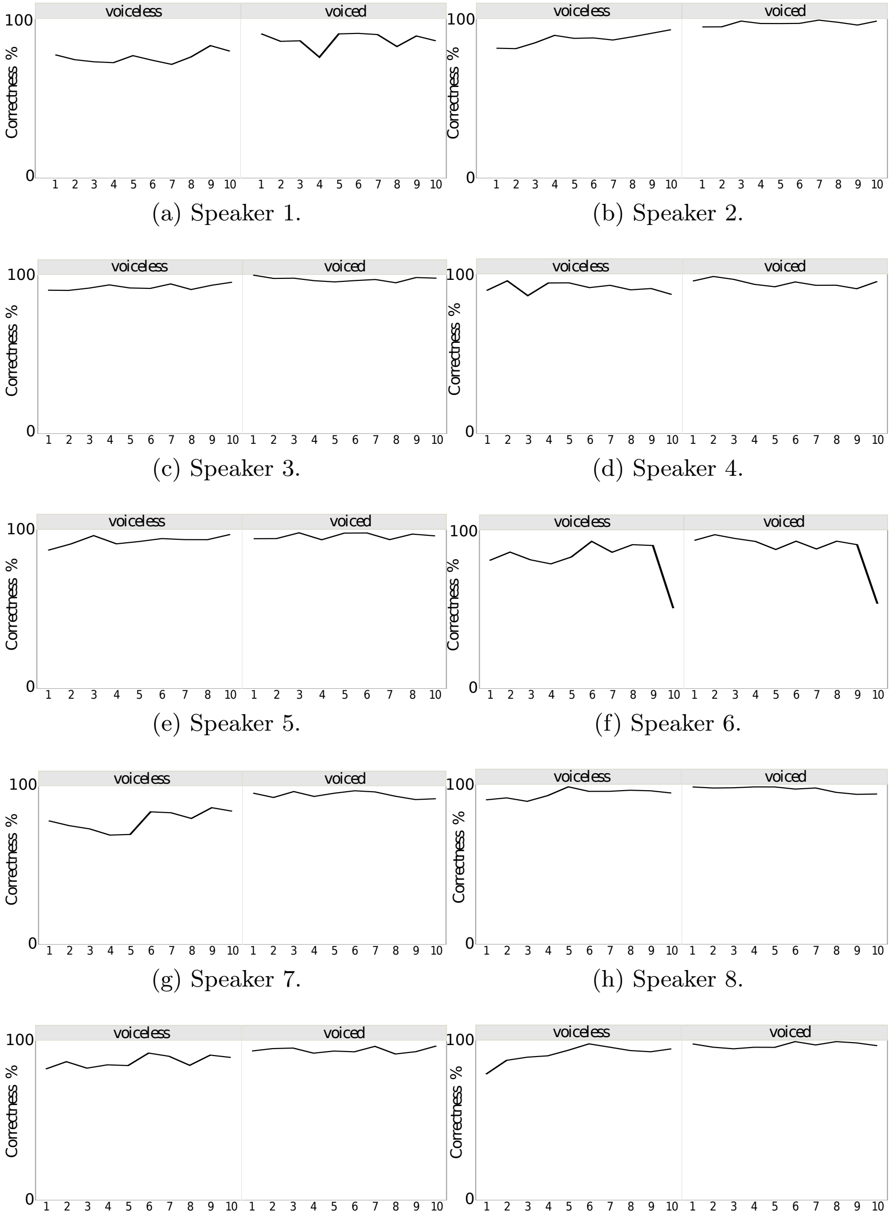


Figure 8.3: Correct identifications (in percent) for trained voiceless items for the Control and Experiment Group.

Individual Behavior During Training

Since each participant of the Experiment Group completed ten training sessions it is interesting to see how well they performed during the individual training units and how and if they improved over time. This is particularly interesting because each learner behaves differently and might apply different strategies while learning. Figure 8.4 shows all individual correctness scores for both voiced and voiceless trained (bilabial, initial) stops. Please note that for speaker 5 no data for the ninth training session was available because the learner accidentally completed the eighth training unit twice. Only data of the first training was included in the analysis.

In general, a steady increase in correctness scores can be observed for voiceless stops. As for the voiced stops, the picture is not as clear. Since the identification for voiced stops is already at ceiling, an improvement is difficult to accomplish. Some speakers show an up and down movement (e.g., speaker 1 or speaker 5) which might be explained by the different sets of words and French speakers in the individual sets, and others seem to get worse over time (e.g., speaker 7). It is also obvious that speaker 6 seemed to have randomly pressed buttons during the last session as the identification scores for both voiced and voiceless stops are at chance level.



(a) Speaker 1.
(b) Speaker 2.
(c) Speaker 3.
(d) Speaker 4.
(e) Speaker 5.
(f) Speaker 6.
(g) Speaker 7.
(h) Speaker 8.
(i) Speaker 9.
(j) Speaker 10.

Figure 8.4: Correctness scores (in %) for voiced (right) and voiceless (left) stops for each participant of the Experiment Group for all ten training sessions.

Table 8.6: Statistical results of the Generalized Linear Mixed Model for correct and incorrect identifications for untrained voiced items of the first set of generalization tests (*generalization I subset*).

source	**df**	χ^2	**p**
GROUP	1	14.594211	0.0001*
POA	2	11.645678	0.003*
POA×GROUP	2	1.6248624	0.4438
WORD POSITION	2	94.623757	<0.0001*
WORD POSITION×GROUP	2	0.145043	0.93
WORD POSITION×POA	4	7.0952623	0.1309
WORD POSITION×POA×GROUP	4	3.091378	0.5627

Voiced Untrained Items

Table 8.6 shows the statistical results of the analysis for voiced untrained items. As can be seen, GROUP shows a main effect ($\chi^2(1)=14.59$, $p<0.001$). Participants of the Experiment Group show a larger percentage of wrong identifications than participants of the Control Group (10.4% vs. 5.1%). It has to be noted that the overall percentage of correct identifications is, despite this difference, still considerably high.

Furthermore, POA shows a main effect ($\chi^2(2)=11.65$, $p<0.01$). Post-hoc tests showed that bilabial and velar stops behave significantly different than alveolar stops, whereby items with alveolar stops showed the highest identification scores (bilabial: 91%, alveolar: 95%, velar: 91%). Why alveolar stops in particular showed higher scores than bilabial and velar stops cannot be answered at this point.

The effect WORD POSITION is also significant ($\chi^2(2)=94.62$, $p<0.0001$). Post-hoc tests revealed that items with final stops have a significantly worse identification rate than items with stops in initial and medial position (initial: 97%, medial: 96%, final: 85%). This result can be interpreted with respect to the final devoicing processes in German. German learners of French might still expect a final stop to be voiceless, independent of the fact that the stop was produced voiced. No other effects were significant.

Regarding the *generalization II subset*, TEST showed a main effect (see Table 8.7). While participants of the Experiment Group identified 89.6% of the items correctly in the first generalization test, they improved by about 4% in the second generalization test (93.5%).

Post-hoc tests of the significant effect POA ($\chi^2(2)=22.51$, $p<0.0001$) showed the same pattern as for the *generalization I subset*. Items with alve-

Table 8.7: Statistical results of the Generalized Linear Mixed Model for correct and incorrect identifications for untrained voiced items of the first and second set of generalization tests (*generalization II subset*).

source	**df**	χ^2	**p**
TEST	1	4.8579292	0.0275*
POA	2	22.513075	<0.0001*
POA×TEST	2	0.4872814	0.7838
WORD POSITION	2	90.663364	<0.0001*
WORD POSITION×TEST	2	1.4352246	0.4879
WORD POSITION×POA	4	4.492326	0.3435
WORD POSITION×POA×TEST	4	4.6711899	0.3227

olar stops received higher correctness scores than items with bilabial and velar stops (bilabial: 91%, alveolar: 94.7%, velar: 89%). Regarding WORD POSITION ($\chi^2(2)$=90.66, p<0.0001), again, a similar picture emerges: final stops show a significantly worse identification than initial and medial stops (initial: 96.3%, medial: 94.7%, final: 84.5%). No other effects were significant.

Voiceless Untrained Items

Regarding the *generalization I subset*, all statistical results can be found in Table 8.8. GROUP shows a main effect ($\chi^2(1)$=15.71, p<0.0001) indicating that participants of the Control Group identified more items correctly than participants of the Experiment Group (87.5% vs. 83.6%). Overall, identification rate is still considerably high but, in general, it can be noted that the identification scores for voiceless stops are lower than for voiced stops.

Furthermore, POA showed main effect ($\chi^2(2)$=17.01, p<0.001) and post-hoc revealed that untrained items with bilabial stops were identified significantly worse than alveolar and velar stops (bilabial: 79.1%, alveolar: 86.8%, velar: 90.1%). It is, however, unclear why items with bilabial stops behaved so much worse than items with alveolar and velar stops. Apparently, training did not help in the identification of untrained bilabial stops and might have had a counter-productive effect. The effect WORD POSITION also showed a significant effect ($\chi^2(2)$=300.02, p<0.0001) and post-hoc tests demonstrated that final stops were identified with a significant lower identification score than initial and medial stops (initial: 92.4%, medial: 97.1%, final: 70.1%).

At last, the interaction WORD POSITION×GROUP revealed a main effect ($\chi^2(2)$=7.84, p<0.05). Post-hoc tests revealed that for the Control Group, identification scores for all word positions differ significantly from each other

Table 8.8: Statistical results of the Generalized Linear Mixed Model for correct and incorrect identifications for untrained voiceless items of the first set of generalization tests (*generalization I subset*).

source	**df**	χ^2	**p**
GROUP	1	15.705039	<0.0001*
POA	2	17.014398	0.0002*
POA×GROUP	2	1.1904435	0.5514
WORD POSITION	2	300.01689	<0.0001*
WORD POSITION×GROUP	2	7.8378017	0.0199*
WORD POSITION×POA	4	3.3083037	0.5076
WORD POSITION×POA×GROUP	4	2.3706516	0.6679

Table 8.9: Statistical results of the Generalized Linear Mixed Model for correct and incorrect identifications for untrained voiceless items of the first and second set of generalization tests (*generalization II subset*).

source	**df**	χ^2	**p**
TEST	1	15.986068	<0.0001*
POA	2	14.864827	0.0006*
POA×TEST	2	0.016697	0.9917
WORD POSITION	2	122.68035	<0.0001*
WORD POSITION×TEST	2	10.92872	0.0042*
WORD POSITION×POA	4	10.17175	0.0376*
WORD POSITION×POA×TEST	4	3.3133403	0.5068

(initial: 92.9%, medial: 99.2%, final: 73.3% correct identifications). For the Experiment Group, initial and medial stops differ from final ones (initial: 91.9%, medial: 95%, final: 66.9% correct identifications). In a direct comparison, both groups only differ in terms of medial stops from each other (99.2% for the Control Group and 95% percent for the Experiment Group). No other effects or interactions were significant.

Regarding the *generalization II subset*, results of the statistical analysis can be found in Table 8.9. TEST shows a main effect ($\chi^2(1) = 15.99$, $p<0.0001$) with better identification scores for the delayed generalization test (GENERALIZATION I: 83.6%, GENERALIZATION II: 92%) which is similar to the results of the untrained voiced items.

Table 8.10: Correctness (in percent) for the first and second generalization tests for the Experiment Group with regard to the different word positions for untrained voiceless stops.

		incorrect	**correct**
initial	Generalization I	8%	92%
	Generalization II	8%	92%
medial	Generalization I	5%	95%
	Generalization II	3%	97%
final	Generalization I	33%	67%
	Generalization II	12%	88%

Furthermore, POA showed a main effect ($\chi^2(2)=14.86$, $p<0.001$) and post-hoc tests demonstrated that, similar to the *generalization I subset*, items with bilabial stops were identified less often than items with alveolar and velar stops (bilabial: 82.4%, alveolar: 88.6%, velar: 92.2%). Again, it is unclear why items with bilabial stops behave differently. One possible explanation might be that training was counter-productive in regard to new, untrained items with the same specification.

Again, WORD POSITION shows a main effect ($\chi^2(2)=122.68$, $p<0.0001$) and post-hoc tests demonstrated that items with final stops received a lower identification score than items with initial and medial stops. Furthermore, words with initial and medial stops also differ significantly from each other (initial: 91.9%, alveolar: 96:3%, final: 77.4%).

In addition, the interactions WORD POSITION×TEST ($\chi^2(2)=10.93$, $p<0.01$) and WORD POSITION×POA ($\chi^2(4)=10.17$, $p<0.05$) are significant. Post-hoc tests of WORD POSITION×TEST revealed that, interestingly, the improvement from GENERALIZATION I to GENERALIZATION II is due to a significant improvement in the identification of items with final stops, as initial and medial contrasts are not significant (see Table 8.10).

Lastly, post-hoc tests of the WORD POSITION×POA interaction showed that regarding initial stops in general, items with bilabial stops were identified less often correctly than items with alveolar and velar stops (bilabial: 84.7%, alveolar: 94.7%, velar: 96.3%). Again, initial bilabial stops is the exact specification of items that were trained but could not be generalized to. It rather resulted in a decline in performance. With regard to final stops, this pattern was not repeated but it was shown that words with final velar stops were identified correctly on more occasions than items with final bilabial and alveolar stops (bilabial: 70.3%, alveolar: 76.8%, velar: 85%).

8.3.2 Production

All productions by the twenty participants were labeled for positive and negative VOT and the data was entered into a Linear Mixed Model using JMP 12 (JMP, 2014). Due to differences in the implementation of voicing between the languages, voiced and voiceless stops were analyzed separately. It was also distinguished between the production of items that received training and untrained items.

For trained items, VOT was entered into the model as dependent factor and SPEAKER and ITEM as random factors. Since the delayed post-test was only carried out for the Experiment Group and to avoid similarities, additional subsets had to be built. The first subset was analyzed in order to see if participants of the Experiment Group improved in the production after training and if subjects of the Control Group did not improve (*training subset*). The next subset, including only data by the Experiment Group, was analyzed in order to see whether potential improvements by participants of this group were maintained even after three months (*sustainability subset*). For the *training subset* FOLLOWING SOUND, TEST (first vs. second production task), and GROUP (Control vs. Experiment) were included as independent factors as well as the interaction TEST×GROUP. For the *sustainability subset* only FOLLOWING SOUND and TEST were entered into the model.

Similar to the trained items, two subsets were analyzed for untrained items which show the same structure as the subsets for the trained items (see Table 8.11). The first subset was analyzed in order to see if participants of the Experiment Group were able to generalize improvements in the production to untrained items and if subjects of the Control Group did not generalize (*generalization I subset*). The second subset, including only data by the Experiment Group, was analyzed in order to see if potential improvements/generalizations by participants of this group were sustained even after three months (*generalization II subset*). To analyze the first generalization subset, FOLLOWING SOUND, GROUP, TEST, WORD POSITION (initial/medial/-final) and POA were entered into the model as independent factors, as well as all possible interactions (without the effect FOLLOWING SOUND and without the four-way interaction). For the second subset an additional model was run only for results of the Experiment Group. This model was carried out without the effect GROUP. WORD POSITION and POA were not included in the model for the trained items because trained items only consisted of items with initial, bilabial stops.

Table 8.11: Summary of the different subsets for the production analysis.

condition	**subset**	**group**	**test**
trained items	*training*	control/experiment	pre/post
	sustainability	experiment	pre/post/delayed post
untrained items	*generalization I*	control/experiment	pre/post
	generalization II	experiment	pre/post/delayed post

Table 8.12: Results of the statistical model for the production of trained voiced stops for the *training subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	1	1216	7.709	<0.0001*
GROUP	1	17.99	0.4002	0.535
TEST×GROUP	1	1216	16.925	<0.0001*
FOLLOWING SOUND	10	22.81	4.8202	0.0009*

Trained Voiced Items

The statistical results for the model of the *training subset* are displayed in Table 8.12. All but one effect was significantly different. FOLLOWING SOUND shows a main effect ($F(10, 22.81)=4.82$, $p<0.001$) confirming that the following context had an influence on the duration of VOT. TEST is also significant ($F(1, 1216)=7.71$, $p<0.01$) confirming that participants across both groups produced significantly longer negative VOT values in the post-test condition (-42 ms) than in the pre-test condition (-34 ms). Post-hoc tests for the interaction TEST×GROUP ($F(1, 1216)=16.93$, $p<0.0001$) revealed that participants of the Experiment Group were able to improve their VOT values significantly from -21 ms to -42 ms (see Figure 8.5). No other contrasts were significant. This, however, indicates that participants of the Experiment Group do not differ from participants of the Control Group after training. In order to improve, the learners were either able to produce lengthened fully voiced closures or increase the number of fully voiced stops or both.

For this reason, the number of fully voiced stops was checked afterwards and a General Linear Mixed Model was carried out with the same specifications as the LMM for VOT (see Table 8.13). It was shown that, although TEST and GROUP did not show a significant effect, FOLLOWING SOUND

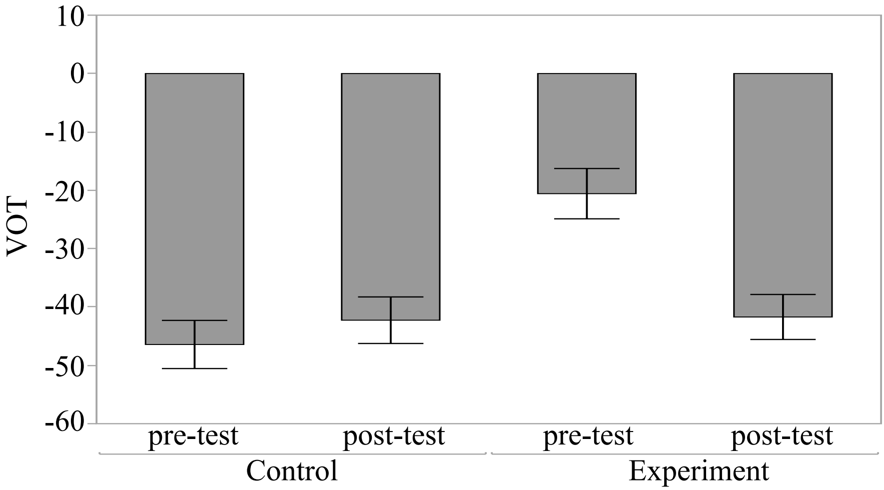


Figure 8.5: Mean VOT (ms) for trained voiced items for the Control and Experiment Group.

($\chi^2(10)$=25.39, p<0.01) and the interaction TEST×GROUP ($\chi^2(1)$=5.9, p<0.05) were significant. Post-hoc tests showed that more fully voiced stops were produced by participants of the Control Group than by learners of the Experiment Group (46% vs. 35%) in the pre-test condition. However, subjects of the Experiment Group were able to significantly increase the number of produced fully voiced stops in the post-test condition to 45% (see also Figure 8.6). Although an improvement could be shown, learners of both groups still produce considerably high numbers of voiceless stops on a phonetic level. It is still interesting to see that participants of the Experiment Group were able to improve their production after training which was only carried out on a perceptual level.

It was noticeable that some speakers actually produced a voiced closure when articulating the voiced bilabial stop but did not manage to maintain this voicing throughout the whole closure. It might be interesting to see whether participants were able, from a phonetic point of view, to shift from a completely voiceless production to a partially voiced production. Stops that were produced partially voiced were marked as such. A General Linear Mixed Model was carried out with PARTIAL VOICING as a dependent factor, SPEAKER and ITEM as random effects and FOLLOWING SOUND, TEST and GROUP as well as the interaction TEST×GROUP as independent effects. The statistical results can be found in Table 8.14. Only FOLLOWING SOUND

Table 8.13: Statistical results of the Generalized Linear Mixed Model for number of fully voiced stops of trained items for the Control and Experiment Group.

source	**df**	χ^2	**p**
TEST	1	1.4151084	0.2342
GROUP	1	2.6657077	0.1025
TEST×GROUP	1	5.904785	0.0151*
FOLLOWING SOUND	10	25.389188	0.0047*

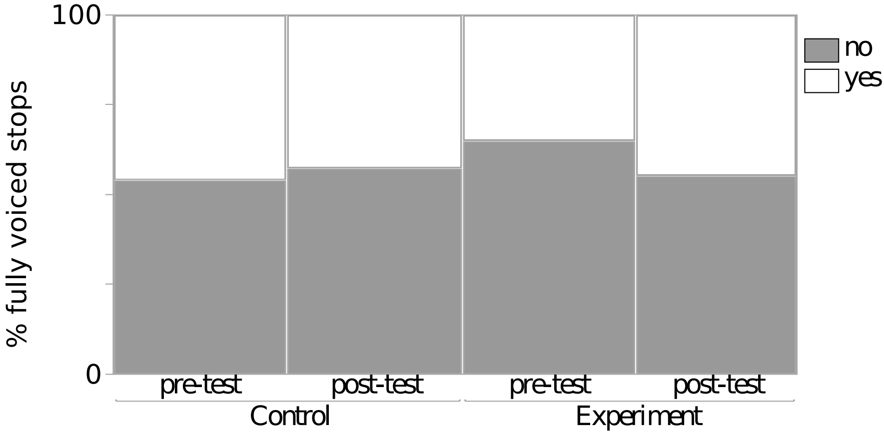


Figure 8.6: Percentage of fully voiced stops of trained items for the Control and Experiment Group for the *training subset*. (yes = fully voiced)

shows a significant effect ($\chi^2(10)$=22.19, p<0.05), no other effects reached significance. This can also be seen in Figure 8.7. Only a considerably small number of stops were produced partially voiced (Control Group: 7.4%, Experiment Group: 8.6%). No difference was found for the pre- and post-test condition for both groups. Please note that it was not checked how many partially voiced stops in the pre-test condition were produced fully voiced in the post-test condition.

With regard to the *sustainability subset*, only TEST reached significance (F(1, 922)=6.34, p<0.01, see Table 8.15). Post-hoc tests showed that, although participants of the Experiment Group managed to improve their VOT values in the post-test condition (from -21 ms to -42 ms), they failed

Table 8.14: Statistical results of the Generalized Linear Mixed Model for number of partially voiced stops of trained items for the Control and Experiment Group for the *training subset*.

source	**df**	χ^2	**p**
TEST	1	0.5902532	0.4423
GROUP	1	0.6754376	0.4112
TEST×GROUP	1	0.2382821	0.6254
FOLLOWING SOUND	10	22.193571	0.0141*

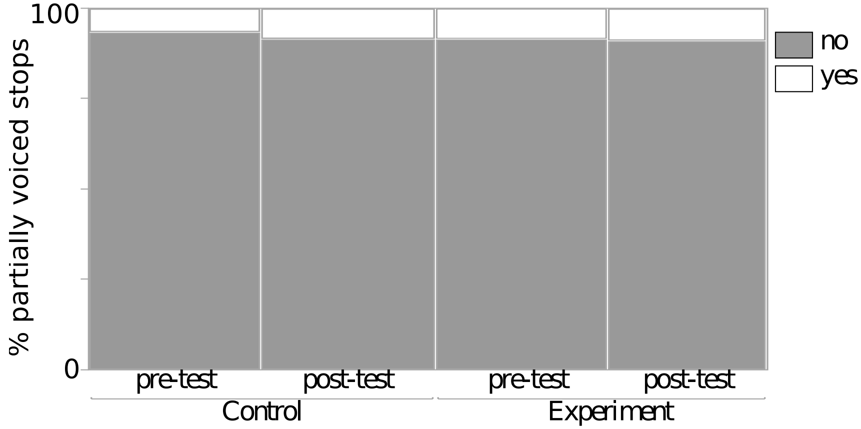


Figure 8.7: Percentage of partially voiced stops of trained items for the Control and Experiment Group. (yes = partially voiced)

to maintain this improvement three months after training (-20 ms, see also Figure 8.8).

For a full picture, the number of produced fully voiced stops were also analyzed again. The Generalized Linear Mixed Model showed that both TEST ($\chi^2(1)=7.14$, $p<0.05$) and FOLLOWING SOUND ($\chi^2(10)=36.75$, $p<0.0001$) showed a main effect. Post-hoc tests for TEST showed a similar picture as the one already described for VOT values. The number of fully voiced stops dropped significantly to the level of the pre-test condition (see Figure 8.9).

Table 8.15: Results of the statistical model for the production of trained voiced stops for the *sustainability subset.*

source	**between-groups df**	**within-groups df**	F	**p**
TEST	2	922	6.3408	0.0018*
FOLLOWING SOUND	10	21.12	1.757	0.1325

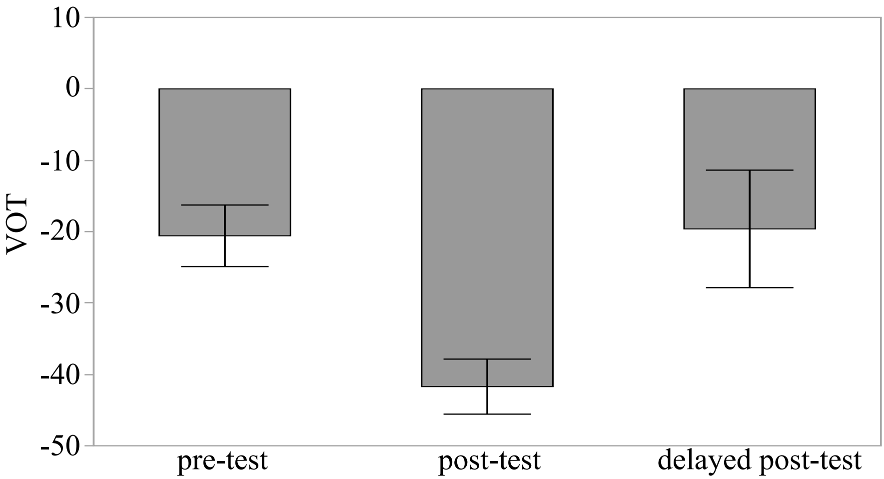


Figure 8.8: Mean VOT (ms) for trained voiced items of the Experiment Group for all three test conditions.

Trained Voiceless Items

The statistical results of the model for the *training subset* for trained voiceless items can be found in Table 8.17. Only the effect FOLLOWING SOUND reached significance, which can be seen in Figure 8.10. Participants of the Experiment Group failed to improve their production of voiceless stops after training. It can be concluded that the perception training was not successful with regard to voiceless stops. A possible explanation will be addressed in the discussion.

Regarding the *sustainability subset*, only FOLLOWING SOUND reach significance ($F(10, 20.67)=2.98$, $p<0.05$). Since participants did not improve in the post-test condition after the perceptual training, they were unlikely to improve after three months.

Table 8.16: Statistical results of the Generalized Linear Mixed Model for number of fully voiced stops of trained items for the Experiment Group for the *sustainability subset*.

source	**df**	χ^2	**p**
TEST	2	7.1354773	0.0282*
FOLLOWING SOUND	10	36.745238	<0.0001*

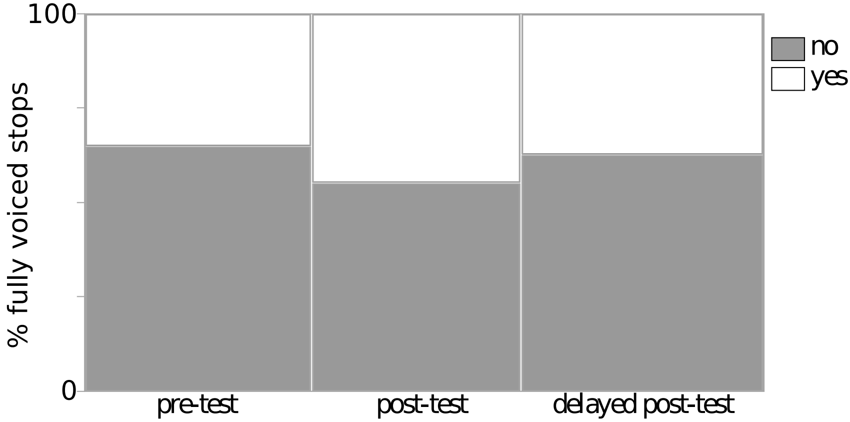


Figure 8.9: Percentage of fully voiced stops of trained items for Experiment Group for all three test conditions (*sustainability subset*). (yes = fully voiced)

Untrained Voiced Items

The results of the statistical model for the *generalization I* subset for untrained voiced stops can be found in Table 8.19. Only four effects reached significance: WORD POSITION ($F(2, 21.66)=5.31$, $p<0.05$), TEST ($F(2, 1433)=6.89$, $p<0.01$), PLACE OF ARTICULATION ($F(2, 20.4)=10.59$, $p<0.001$) and the interaction GROUP× WORD POSITION ($F(2, 1435)=10.71$, $p<0.0001$).

Regarding WORD POSITION, no difference was found between voiced initial and medial stops (-12 ms and -23 ms, respectively). However, final stops behaved differently and showed significant longer (positive) VOT values (22 ms) which is not surprising due to final lengthening processes and the fact that voiced obstruents in final position will be realized voiceless (final devoicing) in German. Post-hoc tests for the interaction GROUP×WORD POSITION showed, that participants of both groups behaved similarly with respect to

Table 8.17: Results of the statistical model for the production of trained voiceless stops for the *training subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	1	1183	0.0025	0.96
GROUP	1	18	0.3845	0.543
TEST×GROUP	1	1183	0.9346	0.3339
FOLLOWING SOUND	10	20.54	3.0672	0.0151*

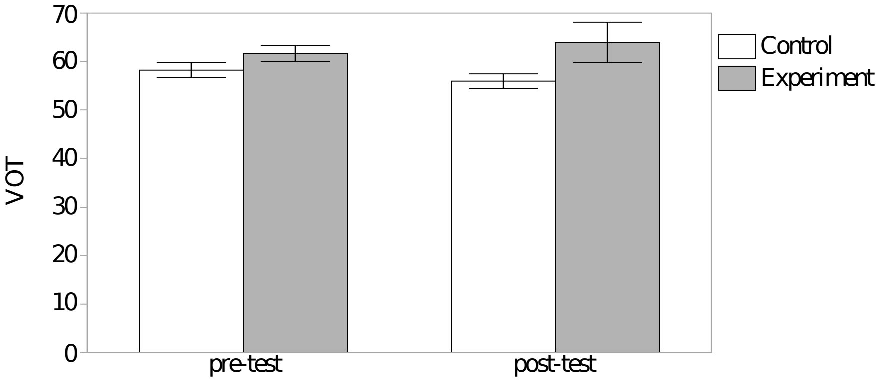


Figure 8.10: Mean VOT (ms) for trained voiceless items for the Control and Experiment Group (*training subset*).

medial stops but differed in their production of initial and final plosives. Learners of the Experiment Group produced initial stops with a positive mean VOT of 4 ms whereas participants of the Control Group produced initial plosives with -27 ms (see Figure 8.11). Furthermore, VOT for final stops produced by the Experiment Group showed much larger values (41 ms) in contrast to the Control Group (5 ms). Participants of the Control Group might have been more advanced from the beginning of the experiment than speakers of the Experiment Group. With respect to PLACE OF ARTICULATION, post-hoc tests showed that bilabial stops were produced with significantly longer negative VOT values (-17 ms) than alveolar (-6 ms) and velar stops (3 ms), the latter was even produced with positive values. With respect to TEST it was shown that the production for the post-test showed a significant reduction of VOT values (-11 ms) in comparison to the pre-test

Table 8.18: Results of the statistical model for the production of trained voiceless stops for the *sustainability subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	2	1183	0.2568	0.7735
FOLLOWING SOUND	10	20.67	2.9842	0.0171*

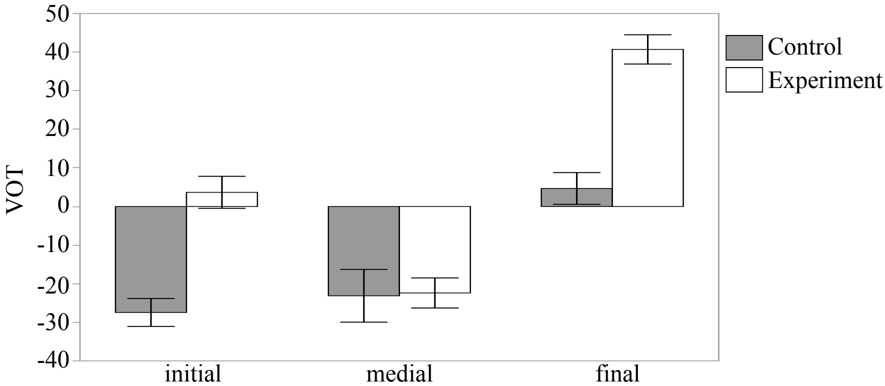


Figure 8.11: Mean VOT (ms) for untrained voiced items for the Control and Experiment Group for the different word positions (*generalization I subset*).

condition (0 ms, which does not mean here that there was no VOT). For this reason, some kind of generalization might have occurred. Unfortunately, no other interaction with the effect TEST was significant, so no clear statements can be made.

Taking a look at the statistical model for the *generalization II subset* (see Table 8.20) it can be seen that only the effect WORD POSITION reached significance ($F(2, 22.72)=8.41$, $p<0.01$). Post-hoc tests showed that all places of articulation differ significantly from each other. Medial stops are produced with a mean negative VOT of -18 ms, initial stops with a mean positive VOT of 5 ms, and final stops with a long positive VOT of 42 ms. The reason for the relatively long VOT values for final stops was discussed before and negative VOT values for medial stops are also not surprising as many stops were enclosed by vowels. Due to coarticulation processes, medial voiced stops tend to be realized fully voiced. No other effect was significant which further confirms that the perceptual training had no effect on the production of voiced

Table 8.19: Results of the statistical model for the production of untrained voiced stops for the *generalization I subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	1	1433	6.8922	0.0087*
GROUP	1	18.27	2.0448	0.1696
WORD POSITION	2	21.71	3.8286	0.0377*
POA	2	20.4	10.5873	0.0007*
TEST×GROUP	1	1434	0.0384	0.8447
TEST×WORD POSITION	2	1433	0.7458	0.4745
TEST×POA	2	1433	0.591	0.5539
TEST×GROUP×WORD POSITION	2	1434	0.6636	0.5151
TEST×GROUP×POA	2	1433	0.7723	0.4621
TEST×WORD POSITION×POA	4	1433	0.8771	0.4769
GROUP×WORD POSITION	2	1435	10.7078	<0.0001*
GROUP×POA	2	1435	1.5616	0.2102
GROUP×WORD POSITION×POA	4	1435	1.1375	0.3371
WORD POSITION×POA	4	21.7	1.9121	0.1464
FOLLOWING SOUND	11	19.75	1.5106	0.2045

untrained stops with different word positions and places of articulation (and also did not improve those items with the same criteria as the trained ones) and, hence, could not be generalized.

Untrained Voiceless Items

The statistical results for untrained voiceless stops of the *generalization I subset* can be found in Table 8.21. Only four effects reached significance. WORD POSITION showed a main effect ($F(2, 22.15)=3.66$, $p<0.05$) and post-hoc tests showed that initial and medial stops (67 ms and 67 ms, respectively) behaved differently than final stops (90 ms) which might have occurred due to final lengthening processes. However, post-hoc tests for the interaction GROUP×WORD POSITION showed that no difference was found between the different word position for the Experiment Group. With regard to Control Group, initial and medial stops behaved similarly in contrast to longer final stops. Furthermore, PLACE OF ARTICULATION was significant ($F(2, 22.2)=5.58$, $p<0.05$) revealing that bilabial stops were produced with a significant smaller VOT (66 ms) than alveolar and velar stops (76 ms and 77 ms, respectively) which might be an indication for generalizability to untrained voiceless bilabial items. However, there is evidence that stops

Table 8.20: Results of the statistical model for the production of untrained voiced stops for the *generalization II subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	1	1064	2.7637	0.0635
WORD POSITION	2	22.72	8.4129	0.0018*
POA	2	20.69	1.4016	0.2686
TEST×WORD POSITION	2	1064	0.6997	0.5922
TEST×POA	2	1063	1.6021	0.1715
WORD POSITION×POA	4	20.79	1.2918	0.3054
TEST×WORD POSITION×POA	4	1063	0.8013	0.6015
FOLLOWING SOUND	11	18.98	2.0475	0.0821

with different places of articulation show different duration behavior (Lisker and Abramson, 1964; Klatt, 1975). Concerning the interaction TEST×GROUP ($F(1, 1573)=5.49$, $p<0.05$) it was noticed that the pre-test condition showed significant lower VOT values than the post-test condition. No other contrast was significant and, therefore, no generalization (across word position and poa) was found for the Experiment Group.

Regarding the *generalization II subset* only two effects reached significance (see Table 8.22). FOLLOWING SOUND showed a main effect ($F(11, 22.74)=5.29$, $p<0.001$) as well as PLACE OF ARTICULATION ($F(12, 22.8)= 6.9238$, $p<0.01$). Post-hoc tests showed that bilabial stops (67 ms) behaved differently from alveolar (77 ms) and velar stops (78 ms), which was already shown for the *generalization I subset*. However, no other effect reached significance and no improvement could be shown for untrained voiceless stops.

8.4 Discussion

The analysis of the production and perception of voiced and voiceless stops showed two different aspects of the behavior of German learners of French. Subjects of the Control and Experiment Group performed relatively well in the identification of stops. Since both French and German have a two-way distinction of stops, developing a strategy to distinguish between voiced and voiceless stops, although phonetically marked differently, seems to be rather straightforward (see Best, 1994; Best and Hallé, 2010; Escudero, 2005; van Leussen and Escudero, 2015). It was shown that trained voiced stops tend to be identified better than voiceless stops. Since participants already achieved

Table 8.21: Results of the statistical model for the production of untrained voiceless stops for the *generalization I subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	1	1573	0.9539	0.3289
GROUP	1	18.53	0.0053	0.943
WORD POSITION	2	22.15	3.6614	0.0423*
POA	2	22.2	5.5791	0.0109*
TEST×GROUP	1	1573	5.4854	0.0193*
TEST×WORD POSITION	2	1573	2.1633	0.1153
TEST×POA	2	1573	1.3335	0.2638
TEST×GROUP×WORD POSITION	2	1573	1.6803	0.1867
TEST×GROUP×POA	2	1573	0.792	0.4531
TEST×WORD POSITION×POA	4	1573	0.7113	0.5842
GROUP×WORD POSITION	2	1573	8.5827	0.0002*
GROUP×POA	2	1573	0.7391	0.4777
GROUP×WORD POSITION×POA	4	1573	0.8704	0.4809
WORD POSITION×POA	4	22.22	1.4583	0.2484
FOLLOWING SOUND	11	22.12	2.0858	0.0682

a correctness score of above 95% for voiced stops, it is difficult to improve even further. Training seems to have helped participants of the Experiment Group to improve the identification of items with voiceless stops by 11%. This improvement was also sustained even three months after training. Concerning untrained stops, items with final stops were identified overall worse than items with initial and medial stops. These results can be interpreted in two ways: a) with respect to the final devoicing processes in German which defines that voiced obstruents will be produced voiceless at the end of a word or syllable. For voiced stops, German learners of French might still expect a final stop to be voiceless, according to the final devoicing, independent of the fact that the stop was actually produced voiced, b) concerning the problematic identification of final voiceless stops, participants might try to ignore final devoicing since this phonological process does not exist in French. Due to the fact that voiceless French stops roughly coincide with German voiced stops, participants might still be troubled by the phonetic differences and therefore wrongly identify items with final voiceless stops. However, especially in regard to voiceless untrained stops, it was shown that participants of the Experiment Group managed to improve in their second generalization test three months after training. This improvement was mainly contributed to the improvement of final stops (from 67% to 88%). How and why this

Table 8.22: Results of the statistical model for the production of untrained voiceless stops for the *generalization II subset*.

source	**between-groups df**	**within-groups df**	F	**p**
TEST	1	1179	0.3017	0.6396
WORD POSITION	2	22.24	1.092	0.3529
POA	2	22.8	6.9238	0.0045*
TEST×WORD POSITION	2	1179	0.2466	0.9118
TEST×POA	2	1179	0.8966	0.4652
WORD POSITION×POA	4	22.88	2.5202	0.0691
TEST×WORD POSITION×POA	4	1179	1.0682	0.3829
FOLLOWING SOUND	11	22.74	5.2943	0.0004*

particular category improved is, however, unclear. A general improvement was also shown for voiced untrained stops but due to a lack of significance of the interaction WORD POSITION×POA no detailed information can be given at this point. Unclear is also why alveolar voiced stops behave significantly better than bilabial and velar stops and why bilabial voiceless stops behave significantly worse than alveolar and velar stops. Especially the second result is interesting as a more detailed look at the interaction WORD POSITION×POA showed that items with initial bilabial stops were identified worse, which coincides with the exact specification of the training condition. It seems as if training did seem to have helped for the words that were included in the training but was counter-productive for untrained items at the same time.

Regarding the production, it was found that training helped to improve the production of voiced stops by producing more fully voiced stops. Although the interaction TEST×GROUP did not reach significance for the *training subset*, the *sustainability subset* showed a main effect for TEST. The improvement is displayed in Figure 8.8 and results can be explained by the increased number of produced fully voiced stops (and therefore negative VOT). This may be an indication for developing an awareness of the correct production of voiced stops in French. However, participants of the Experiment Group were not able to sustain this improvement after three months of training. Furthermore, no improvement and generalization was found with regard to the production of voiceless stops.

In general, it is concluded that high variability training seemed to have a beneficial effect on the production and perception of French voiced and voiceless stops by German native speakers. However, training seemed to be only beneficial for the perception of voiceless stops (since the perception of voiced stops is already at ceiling), but concerning the production, it only seems helpful for voiced stops. It is interesting to see that participants were able to sustain their improvement after three months of training for perception, but fell back into old habits with regard to the production. This is confirmation that if a learner wants to succeed and improve with the aim of a correct or more native-like production, the learner must keep practicing. Since the difference between the production of voiced and voiceless stops is marginal with respect to the small durational changes in VOT and the presence/absence of vocal fold vibration during the closure, a training over a few weeks only is not sufficient for a sustainable improvement with respect to these fine details. It has also to be kept in mind that training here only refers to perceptual training and that no articulatory training was carried out.

The results reported here may be seen as a challenge for accounts claiming a close link between production and perception, because the behavioral patterns of the participants were inverted – rather than parallel, as one might expect - for the production and perception tasks. It is not argued that this close link does not exist since perceptual training clearly helped to improve pronunciation. But it is unclear how, in particular, this link is defined. However, this question would go beyond the scope of this dissertation.

In terms of a possible explanation for the encountered perceptual results, it could be argued that participants might have been able to use the perceptual cue of vocal fold vibration during the closure duration in order to decide whether they heard a word starting with a voiced or voiceless stop. Although German native speakers do not pay attention to this feature when speaking German, they might still be able to perceive it (especially at the beginning of a word in isolation) and use it to distinguish between the two stop categories. However, if the vocal fold vibration is absent, learners might still get confused by the different phonological classifications of a voiceless French stop and a voiced German stop with similar phonetic realizations. However, when receiving feedback about the correctness of the identification, learners are able to shift their categorical boundaries in order to identify especially French voiceless stops more often correctly after the training. Apparently, no additional training is necessary to maintain this shift in categorical perception since participants of the Experiment Group were able to identify voiceless French stops equally well three months after training. Note that these results only apply to the identification of isolated words.

According to the Speech Learning Model (SLM, Flege, 1995) the difficulty of acquiring non-native sounds can be predicted by the perceptual similarity between L1 and L2 sounds. Therefore, non-native sounds that are identical to the sounds of the native language are acquired easily, non-native sounds that are different but similar to sounds of the native language (i.e. do not fall into already existing phonological categories but create their own categories) are also relatively easy to learn, but non-native sounds that are similar but not identical to sounds of the native language (i.e. overlap with an existing category but do not share the same thresholds) are very hard to acquire. Since both German and French distinguish between voiced and voiceless stops, a mapping of two categories onto two categories which are different from a phonetic point of view and overlap to a certain point, the acquisition of the correct realization of the L2 sounds is rather difficult for learners to achieve. Especially since German native speakers do not pay attention to the phonological feature of voicing which is crucial for the correct pronunciation of French voiced stops. It was already shown that learners were able to perceive this feature. However, an active articulation is challenging for many learners. This is especially true since there are occurrences of pre-voicing in the data, but participants were often not able to maintain the vocal fold vibration throughout the complete closure. For one speaker in particular a certain strategy seemed to have helped him to articulate a voiced closure: producing the voiced bilabial nasal [m] before producing voiced bilabial stop sounds. However, he also often produced an additional positive VOT.

With regard to the results of the production data it can be argued that participants developed the strategy to change one category first in order to avoid a clash of new (L2) and old (L1) categories. It should be easier for the learner to reduce VOT values of voiceless stops since German voiced stops are already characterized as having a short VOT (see results for experiment II). However, it is more difficult for the learner to produce a fully voiced closure without any VOT for voiced stops. If the learner is not able to make this transition or falls back into native habits after discontinuing the training, both phonological categories would clash into one phonetic category: a phonetically voiceless stop with a short VOT. For this reason, it seems plausible to first change and improve the more challenging category including the fully voiced closure. Interestingly, Flege (1987) found evidence that L2 learning also influenced the production of /t/ in the native language. According to Flege, French learners of English showed longer, more English-like VOT values than French speakers who did not speak another language. Similarly, English learners of French showed smaller, more French-like VOT values. It would have been interesting to see if the French learning participants behaved differently when speaking German than monolingual German speakers. However, this was beyond the scope of this dissertation.

Chapter 9

Summary and Conclusion

In this thesis the effect of different training procedures on the pronunciation of French stops by German native speakers was investigated. In general, improving the pronunciation of a second language is nothing that can be acquired over night and is a complex process. In order to become successful, you need to be disciplined and persistent. When learning a foreign language most learners usually retain a foreign accent which results from interferences of the phonological and phonetic system of the native and non-native language (e.g., Best, 1994; Flege, 1995; Kuhl and Iverson, 1995; Escudero, 2005). Four of the most important models of second language acquisition with regard to the acquisition of perception and production of non-native sounds were shortly discussed and compared in this thesis (Perceptual Assimilation Model, (Best, 1994), Speech Learning Model, (Flege, 1995), Second Language Linguistic Perception Model, (Escudero, 2005; van Leussen and Escudero, 2015), Native Language Magnet Model, (Kuhl and Iverson, 1995; Kuhl et al., 2008)). These models try to explain why most speakers of a second language retain a foreign accent and why it is difficult to acquire a more native-like perception and production.

The main challenge for the learner is to become aware of own mistakes and to perceive the deviations in the native pronunciation of the respective L2 in comparison to own realizations (Barry, 2007). A correct pronunciation is essential for speaking a foreign language, since a poor pronunciation will make it harder to be understood by other interlocutors (e.g., Mennen et al., 2007). Unfortunately, pronunciation training does not receive much attention in second language classrooms and most teachers are often not properly trained in phonetics in order to create helpful exercises and give valuable feedback (Baker, 2011).

Butler and Winne (1995) described in their model of self-regulated learning how important feedback is for the learning process. The model illustrates the mental processes that are undergone to cope with a given task and how feedback can help to improve the learner's performance and outcome. In a study by Lyster and Ranta (1997) six different types of feedback were identified in a second language classroom environment: 1) explicit correction, 2) recast, 3) clarification request, 4) metalinguistic feedback, 5) elicitation, and 6) repetition. They showed that recast, which is defined as corrected reformulations by the teacher, was the method that was used most frequently. However, recasts do not encourage the learner to react to the feedback. This way, it cannot be said with certainty that the learner actually realized that the production contained a mistake which was reformulated. In comparison, elicitation indicates implicitly that a mistake was made and encouraged an uptake by learners in 43% of the cases in the study by Lyster and Ranta.

When lessons do not include extensive pronunciation training, computer-assisted pronunciation training (CAPT) systems can be useful for the learner to improve their pronunciation of a second language. The advantage of CAPT systems is that learners are able to train as long and as much as they want on specific pronunciation problems. And students that are self-conscious in classroom interactions are able to actively engage with the system without feeling judged by classmates or the teacher. However, an evaluation of 20 commercial and non-commercial CAPT systems revealed that many systems still include feedback methods that are difficult to interpret. For example, in order to interpret a display of an oscillogram or even a spectrogram, the learner needs to have some kind of knowledge in acoustic phonetics. Furthermore, it gives the impression that the wave form of the native speaker and of the learner need to look the same in order to be correct. However, it most likely will not even look the same when the reference speaker produces the same utterance twice. And the learner's utterance can still be perfectly fine, even though the oscillogram looks different. It is of importance that the implemented feedback gives all necessary information on what was wrong (e.g., highlighting the specific sound in a word or longer phrase) and on how to reduce or improve the mistake. Without this knowledge, given feedback cannot be interpreted accordingly and is most likely useless (Kulhavy, 1977; Hattie and Timperley, 2007).

One example of implicit feedback that aims at helping learners of German to improve their vowel productions in a playful manner, was presented with the vowel training software 'Hüpf'. The name of the software is based on the graphical user interface of a frog that jumps to specific locations of water lilies in a pond. The training software was intendedly created as a game. This way, both adults and children would hopefully like to use the software as it is both helpful and entertaining. Simultaneously, feedback can be presented in an intuitive way so users without any background in phonetics can use the system equally well as users with a background in phonetics. Each water lily represents mean values of F1/F2 coordinates which were fed into the system. If the frog lands on the correct water lily, the production of the vowel was correct. If, however, the frog lands on a different water lily or in the water, the production was incorrect. Regarding vowel duration, two bars are displayed around each water lily. An evaluation of usability and effectiveness is still to be done.

A short description of the IFCASL (Individualized Feedback in Computer-Assisted Second Language Learning) French-German learner corpus gave information about the most common problems German learners of French and French learners of German face when learning French and German, respectively. About 50 German and 50 French native speakers were recorded both in their native as well as in their non-native language which allows for cross-language comparisons. In order to see which pronunciation problems disappear with an increasing language level and which problems are still existent with a high language level, both beginners and advanced speakers were recorded. To generate corpus material which was as variable as possible for read speech, different recording conditions were included (read and repeated short sentences, focus conditions, story “The three little pigs”). Subsequently, the complete corpus was forced-aligned (Jouvet et al., 2011; Fohr and Mella, 2012) and a large part of the corpus was manually checked and corrected by student assistants.

Based on the recordings of “The three little pigs” the first experiment was conducted, and investigated the effect of prosody manipulation on perceived foreign accentedness. In general, evidence suggests that L1 speech is produced faster than L2 speech (e.g., Munro and Derwing, 1995b; Guion et al., 2000 regarding mean utterance duration and Raupach, 1980; Gut, 2009; Baese-Berk and Morrill, 2015; Trouvain and Möbius, 2014b regarding speaking/articulation rate). Concerning pitch in L2, a number of studies showed that learners have difficulties concerning (global) long term distributional pitch profiles reflected by differences in pitch range and the correct alignment of pitch accents (e.g., Mennen, 1998; Ullakonoja, 2007; Hincks and Edlund, 2009; Busà and Urbani, 2011; Busà and Stella, 2012; Zimmerer et al., 2014). Overall, the role of prosody in foreign accent perception has rarely been studied. Boula de Mareüil and Vieru-Dimulescu (2006) applied a prosody transplantation paradigm on Spanish and Italian native utterances which transferred phoneme duration and pitch contours of one language to the other. They found evidence that listeners were equally influenced by segmental and suprasegmental features for natural speech. However, prosody transplantation was only applied on recordings of native speech. Other experiments investigated the effect of transplantation of native prosody on foreign accented productions (e.g., Winters and Grantham O'Brien, 2013; Rognoni and Busà, 2013; Ulbrich and Mennen, 2015). They found that manipulated utterances were rated as less accentuated than original recordings by native listeners.

Recordings of "The three little pigs" by French learners of German were manipulated using the proposed technique by Boula de Mareüil and Vieru-Dimulescu (2006). Based on one German male and female reference speaker, syllable duration and pitch contour were transplanted on the French accented utterances. An acoustic analysis of the articulation rate showed, that French speakers spoke French with a higher rate than German native speakers spoke German. However, when French speakers spoke German they were only half as fast as in their L1. With regard to pitch range it was shown that French female speakers have a narrower pitch range in their native language than German native speakers in their L1. However, when French speakers spoke German, they actually adjusted their pitch range to the native German speakers. No difference was found for male speakers.

In a perception experiment German native speakers were instructed to listen to each utterance and decide how accented each recording was (1 = not accented, 7 = heavily accented) and whether the recording sounded natural or artificial. Results revealed that manipulated sentences received lower accentedness scores than the original accented recordings. However, the manipulated utterances were often rated as sounding artificial. And although foreign accent rating decreased significantly, the rating for manipulated utterances is still considerably high. This might be an impact of non-native segmental cues which were not manipulated or even the manipulation procedure itself.

It was shown that manipulation can help to reduce the perceived accent. The following experiment investigated the effect of exposure to manipulated speech and to recordings of a native speaker on the pronunciation of French voiced and voiceless stops by German learners of French. Since French and German share the same set of voiced and voiceless stops /b p d t g k/ but differ in terms of the phonetic realization of the sounds, it is particularly problematic for German learners of French to produce the stops correctly (Best, 1994; Flege, 1995; Kuhl and Iverson, 1995; Escudero, 2005). French speakers differentiate between fully voiced plosives and voiceless unaspirated ones with a short Voice Onset Time (VOT). In contrast, German shows a distinction by voiceless unaspirated plosives with a short VOT and voiceless aspirated ones with a long VOT (e.g., Beyer, 1908; Künzel, 1977; Hammarström, 1998; Pustka, 2011). While participants of one group trained with their own manipulated voice, subjects of the other group trained with recordings of a native speaker. Manipulation was carried out for VOT only. Additionally, a control group was also tested but did not receive any feedback. On the first day, participants recorded short French sentences for all six stops. Since manipulation was carried out manually, the training and post-test had to be done on the second day. The training consisted of a short imitation phase for each stop in which participants had to imitate either their own manipulated voice or a native speaker. Subsequently, a transfer phase followed in which subjects could not to listen to their voice or the native speaker before speaking.

Although it was shown that training with the own voice helped to improve the pronunciation of voiceless stops, training with the native speaker was more beneficial and outperformed the manipulation condition. It has to be considered that many people do not like to hear their own voice which might be one reason why the manipulation condition did not yield as good results as the native speaker condition. A subsequent perception experiment with French native listeners did not confirm the improvement by participants of the Native Speaker Group. French native listeners had problems to identify the produced words correctly in an identification experiment in which they had to state whether they heard the word with the voiced or voiceless stop. This result can be explained by the fact that when VOT was reduced for voiceless stops, they might have coincided with voiced stops which did not improve and were still produced with a short VOT as well. It is interesting to see that participants only improved for voiceless stops. It might be argued that manipulation of VOT is not a sufficient method to have an effect on voiced stops since it does not focus on the phonation state of the closure. However, recordings of the French native speaker included information about the voiced closure but participants of the Native Speaker Group still did not improve for voiced stops either. It is unclear whether learners were not able to perceive voicing or whether they were not able to implement this feature.

For this reason, another experiment was conducted as a pure perceptual training with recordings of six French native speakers. This type of perception training is called High Variability Training (HVT) because it includes stimuli by several native speakers and was found to be effective in a number of studies for both perception and production (e.g., Logan et al., 1991; Wang et al., 1999; Sadakataa and McQueen, 2013). In a pre-test participants were tested for their production and perception skills with regard to French voiced and voiceless stops. The perception test was carried out by means of an identification test in which they heard single words and had to decide whether they heard the word with the voiced or voiceless stop. A training phase of three weeks followed in which participants carried out additional identification tests for words with initial, bilabial stops and they received simple 'correct'/'incorrect' feedback. A post-test was conducted shortly after the training for perception and production skills including an additional set of generalization tests to see whether any improvements for initial, bilabial stops could be transferred to different places of articulation (alveolar, velar) and word positions (medial, final). The same tests were also carried out three months after training to see whether any improvements were maintained. An additional control group was tested but participants did not receive any training.

It was shown that participants were quite good from the beginning in identifying French stops. However, the identification was better for voiced than for voiceless stops, but participants of the experiment group were able to improve their identification of voiceless stops. This improvement was maintained even three months after training. With regard to the production of French stops, only voiced stops improved. However, this improvement was not maintained after three months. Furthermore, no generalizations to other places of articulation and word positions were found.

But why did the two tested methods result in different outcomes? Regarding the imitation experiment it can be argued that since the manipulation of the learner's voice only focused on VOT, participants of this group did not receive all relevant information necessary to produce a French fully voiced stop. Reducing the VOT of German accented voiceless stops seems easier to be realized since German voiced stops are already characterized by a short VOT. This means that learners are already familiar with the articulatory gestures to produce unaspirated stops. However, reducing VOT even more does not seem to be straightforward. Although participants who trained with recordings of the native speaker did have access to all acoustic information of voiced stops, they did not show any improvement for voiced stops either. It was argued before that it is unclear whether participant were not able to hear the voiced closure (regarding the production of stops, voicing is not distinctive in German) or whether they were not able to implement the voicing feature in their speech. After carrying out the HVT experiment, this question can be answered. Participants were able to hear fully voiced closures but it seems easier to be perceived in single words. As the imitation method included words embedded in sentences, participants who trained with the native speaker might have actually had troubles focusing on or perceiving the voiced closure. However, since participants of the HVT focused on single words, they were able to develop a strategy to change their pronunciation of voiced stops accordingly.

However, it is interesting why the production of voiceless stops did not improve for HVT but voiced stops did. It is argued that participants developed a strategy to change one category first in order to avoid a clash of new (L2) and old (L1) categories. It should be easier for the learner to reduce VOT values of voiceless stops since German voiced stops are already characterized as having a short VOT (see results for experiment II). However, it is more difficult for the learner to produce a fully voiced closure without any VOT for voiced stops. If the learner is not able to make this transition or falls back into native habits after discontinuing the training, both phonological categories would clash into one phonetic category: a phonetically voiceless stop with a short VOT. For this reason, it seems plausible to first change and improve the more challenging category including the fully voiced closure. And as it was shown for HVT, participants were tested three months after training but failed to maintain the improvement for French voiced stops which would have indeed resulted in the clash of categories. In general it can be said that intensive training seems to help and shows improvements directly after the training. However, if the learner does not proceed with the training, improvements will not be maintained and will disappear.

To summarize, the discussed results indicate that not all training methods are equally well suited for certain pronunciation problems. For this reason, it is important to test these methods beforehand since training procedures can also be counter-productive, for example if they confuse the learner. This way, pronunciation problems might be reinforced. Unfortunately, developing well-conceived training methods is time-consuming and elaborate and should always focus on a specific problem at a time. This does not only ensure that the used method does not interfere with other problems but gives the learner time to focus on one problem at a time. Another feature that is often ignored in the development of commercial CAPT systems is the native language of the learner. Not all learners show the same pronunciation problems in a given foreign language. For example, English native speaker have problems to produce the rounded, closed, front vowel [y] when learning French (Levy and Law, 2010). However, German native speakers would probably not show this problem since the sound already exists in German. This is something that especially commercial CAPT systems need to consider when providing training procedures that are computer-assisted.

This thesis contributed a small part to the various training possibilities that can be provided to help learners become aware of their deviations and to help them to improve. Learners need to be aware of the fact that second language learning and pronunciation training is something that cannot be learned over night if they really want to be successful. Pronunciation training needs endurance by the learner but also thoughtful developed exercises and techniques with pedagogical use.

Appendix A

Recording Material for the IFCASL Corpus

(1) Elle habite dans un beau village en France.
(2) Le champagne est rangé dans la cave en terre.
(3) Les avions sont rentrés à la base apr ès le vol.
(4) Maman a perdu sa bague en argent.
(5) Les enfants sont partis en balade en forêt.
(6) Le panier est rempli de crabes abîmés.
(7) Le bateau à vapeur a quitté le port de Nice.
(8) Le bateau à vapeur a viré de bord dimanche.
(9) Le bateau à vapeur s'appelle Igor.
(10) La police a découvert le gosse dans le parc.
(11) La police a découvert le corps à terre.
(12) Le silence se fait quand elle dort l'après-midi.
(13) Les accusés ont toujours tort dans ce cas.
(14) Le train a quitté la gare de Paris.
(15) Le garçon a pris le car à Berlin.
(16) Le parrain a quitté le bar ce matin.
(17) Le garçon mange deux parts de gâteau.
(18) L'abeille n'a pas de dard sur le corps.
(19) Les poids ont fait la tare sur la balance.
(20) Elle a mis une robe rose faite dans un tissu doux.
(21) Ils ont mangé une pomme.
(22) Ils sont rentrés tôt.

(23) La voiture est belle.
(24) Le prince charmant a un bel habit.
(25) Le prince charmant a un bel ami.
(26) Il boit une bière au bar.
(27) La BMW bleue est immatriculée AM 525 YY.
(28) Barack Obama est le 44ème président des USA.
(29) En 2013, il y aura 26 millions d'habitants dans l'UE.

Table A.1: French sentences of the *Sentence Heard* (SH) recording condition in the IFCASL corpus.

(1) Mon ami a perdu ses bagages à la gare.
(2) Les enfants sont braves à l'école.
(3) Les élèves doivent cocher la bonne case avec un feutre.
(4) Son ami comprend la blague en anglais.
(5) Vous aimez la salade aux anchois.
(6) Les idiots chassent les crabes en Australie.
(7) La piscine est couverte d'une bâche en plastique.
(8) Maman a acheté une table basse orange.
(9) Marc a pris une baffe àcause de Marie.
(10) Les enfants ont peur du loup.
(11) La voiture s'est arrêtée au feu rouge.
(12) Il y a deux ans, vous y êtes arrivés.
(13) Il les en a empêchés.
(14) J'ai acheté mon armoire à Rome.
(15) Mon enfant est aimable.
(16) Les chiffres indiquent neuf heures.
(17) Prenez-en soin.
(18) Aimez-vous mon ami?
(19) Est-ce que vous aimez mon ami?
(20) Vous aimez mon ami?
(21) Où allons-nous en mai?
(22) La méchante sorcière est le personnage principal du spectacle de ce soir.
(23) Nous devrions porter quelques fleurs à la servante au grand cœur dont vous étiez jalouse.
(24) Marc fait de la poutre en cours de sport.
(25) En cette saison, Paul commande souvent du gratin au restaurant.
(26) Mon cousin a beaucoup de chance, il habite dans un grand pavillon.
(27) Cent moins deux est égal à quatre-vingt-dixhuit.
(28) Les 5 anneaux sur le drapeau du CIO sont le symbole des 5 continents.
(29) Mitterrand chassait le canard avec un fusil à plomb.
(30) Je n'ai pas le temps.
(31) Je ne sais pas s'il y a des vacances.

Table A.2: French sentences of the *Sentence Read* (SR) recording condition in the IFCASL corpus.

Marc am ène un ami?	Yvonne am ène un ami.
Yvonne am ène ma m ère?	Yvonne am ène un ami .
Yvonne dépose un ami?	Yvonne am ène un ami.
Que fait Yvonne?	Yvonne am ène un ami.
Qui met un noeud?	Marie met un noeud.
Que met Marie?	Marie met un noeud.
Que fait Marie?	Marie met un noeud.

Table A.3: French sentences of the *Focus* (FC) recording condition in the IFCASL corpus.

Les trois petits cochons s'en vont de chez eux pour construire leurs maisons. Le premier petit cochon construit une maison en paille, le deuxième construit une maison en bois et le troisième construit une maison en brique. Le loup aperçoit les petits cochons et décide de manger celui dans la maison en paille en premier.Il frappe à la porte mais le petit cochon ne le laisse pas entrer. Le loup gonfle alors ses joues, souffle de toutes ses forces et la maison s'envole. Le petit cochon court alors chez son frère celui dans la maison en bois. Le loup frappe à la porte mais les petits cochons ne le laissent pas entrer. Le loup gonfle alors ses joues, souffle de toutes ses forces et la maison s'envole. Les deux petits cochons courent alors chez leur frère celui à la maison en brique. Le loup frappe à la porte mais les petits cochons ne le laissent pas entrer. Le loup gonfle alors ses joues, souffle de toutes ses forces mais la maison ne s'envole pas.Le loup décide alors de passer par la cheminée mais les petits cochons ont préparé un chaudron d'eau bouillante. Le loup tombe dedans, pousse un hurlement et s'enfuit en courant.

Table A.4: French story of the *Conté* (CT) recording condition in the IFCASL corpus.

(1) Ein kleines Boot nennt man Kahn.
(2) An Karneval tragen die Kinder auf der Gasse lustige Hüte.
(3) Wir leiten einen Ausflug nach St. Tropez.
(4) Der Hase lebt auf dem Feld, das Kaninchen in einer Höhle.
(5) Die Flagge des IOC besteht aus 5 Ringen als Symbol der 5 Kontinente.
(6) Bei der nächsten Wahl hoffen SPD und Grüne auf eine Revanche gegen CDU, CSU und FDP.
(7) Schnecken haben ihr Haus immer dabei.
(8) Du machst ständig Quatsch.
(9) Obama ist der 44. Präsident der USA.
(10) Die Oase liegt in der Wüste.
(11) Bring bitte noch Teller in den Garten.
(12) Ein paar Journalisten möchten mit dir sprechen.
(13) In der Garage steht ein gelbliches Auto.
(14) In einer Bar hört man viel Jugendsprache.
(15) Mitte des Jahres 2013 waren in der EU 26 Millionen Menschen arbeitslos.
(16) Im Gegensatz zum Schreibblock ist der Schreibtisch 80 cm lang.
(17) Mein Cousin isst häufig Bratwürstchen oder Hähnchen.

(18) „Zeitgeist“ ist ein Wort das in vielen Sprachen existiert.
(19) Für manche sind Haare von Katzen die Hölle.
(20) Auch Täler findet man in Karten.
(21) Wir treffen uns um 4 Uhr in Essen.
(22) Alte Häuser sind charmant.
(23) Wer segelt mit deinem Schiff?
(24) Hängt das Bild denn schief?
(25) Braucht man zum Reisen einen Pass?
(26) Braucht man in einer Band einen Bass?
(27) An Weihnachten gibt es viel Gebäck.
(28) Du reist mit viel Gepäck.
(29) Du reißt es mit großem Eifer aus dem Boden.

Table A.5: German sentences of the *Sentence Heard* (SH) recording condition in the IFCASL corpus.

(1) Ich wüsste nicht, wie der schnellste Weg nach Polen ist.
(2) Pflanzen können Licht zu Energie verarbeiten.
(3) Computer sind schnell veraltet.
(4) In Berlin zahlt man wenig Miete.
(5) In jeder Saison sitzen wir im Pavillon.
(6) Die Sonne versinkt am Abend.
(7) Wir essen Gemüse aus unserem Beet.
(8) Die riesigen Risse am Haus gehen nicht von alleine weg.
(9) Zum Geburtstag kaufe ich dir Pflaumen.
(10) 96 sind 4 weniger als 100.
(11) Für manche sind Internet und Email eine Sucht.
(12) Nicht jeder kann Goethe leiden.
(13) Hast Du Schmerzen am Knie oder am Ohr?
(14) Auf der Erde ist der Blauwal das größte Tier.
(15) Die Tschechen halten den Weltrekord im Biertrinken.
(16) In jeder Bank gibt es eine Kasse.
(17) Der Schwarzwald ist gar nicht schwarz.
(18) „La Boum“ ist ein bekannter französischer Film.
(19) Im Restaurant bestellt Paul oft Gratin.
(20) Frankfurt liegt in Hessen.
(21) Das Kfz-Kennzeichen des blauen BMW lautet HA-BT 521.
(22) Eine Metro-Station in Paris heißt „Oberkampf“.
(23) Im Frühling fliegen Pollen durch die Luft.
(24) Wieviele Kinder fahren mit dem Rad?
(25) Braucht sie deinen Rat?
(26) Die Wildpferde in der Camargue sind schön.
(27) Ich hasse deinen Regenschirm.
(28) Du mietest mit ihr eine Hütte.
(29) Bei hohem Fieber legst Du dich ins Bett.
(30) Die Chance im Lotto zu gewinnen ist sehr klein.
(31) Der Mörder hat den Tatort wieder besucht.

Table A.6: German sentences of the *Sentence Read* (SR) recording condition in the IFCASL corpus.

Peter vertraut dem Kamel?	Yvonne vertraut dem Kamel.
Yvonne vertraut dem Hund?	Yvonne vertraut dem Kamel .
Yvonne glaubt dem Kamel?	Yvonne vertraut dem Kamel.
Was macht Yvonne?	Yvonne vertraut dem Kamel.
Wer schwimmt im See?	Marie schwimmt im See.
Wo schwimmt Marie?	Marie schwimmt im See .
Was macht Marie?	Marie schwimmt im See.

Table A.7: German sentences of the *Focus* (FC) recording condition in the IFCASL corpus.

Die drei kleinen Schweinchen gehen heim, um ihre Häuser zu bauen. Das erste kleine Schweinchen baut ein Haus aus Stroh, das zweite baut ein Haus aus Holz, und das dritte baut ein Haus aus Backsteinen. Der Wolf entdeckt die drei kleinen Schweinchen und beschließt, das kleine Schweinchen mit dem Haus aus Stroh als erstes zu fressen. Er klopft an der Türe, aber das kleine Schweinchen lässt ihn nicht ins Haus. Der Wolf bläht daraufhin seine Backen auf und bläst mit all seiner Kraft, woraufhin das Haus weg fliegt. Das kleine Schweinchen rennt zu seinem Bruder mit dem Haus aus Holz. Der Wolf folgt dem Schweinchen und klopft wieder an die Türe, aber die kleinen Schweinchen lassen ihn nicht hinein. Der Wolf bläst daraufhin seine Backen auf, pustet mit all seiner Kraft und das Haus fliegt davon. Die zwei kleinen Schweinchen rennen zu ihrem Bruder mit dem Haus aus Backsteinen. Der Wolf klopft an die Türe, aber die kleinen Schweinchen lassen ihn nicht hinein. Der Wolf bläst daraufhin seine Backen auf, pustet mit all seiner Kraft aber das Haus fliegt nicht davon. Da beschließt der Wolf, durch den Schornstein in das Haus zu steigen, aber die kleinen Schweinchen haben einen Kessel mit kochendem Wasser vorbereitet. Der Wolf fällt hinein, stößt einen Schrei aus und rennt davon.

Table A.8: German story of the *Conté* (CT) recording condition in the IFCASL corpus.

Appendix B

Recording Material for Experiment II

Der schwere **Ballast** ist zu viel für mich.	The heavy **ballast** is to much for me.
Der große **Palast** des Prinzen ist umwerfend.	The **prince**'s large palace is stunning.
Ich spiele **Bass** und Klavier.	I play **bass** and piano.
Der gute **Pass** an Peter kam von Thomas.	The good **pass** to Peter came from Thomas.
Morgen Abend gehen wir in eine **Bar** in Frankfurt.	We will go to a **bar** in Frankfurt tomorrow evening.
Das glückliche **Paar** küsste sich bei Mondlicht.	The happy **couple** kissed each other in the moonlight.
Das frische **Gebäck** meiner Oma ist lecker.	My grandma's fresh **pastry** is delicious.
Das große **Gepäck** von Lisa ist schwer.	The large **luggage** from Lisa is heavy.
Das Baby hat süße **Bäckchen** und schwarze Haare.	The baby has sweet **cheeks** and black hair.
Das kleine **Päckchen** ist gestern angekommen.	The small **package** came yesterday.
Die Flasche **Bier** musst du bezahlen.	You have to pay the **beer** bottle.
Der hölzerne **Pier** bietet viel Platz für Schiffe.	The wooden **pier** provides plenty of room for ships.
Das linke **Bein** von Anton ist gebrochen.	Anton's left **leg** is broken.
Sie litt furchtbare **Pein** bei dieser Vorstellung.	She was in complete **agony** at the very thought.
Anne hat ihre **Daten** gelöscht.	Anne has deleted her **data**.
Diese **Taten** sind nicht zu entschuldigen.	These **actions** are inexcusable.
Das kleine **Dorf** ist sehr ruhig.	The small **village** is very quiet.

Der trockene **Torf** brennt gut.	The dry **turf** burns well.
Der große **Dank** geht an Thomas.	The biggest **thanks** goes to Thomas.
Der kleine **Tank** des Motorrads ist schon wieder leer.	The motorbike's small **tank** is empty again.
Die sogenannte **Dur**-Tonleiter findet man in der Musik.	The so-called **major** scale can be found in music.
Wir haben eine neue **Tour** gebucht.	We have book a new **tour**.
Der große **Deich** schützt uns.	The large **dike** protects us.
Der schöne **Teich** vor unserem Haus ist neu.	The nice **pond** in front of our house is new.
Diese **Delle** in meinem Auto ist alt.	The **dent** in my car in old.
Der blaue **Teller** ist kaputt gegangen.	The blue **plate** broke.
Der kurze **Docht** in der Kerze brennt nicht.	The short **wick** of the candle does not burn.
Meine **Tochter** hat morgen Geburtstag.	My **daughter**'s birhtday is tomorrow.
Die dunkle **Gasse** ist mir unheimlich.	The dark **alley** is scary.
Die schwere **Kasse** steht auf dem Tisch.	The heavy **cash register** is placed on the table.
Ich brauche eine **Gabel** zum Essen.	I need a **fork** to eat.
Ich brauche das rote **Kabel** für den Fernseher.	I need the red **cable** for the tv.
Der große **Garten** von Heidi ist schön.	The large **garden** of Heidi is nice.
Paul hat seine **Karten** für das Konzert vergessen.	Paus forgot his **tickets** for the concert.
Der weiße Guss auf dem Kuchen ist lecker.	The white **icing** on the cake is delicious.
Der erste **Kuss** ist etwas Besonderes.	The first **kiss** is something special.
Er hat ihr seine **Gunst** geschenkt.	He gave her his **love**.
Die gegenwärtige **Kunst** mag ich nicht.	I do not like contemporary **art**.
Einflussreiche **Gönner** haben viel Macht.	Influental **patrons** are in power.
Nur richtige **Könner** werden diese Arbeit schaffen.	Only **skilled persons**can do this job.
Ich habe das starke **Begehren** nach Schokolade.	I have the strong **desire** for chocolate.
Das christiliche **Bekehren** ist deine Entscheidung.	It is your decision to **convert** to Christianity.

Das kleine **Boot** schwimmt in der Mitte des Sees.	The small **boat** swims in the middle of the lake.
Der schmale **Bach** fließt vor unserem Haus.	The narrow **stream** runs in front of our house.
Das dicke **Buch** steht im Schrank.	The thick **book** is stored away.
Der braune **Bär** macht Winterschlaf.	The brown **bear** hibernates.
Ihre **Puppe** hat braune Haare.	Her **doll** has brown hair.
Der alte **Pirat** hat einen Papagei.	The old **pirate** has a parrot.
Ich habe **Panik** vor der Prüfung.	I have exam **anxiety**.
Seine **Post** ist angekommen.	His **mail** came.
Der geschickte **Dieb** hat meine Geldbörse gestohlen.	The skilled **thief** stole my purse.

Das genaue **Datum** steht noch nicht fest.	The exact **date** has not been set yet.
Die warme **Decke** ist in der Wäsche.	The warm **blanket** is in the wash.
Der linke **Daumen** ist kürzer als der rechte.	The left **thumb** is shorter than the right one.
Diese **Tasse** hat mir mein Bruder geschenkt.	My brother gave me this **cup**.
Diese **Täter** sind schuldig.	These **offenders** are guilty.
Das wilde **Tier** hat scharfe Zähne.	The wild **animal** has sharp teeth.
Das große **Theater** steht in Frankfurt.	The big **theater** is in Frankfurt.
Unsere **Gans** hat einen Namen.	Our **goose** has a name.
Seine **Geige** ist kaputt.	His **violin** is broken.
Der hohe **Gipfel** ist mit Schnee bedeckt.	The high **peak** is covered with snow.
Die grüne **Gurke** ist lecker.	The **cucumber** is delicious.
Wir haben eine **Katze** mit weißen Flecken.	We have a **cat** with white spots.
Meine **Kette** ist rot.	My **necklace** is red.
Der neue **Komiker** ist witzig.	The new **comedian** is funny.
Die braune **Kuh** gibt viel Milch.	The brown **cow** produces a lot of milk.

Table B.1: German recording material of experiment II. The first part of the table shows the experiment targets, the second part of the table shows the training targets.

Je prendrai des **bains** pour me détendre la semaine prochaine.	I will take the relaxing **baths** next week.
Les **pains** au chocolat sont délicieux.	The chocolate **rolls** are delicious.
Les **boules** sont lourdes.	The **balls** are heavy.
Les **poules** âgée pondent un oeuf tous les jours.	The old **hens** lay an egg every day.
Il fait des **bonds** de joie.	He **jumps for joy**.
Les **ponts** sont vieux.	The **bridges** are old.
Les **biques**[11] ont assez à manger.	The **goats** have enough to eat.
Les **piques** sont faites en bois.	The **spears** are made of wood.
Les **battes** de baseball sont très chère.	The baseball **bats** are expensive.
Les **pattes** de mon chat sont blances.	My cat's **paws** are white.
Les **bars** à Francfort ne sont pas bon marché.	The **bars** in Frankfurt are not cheap.
Les **parts** de gâteau sont grandes	The cake **slices** are big.
Les **bas** ont des trous.	My **socks** have holes.
J'entends mes **pas** dans la neige.	I hear my **steps** in the snow.
Les **douches** sont propres.	The **showers** are clean.
Appuyes sur les **touches** noires.	Press the black **buttons**.
L'animal a des **dents** pointues.	The animal has sharp **teeth**.
Les **temps** sonst bons.	**Times** are good[12]
Les **daims** vivent dans la forêt	The **fallow deers** live in the forrest.

[11] Not the standard word for goat.
[12] Means: Grammar is correct.

Vous avez un mauvais **teint** une mauvaise peau.	You have a bad **complexion** and bad skin.
Je ne peux pas trouver les **dés** dans la boîte.	I cannot find the **dices** in th box.
Les **thés** sont épuisés.	The **tea** is sold out.
Les **dos** des lapins sont sensibles.	The **backs** of rabbits are sensitive.
Les **taux** sont trop élevés.	The **rates** are too high.
Nous recueillons des **dons** pour un hôpital.	We collect **donations** for a hospital.
Les **thons** sont des gros poissons.	**Tunas** are big fish.
J'ai des **dettes** de jeu.	I have gambling **debts**.
Les **têtes** des éléphants sont grandes.	The **heads** of elephants are big.
Les **gaz** d'échappement sont toxiques.	The **fumes** are toxic.
Les **cases** sont petites.	The **boxes** are small.
Toutes les **gares** sont bruyantes.	All **train stations** are noisy.
Les **cars** ne sont pas confortables.	The **coaches** are not comfortable.
Les **gants** sont beaux.	The **gloves** are beautiful.
Les **camps** d'été sont pour les enfants.	Youth **camps** are for children.
Les **goûts** sont très élevés.	The **costs** are really high.
Les **coûts** sont différents.	**Tastes** differ.
Les **gages** sont payés à la fin du mois.	The **wages** will be payed at the end of the month.
Les **cages** à oiseaux sont très petites.	The **cages** for birds are really small.
Les **gars** de la marine sont beaux.	The **boys** from the navy are handsome.
Ces sont des **cas** difficiles.	These are difficult **affairs**.
Les **garrots** de pression peuvent étre vitaux.	The **tourniquets** can be vital.
Les **carreaux** sont sales.	The **window panes** are dirty.

Les **balles** ont des couleurs différentes.	The **balls** have different colours.
Les **bonbons** ne sont pas bons.	The **candies** do not taste good.
Les **billet** sont gratuits.	The **tickets** are for free.
Les **bus** sont pleins.	The **buses** are crowded.
Les **poux** sont résistants.	The **lice** are persitant.
Les **pommes** sont m̂ures.	The **apples** are mellow.
Les **personnes** sont dans la cuisine.	The **persons** are in the kitchen.
Les **palais** des rois sont magnifiques.	The king's **palaces** are magnificent.
J'ai des **doutes** sur le mariage.	I have **doubts** about the wedding.
Mes **dents** sont de travers.	My **teeths** are crooked.
Les **documents** sont des faux.	The **documents** are forged.
Les **disciplines** des Jeux Olympiques sont variées.	The **disciplines** of the Olympic Games are diverse.
J'ai des **taches** sur mon pantalon.	I have **stains** on my trousers.
Les **tulipes** sont orange.	The **tulips** are orange.
Les **tomates** sont rouges.	The **tomatoes** are red.
Les **téléviseurs** sont de moins en moins chers.	The **television sets** become cheaper and cheaper.
Les **gouttes** d'eau sont chaudes.	The water **drops** are warm.
Les **garages** sont sombres.	The **garages** are dark.
Les **goélands** sont blancs.	The large **seagulls** are white.
Les **gazettes** sont démodées.	The **journals** are old.
Les **cours** sont ennuyeux.	The **courses** are boring.

Les **carpes** sont délicieuses.	The **carps** are delicious.
Les **comédiens** sont drô	The **comedians** are funny.
Les **canards** sont jaunes.	the **ducks** are yellow.

Table B.2: French recording material of experiment II. The first part of the table shows the experiment targets, the second part of the table shows the training targets.

Appendix C

Recording Material for Experiment IV

	bilabial	alveolar	velar
initial	baie-paix	dé-thé	gage-cage
	bail-paille	dette-tête	galle-cale
	bain-pain	don-thon	galop-calot
	bal-pale	dos-taux	garrot-carreau
	ballot-palot	*daim-teint	*gable-câble
	banc-paon	*dard-tare	*gosse-cosse
	bar-part	*dent-temp	*gaz-case
	barque-parc	*douche-touche	*goût-coût
	bas-pas		
	basse-passe		
	bateau-pataud		
	batte-patte		
	battée-pâtée		
	baume-paume		
	baux-peau		
	bêche-pêche		
	beigne-peigne		
	benne-peine		
	berme-perme		
	beurre-peur		
	bile-pile		
	bique-pique		
	bond-pont		
	bonnet-poney		
	bord-port		
	botte-pote		
	boue-pou		
	bouffe-pouf		

	bouille-pouilles		
	boule-poule		
	bulle-pull		
	burin-purin		
	*ballet-palais		
	*baquet-paquet		
	*belle-pelle		
	*bus-put		
medial	*abord-apport	*badaud-bateau	*agui-acquis
	*débit-dépit	*édile-éthyle	*égard-écart
	*rabbin-rappin	*radeau-râteau	*hangar-encart
final	*cab-cape	*bled-blette	*bague-bac
	*crambe-crampe	*code-côte	*bègue-bec
	*rab-rap	*lad-latte	*bigue-bique
	*trombe-trompe	*ride-rite	*dogue-dock

Table C.1: French recording material for experiment IV. The minimal pairs marked with an asterix represent words included in the generalization tests.

Bibliography

Ammar, Ahlem (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research*, 12: 183–210.

Andreeva, Bistra, Grażyna Demenko, Bernd Möbius, Frank Zimmerer, Jeanin Jügler and Magdalena Oleskowicz-Popiel (2014a). Differences of pitch profiles in Germanic and Slavic Languages. *Proc. Interspeech 2014, Singapore*, 1307–1311.

Andreeva, Bistra, Grazyna Demenko, Magdalena Wolska, Bernd Möbius, Frank Zimmerer, Jeanin Jügler and Jürgen Trouvain (2014b). Comparison of pitch range and pitch variation in Slavic and Germanic languages. *Proc. Speech Prosody 2014, Dublin*, 776–780.

Andreeva, Bistra, Bernd Möbius, Grażyna Demenko, Frank Zimmerer and Jeanin Jügler (2015). Linguistic measures of pitch range in Slavic and Germanic languages. *Proc. Interspeech 2015, Dresden*, 968–972.

Baese-Berk, Melissa M. and Tuuli H. Morrill (2015). Speaking rate consistency in native and non-native speakers of English. *Journal of the Acoustical Society of America*, 138: EL223–EL228.

Baker, Amanda A. (2011). ESL teachers and pronunciation pedagogy: Exploring the development of teachers' cognitions and classroom practices. *Proc. Pronunciation in Second Language Learning and Teaching 2011, Ames*, 82–94.

Ball, Martin and Joan Rahilly (2013). *Phonetics: The Science of Speech*. London, New York: Routledge.

Barry, William J. (2007). Rhythm as an L2 problem: How prosodic is it? Jürgen Trouvain and Ulrike Gut (Editors), *Non-Native Prosody. Phonetic Description and Teaching Practice*, Berlin: Mouton de Gruyter, 97–120.

Bartram, Mark and Richard Walt (1991). *Correction: A Positive Approach to Language Mistakes*. Hove: Language Teaching Publications.

Best, Catherine T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. J. Goodman and H.C. Nusbaum (Editors), *The development of speech perception: The transition from speech to spoken words*, Cambridge, MA: MIT Press, 167–224.

Best, Catherine T. (1995). A direct realist perspective on cross-language speech perception. Winifred Strange (Editor), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, Baltimore, MD: York Press, 167–200.

Best, Catherine T. and Pierre A. Hallé (2010). Perception of initial obstruent voicing is influenced by gestural organization. *Journal of Phonetics*, 38: 109–126.

Beyer, Franz (1908). *Französische Phonetik für Lehrer und Studierende*. Cöthen: Otto Schulze.

Bissiri, Maria P. and Hartmut R. Pfitzinger (2009). Italian speakers learn lexical stress of German morphologically complex words. *Speech Communication*, 51: 933–947.

Boersma, Paul and David Weenink (2013). Praat: Doing phonetics by computer. [Version 5.3.51]. URL `http://www.praat.org`.

Bonneau, Anne (2015). Realization of French voiced fricatives by German learners as a function of speaker level and prosodic boundaries. *Proc. International Congress of Phonetic Science 2015, Glasgow.*

Bonneau, Anne and Martine Cadot (2015). German non-native realizations of French voiced fricatives in final position of a group of words. *Proc. Interspeech 2015, Dresden*, 1927–1931.

Boula de Mareüil, Philippe and Bianca Vieru-Dimulescu (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, 63: 247–267.

Bouselmi, Ghazi, Dominique Fohr and Irina Illina (2012). Multilingual recognition of non-native speech using acoustic model transformation and pronunciation modeling. *International Journal of Speech Technology*, 15: 203–213.

Bradlow, Ann R., Reiko Akahane-Yamada, David B. Pisoni and Yoh'ichi Tohkura (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustic Society of America*, 101: 2299–2310.

Bradlow, Ann R., Reiko Akahane-Yamada, David B. Pisoni and Yoh'ichi Tohkura (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Percept Psychophys*, 61: 977–985.

Bresnahan, Mary Jiang, Rie Ohashi, Reiko Nebashi, Weng Ying Liu and Sachiyo Morinaga Shearman (2002). Attitudinal and affective response towards accented English. *Language and Communication*, 22: 171–185.

Brown, Dan (2016). The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research*, 20: 436–458.

Busà, Maria Grazia and Antonio Stella (2012). Intonational variations in focus marking in the English spoken by North-East Italian speakers. Methodological perspectives on second language prosody. Maria Grazia Busà and Antonio Stella (Editors), *Papers from ML2P 2012*, 31–36.

Busà, Maria Grazia and Martina Urbani (2011). A cross linguistic analysis of pitch range in English L1 and L2. *Proc. International Congress of Phonetic Sciences 2011, Hong Kong*, 380–383.

Butler, Deborah L. and Philip H. Winne (1995). Feedback and self-regulated learning: A theoretic synthesis. *American Educational Research Association*, 64: 245–281.

Cauldwell, Richard (2012). Cool Speech: Hot Listening, Cool Pronunciation. iPad Application. Last access: 04/10/2016, URL `https://itunes.apple.com/de/app/cool-speech/id501050883?mt=8`.

Chervinski, Juri and Elissa Pustka (2010). *IPFC-allemand: Une pré-enquête auprès de quelques étudiants munichois. Journée IPFC2010: Interphonologie, corpus et français langue étrangère*. Paris: MSH.

Colotte, Vincent (2013). Corpus Recorder. URL `https://raweb.inria.fr/rapportsactivite/RA2011/parole/uid63.html`.

Curlee, Richard F. and William H. Perkins (1963). Effectiveness of a day conditioning program for adolescent and adult stutterers. *Behaviour Research and Theory*, 11: 395–401.

De Jong, Nivja H., Rachel Groenhout, Rob Schoonen and Jan H. Hulstijn (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36: 223–243.

De Meo, Anna, Marilisa Vitale, Massimo Pettorino, Francesco Cutugno and Antonio Origlia (2016). Imitation/self-imitation in computer-assisted prosody training for Chinese learners of L2 Italian. *Proc. Pronunciation in Second Language Learning and Teaching Conference, Ames*, 90–100.

Delattre, Pierre (1964). Comparing the vocalic features of English, German, Spanish and French. *IRAL - International Review of Applied Linguistics in Language Teaching*, 2: 77–97.

Demenko, Grażyna, Natalia Dylwik and Agnieszka Wagner (2009). Applying speech and language technology to foreign language education. *Proc. International Multiconference on Computer Science and Information Technology 2009, Mrągowo, Poland*, 457–462.

Di Cristo, Albert (1998). Intonation in French. Daniel Hirst and Albert Di Cristo (Editors), *Intonation Systems: A Survey of Twenty Languages*, Cambridge, England: Cambridge University Press, 195–218.

Dolson, Mark (1994). The pitch of speech as function of linguistic community. *Music Perception*, 11: 321–331.

Draxler, Christoph (2011). Percy – an HTML5 framework for media rich web experiments on mobile devices. *Proc. Interspeech 2011, Florence*, 3339–3340.

Draxler, Christoph (2014). Online experiments with the Percy software framework – experiences and some early results. *Proc. Language Resources and Evaluation Conference 2014, Reykjavik*, 235–240.

Ellis, Rod, Shawn Loewen and Rosemary Erlamd (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28: 339–368.

EnglishCentral, Inc. (2016). EnglishCentral. Last access: 04/10/2016, URL `http://www.englishcentral.com`.

Escudero, Paola (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization.* Ph.D. thesis, Universiteit Utrecht.

Eskenazi, Maxine and Scott Hansma (1998). The fluency pronunciation trainer. *Proc. Speech Technology in Language Learning 1998, Marholmen*, 77–80.

Eskenazi, Maxine, Yan Ke, Jordi Albornoz and Katharina Probst (2000). The fluency pronunciation trainer: Update and user issues. *Proc. inSTILL 2000, Dundee*, 73–76.

EyeSpeak (2016). EyeSpeak Online Free Trial. Last access: 04/10/2016, URL `http://www.eyespeakpro.com`.

Fagyal, Zsuzsanna, Douglas Kibbee and Frederic Jenkins (2006). *French: A Linguistic Introduction.* Cambridge: Cambridge University Press.

Fauth, Camille, Anne Bonneau, Frank Zimmerer, Jürgen Trouvain, Bistra Andreeva, Vincent Colotte, Dominique Fohr, Denis Jouvet, Jeanin Jügler, Odile Mella, Bernd Möbius and Frank Zimmerer (2014). Designing a bilingual speech corpus for French and German language learners: A two-step process. *Proc. Language Resources and Evaluation Conference 2016, Reykjavik*, 1477–1482.

Felps, Daniel, Heather Bortfeld and Ricardo Gutierrez-Osuna (2009). Foreign accent conversion in computer assisted pronunciation training. *Speech Communication*, 51: 920–932.

Féry, Caroline (1998). German word stress in optimality theory. *Journal of Comparative Germanic Linguistics*, 2: 101–142.

Féry, Caroline (2001). Focus and Phrasing in French. Caroline Féry and Wolfgang Sternefeld (Editors), *Audiatur Vox Sapientiae. A Festschrift for Arnim von Stechow*, Berlin: Akademie-Verlag, 153–181.

Féry, Caroline (2014). Final compression in French as a phrasal phenomenon. Stacey Katz Bourns and Lindsy L. Myers (Editors), *Perspectives on Linguistic Structure and Context: Studies in Honour of Knud Lambrecht*, Amsterdam: John Benjamins, 133–156.

Féry, Caroline, Robin Hörnig and Serge Pahaut (2011). Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. Christoph Gabriel and Conxita Lleó (Editors), *Intonational Phrasing in Romance and Germanic: Cross-linguistic and bilingual studies*, Amsterdam: John Benjamins, 11–41.

Flege, James E. (1979). *Phonetic Interference in Second Language Learning.* Department of Linguistics. Indiana University. Bloomington: unpublished doctoral dissertation.

Flege, James E. (1995). Second-language speech learning: Theory, findings, and problems. W. Strange (Editor), *Speech Perception and Linguistic Experience: Issues in Cross-Language research*, Timonium, MD: York, 229–273.

Flege, James Emil (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15: 47–65.

Fohr, Dominique and Odile Mella (2012). A software for comparing automatic labeling tools. *Proc. Language Resources and Evaluation Conference 2012*, 325–332.

Gass, Susan M. (1991). Grammar instruction, selective attention, and learning processes. Robert Phillipson, Eric Kellerman, Mike Sharwood Smith and Merrill Swain (Editors), *Foreign/Second Language Pedagogy Research*, Clevdon, UK: Multilingual Matters, 134–141.

Gass, Susan M. (1997). *Input, Interaction, and the Second Language Learner*. Hillsdale, NJ: Erlbaum.

Geckeler, Horst and Wolf Dietrich (2012). *Einführung in die französische Sprachwissenschaft*. Berlin: Erich Schmidt Verlag.

Ghosh, Sucheta, Camille Fauth, Aghilas Sini and Yves Laprie (2016). L1-L2 interference: The case of final devoicing of French voiced fricatives in final position by German learners. *Proc. Interspeech 2016, San Francisco*, 3156–3160.

Goossens, Diane and Sylviane Granger (2016). Learner corpora around the world. Last access: 05/06/2016, URL `https://www.uclouvain.be/en-cecl-lcworld.html`.

Goronzy, Silke, Marina Sahakyan and Wolfgang Wokurek (2001). Is nonnative pronunciation modeling necessary. *Proc. Eurospeech, Aalborg*, 209–312.

Guion, Susan G., James E. Flege, Serena H. Liu and Grace H. Yeni-Komshian (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21: 205–228.

Gut, Ulrike (2009). *Non-Native Prosody. A Corpus-Based Analysis of the Phonetic and Phonological Properties of L2 English and L2 German*. Frankfurt: Peter Lang.

Hammarström, Göran (1998). *Französische Phonetik*. Tübingen: Narr.

Harmer, Jeremy (2001). *The Oractice of English Language Teaching*. Essex: Person Education Limited.

Hattie, John (1999). Influences on student learning. *Inaugural professorial address, University of Auckland, New Zealand*. Retrieved from http://projectlearning.org/blog/wp-content/uploads/2014/02/Influences-on-Student-Learning-John-Hattie.pdf.

Hattie, John and Helen Timperley (2007). The Power of feedback. *Review of Educational Research*, 77: 81–112.

Hayes-Harb, Rachel and Johanna Watzinger-Tharp (2012). Accent, intelligibility, and the role of the listener: Perceptions of English-accented German by native German Speakers. *Foreign Language Annals*, 45: 260–282.

Hendrickson, James M. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *The Modern Language Journal*, 62: 387–398.

Henry, Guillaume, Anne Bonneau and Vincent Colotte (2007). Tools devoted to the acquisition of the prosody of a foreign language. *Proc. International Congress of Phonetic Sciences, Saarbrücken*, 1593–1596.

Hincks, Rebecca and Jens Edlund (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology*, 13: 32–50.

Hirose, Keikichi, Frédéric Gendrin and Nobuaki Minematsu (2003). A pronunciation training system for Japanese lexical accents with corrective feedback in learner's voice. *Proc. European Conference on Speech Communication and Technology 2003, Geneva*, 3149–3152.

Hirschfeld, Ursula (2012). Bewegte Phonetik. Ausspracheübungen mit französischen Schülern im Deutsch- und Englischunterricht (classes bilangues). Goethe-Institut Paris. März 2014. Last access: 05/10/2016, URL `http://www.goethe.de/ins/fr/lp/prj/clb/unt/pho/deindex.htm`.

Hirschfeld, Ursula and Jürgen Trouvain (2007). Teaching prosody in German as a foreign language. Jürgen Trouvain and Ulrike Gut (Editors), *Non-Native Prosody. Phonetic Description and Teaching Practice*, Berlin: Mouton de Gruyter, 171–187.

Horwitz, Elaine K., Michael B. Horwitz and Joann Cope (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70: 125–132.

Ingham, Roger J. and Gavin Andrews (1971). Stuttering: The quality of fluency after treatment. *Journal of Communication Disorders*, 4: 279–283.

IPDS (1994). The Kiel Corpus of Read Speech. Volume 1, CD-ROM #1, URL `http://www.ipds.uni-kiel.de/forschung/kielcorpus.de.html`.

Jilka, Matthias (2000). Testing the contribution of prosody to the perception of foreign accent. *Proc. New Sounds 2000, Amsterdam*, 199–207.

JMP (2014). [Version 11.0.0], SAS Institute Inc., Cary, NC.

Jokisch, Oliver, Uwe Koloska, Diane Hirschfeld and Rüdiger Hoffmann (2005). Pronunciation learning and foreign accent reduction by an audiovisual feedback system. Jianhua Tao, Tieniu Tan and Rosalind W. Picard (Editors), *Affective Computing and Intelligent Interaction*, Berlin: Springer, 419–425.

Jouvet, Denis, Anne Bonneau, Jürgen Trouvain, Frank Zimmerer, Yves Laprie and Bernd Möbius (2015). Analysis of phone confusion matrices in a manually annotated French-German learner corpus. *Proc. Workshop on Speech and Language Technology for Education (SLaTE), Leipzig*, 107–112.

Jouvet, Denis, Larbi Mesbahi, Anne Bonneau, Dominique Fohr, Irina Illina and Yves Laprie (2011). Impact of pronunciation variant frequency on automatic non-native speech segmentation. *Proc. Language & Technology Conference 2011, Poznan*, 145–148.

Jügler, Jeanin and Bernd Möbius (2015). Auditory feedback methods to improve the pronunciation of stops by German learners of French. *Proc. International Congress of Phonetic Sciences 2015, Glasgow.*

Jügler, Jeanin, Frank Zimmerer, Bernd Möbius and Christoph Draxler (2015). The effect of high variability training on the perception and production of French stops by German native speakers. *Proc. Interspeech 2015, Dresden*, 806–810.

Jügler, Jeanin, Frank Zimmerer, Jürgen Trouvain and Bernd Möbius (2016). The perceptual effect of L1 prosody transplantation on L2 Speech: The case of French accented German. *Proc. Interspeech 2016, San Francisco*, 67–71.

Jun, Sun-Ah and Cécile Fougeron (2000). A phonological model of French intonation. A. Botinis (Editor), *Intonation: Analysis, Modeling and Technology*, Dordrecht: Kluwer, 209–242.

Kaglik, Anna and Philippe Boula de Mareüil (2010). Polish-accented French prosody in perception and production: Transfer or universal acquisition process? *Proc. Speech Prosody 2010, Chicago*, 1–4.

Kalinowski, Joseph, Joy Armson, Andrew Stuart, Marek Roland-Mieszkowski and Vincent L. Gracco (1993). Effects of alterations in auditory feedback and speech rate on stuttering frequency. *Language and Speech*, 36: 1–16.

Kalinowski, Joseph, Andrew Stuart, Sarah Sark and Joy Armson (1996). Stuttering amerlioration at various auditory feedback delays and speech rates. *European Journal of Disorders of Communication*, 31: 259–269.

Kamiyama, Takeki, Barbara Kühnert and Jacqueline Vaissière (2011). Do French-speaking learners simply omit the English /h/? *Proc. International Congress of Phonetic Sciences 2011, Hong Kong*, 1010–1013.

Kang, Eunyoung and Zhaohong Han (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *Modern Language Journal*, 99: 1–18.

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Frauenfelder and Narly Golestani (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America*, 138: 817–832.

Kingston, John (2003). Learning foreign vowels. *Language and Speech*, 46: 295–349.

Klatt, Dennis H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18: 686–706.

Kleiner, Stefan and Ralf Knöbel (2015). *Duden - Das Aussprachewörterbuch.* Berlin: Dudenverlag. 7., komplett überarbeitete und aktualisierte Auflage.

Kohler, Klaus J. (1995). *Einführung in die Phonetik des Deutschen.* Berlin: Erich Schmidt Verlag.

Komissarchik, Edward and Julia Komissarchik (2000a). BetterAccent Tutor – analysis and visualization of speech prosody. *Proc. InSTILL 2000, Dundee*, 86–89.

Komissarchik, Julia and Edward Komissarchik (2000b). Application of knowledge-based speech analysis to suprasegmental pronunciation. *Proc. AVIOS 2000, San Jose*, 243–248.

Kuhl, Patricia K. (1992). Infants' perception and representation of speech: Development of a new theory. *Proc. International Conference on Spoken Language Processing, Banff*, 449–456.

Kuhl, Patricia K., Barbara T. Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola and Tobey Nelson (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, 363: 979–1000.

Kuhl, Patricia K. and Paul Iverson (1995). Linguistic experience and the "Perceptual Magnet Effect". Winifred Strange (Editor), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, Baltimore, MD: York Press, 121–154.

Kuhl, Patricia K., Erica Stevens, Akiko Hayashi, Toshisada Deguchi, Shigeru Kiritani and Paul Iverson (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9: F13–F21.

Kulhavy, Raymond W. (1977). Feedback in written instruction. *Review of Educational Research*, 47: 254–284.

Kulik, James A. and Chen-Lin C. Kulik (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58: 79–97.

Künzel, Hermann J. (1977). Signalphonetische Untersuchung deutsch-französischer Interferenzen im Bereich der Okklusive. *Forum Linguisticum*, 10.

Kupisch, Tanja, Dagmar Bartonand, Giulia Bianchi and Ilse Stangen (2012). The HABLA-corpus (German-French and German-Italian). Thomas Schmidt and Kai Wörner (Editors), *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam: Benjamins, 163–179.

Lambacher, Stephen G., William L. Martens, Kazuhiko Kakehi, Chandrajith A. Marasinghe and Garry Molholt (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26: 227–247.

Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner and Samuel Fillenbaum (1960). Evaluational reactions to spoken languages. *Journal of Abnormal Social Psychology*, 60: 44–51.

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

Lee, Bernard S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America*, 22: 824–826.

Levelt, Willem J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA, London: MIT Press.

Levine, Marvin (1975). *A Cognitive Theory of Learning: Research in Hypothesis Testing.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Levy, Erika S. and Franzo F Law (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship). *The Journal of the Acoustical Society of America*, 128: 1290–1305.

Li, Shaofeng (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60: 308–365.

Li, Shaofeng, Yan Zhu and Rod Ellis (2016). The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *Modern Language Journal*, 100: 276–295.

Lisker, Leigh and Arthur S. Abramson (1964). A cross-language study of voicing in initial stops. *Word*, 20: 527–565.

Lister, Roy and Kazuya Saito (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32: 265–302.

Liu, Qiandi and Dan Brown (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30: 66–81.

Logan, John S., E. Lively, Scott and David B. Pisoni (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustic Society of America*, 89: 874–886.

Long, Michael H (1977). Teacher feedback on learner error: Mapping cognitions. H. D. Brown, C A Yorio and R H Crymes (Editors), *On TESOL '77*, Washington: TESOL, 278–293.

Long, Michael H. (1996). The role of the linguistic environment in second language acquisition. W. C. Ritchie and T. K. Bhatia (Editors), *Handbook of Second Language Acquisition*, San Diego, CA: Academic Press, 413–468.

Lu, Jingli, Ruili Wang, Liyanage C. De Silva, Yang Gao and Jia Liu (2010). CASTLE: A computer-assisted stress teaching and learning environment for learners of English as a second language. *Proc. Interspeech 2010, Makuhari, Japan*, 606–609.

Luce, Paul A. and Jan Charles-Luce (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, 78: 1949–1957.

Lyster, Roy (1998). Recasts, repetition and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20: 51–81.

Lyster, Roy and Leila Ranta (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19: 37–66.

Lyster, Roy, Kazuya Saito and Masatoshi Sato (2013). State-of-the-art article: Oral corrective feedback in second language classrooms. *Language Teaching*, 46: 1–40.

Mackey, Alison and J Goo (2007). Conversational interaction in second language acquisition. Oxford applied linguistics. *Interaction Research in SLA: A Meta-Analysis and Research Synthesis*, Oxford: Oxford University Press, 407–453.

Mackey, Alison and Jenefer Philp (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal*, 82: 338–356.

Mak, Brian, Manhung Siu, Mimi Ng, Yik-Cheung Tam, Yu-Chung Chan, Kin-Wah Chan, Ka-Yee Leung, Simon Ho, Fong-Ho Chong, Jimmy Wong and Jacqueline Lo (2003). PLASER: Pronunciation learning via automatic speech recognition. *Proc. HLT-NAACL 2003 Workshop on Building Educational Applications using Natural Language Processing*, 23–29.

Martin, Philippe (2004). WinPitch LTL II, a multimodal pronunciation software. *Proc. InSTILL 2004, Venice*, 177–182.

Martin, Philippe (2010). Learning the prosodic structure of a foreign language with a pitch visualizer. *Proc. Speech Prosody 2010, Chicago.*

McClaskey, Cynthia L., David. B. Pisoni and Thomas D. Carrell (1983). Transfer of training of a new linguistic contrast in voicing. *Percept Psychophys*, 34: 323–330.

Meisenburg, Trudel and Maria Selig (2006). *Phonetik und Phonologie des Französischen.* Stuttgart: Klett.

Mennen, Ineke (1998). Can language learners ever acquire the intonation of a second language? *Proc. STiLL 1998, Marholmen*, 17–20.

Mennen, Ineke, Felix Schaeffler and Gerard Docherty (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. *Proc. International Congress of Phonetic Sciences 2007, Saarbrücken*, 1769–1772.

Mennen, Ineke, Felix Schaeffler and Gerard Docherty (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, 131: 2249–2260.

Meron, Yoram and Keikichi Hirose (1996). Language training system utilizing speech modification. *Proc. International Conference on Spoken Language 1996, Philadelphia*, 1449–1452.

Mikuteit, Simone and Henning Reetz (2007). Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech*, 50: 247–277.

Miller, Paul Chamness (2003). *The Effectiveness of Corrective Feedback: A Meta-Analysis.* Ph.D. thesis, Purdue University, West Lafayette, IN. Unpublished thesis.

Möbius, Bernd (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica*, 38: 5–26.

Moulines, Eric and Francis Charpentier (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9: 453–468.

Munro, Murray J. and Tracey M. Derwing (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, 11: 253–266.

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

Munro, Murray J. and Tracey M. Derwing (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38: 389–406.

Munro, Murray J., Tracey M. Derwing and Kyoko Sato (2006). Salient accents, covert attitudes: Consciousness raising pre-service second language teachers. *Prospect*, 21: 67–79.

Nagano, Keiko and Kazunori Ozawa (1990). English speech training using voice conversion. *Proc. International Conference on Spoken Language Processing 1990, Kobe*, 1169–1172.

Neri, Ambra, Catia Cucchiarini and Helmer Strik (2008a). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20: 225–243.

Neri, Ambra, Catia Cucchiarini, Helmer Strik and Lou Boves (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15: 441–467.

Neri, Ambra, Ornella Mich, Matteo Gerosa and Diego Giuliani (2008b). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21: 393–408.

Neuhauser, Sara (2012). *Phonetische und linguistische Aspekte der Akzentimitation im forensischen Kontext*. Tübingen: Narr.

Neumeyer, Leonardo, Horacio Franco, Victor Abrash, Luc Julia, Orith Ronen, Harry Bratt, Jehan Bing, Vissilis Digalakis and Marikka Rypa (1998). Webgrader TM: A multilingual pronunciation practice tool. *Proc. STiLL Workshop 1998, Marholmen*, 61–64.

Nicol, David J. and Debra Macfarlane-Dick (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31: 199–218.

Nicolaidis, Katerina, George Papanikolaou, Evia Kainada and Konstantinos Avdelidis (2015). Speak Greek: An online speech training tool for L2 pedagogy and clinical intervention. *Proc. International Congress of Phonetic Sciences 2015, Glasgow.*

Nishi, Kanae and Diane Kewley-Port (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, 50: 1496–1509.

Nishi, Kanae and Diane Kewley-Port (2008). Non-native speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*, 51: 1480–1493.

Peabody, Mitchell and Stephanie Seneff (2006). Towards Automatic Tone Correction in Non-native Mandarin. *Proc. International Symposium Chinese Spoken Language Processing 2006, Singapore*, 602–613.

Pillot-Loiseau, Claire, Takeki Kamiyama and Tanja Kocjančič Antolik (2015). French /y/-/u/ contrast in Japanese learners with/without ultrasound feedback: Vowels, non-words and words. *Proc. International Congress of Phonetic Sciences 2015, Glasgow*, 1–5.

Pisoni, David B., Richard N. Aslin, Alan J. Perey and Beth. L. Hennessy (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8: 297–314.

Precoda, Kristin, Christine A. Halverson and Horacio Franco (2000). Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability. *Proc. InSTILL 2000, Dundee, Scotland*, 102–105.

Prince, Alan and Paul Smolensky (1993). Optimality theory: Constraint interaction in generative grammar. *Rutgers University Center for Cognitive Science Technical Report 2*.

Probst, Katharina, Yan Ke and Maxine Eskenazi (2002). Enhancing foreign language tutors - In search of the golden speaker. *Speech Communication*, 37: 161–173.

Pustka, Elissa (2011). *Einführung in die Phonetik und Phonologie des Französischen*. Berlin: Erich Schmidt Verlag.

Quinn, Paul (2014). *Delayed Versus Immediate Corrective Feedback on Orally Produced Passive Errors in English*. Ph.D. thesis, University of Toronto. Unpublished thesis.

Raupach, Manfred (1980). Temporal variables in first and second language speech production. H Dechert and M Raupach (Editors), *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton, 263–270.

Reetz, Henning and Allard Jongman (2009). *Phonetics. Transcription, Production, Acoustics, and Perception*. Oxford: Wiley-Blackwell.

Riggenbach, Heidi (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes*, 14: 423–441.

Ringbom, Håkan (2007). *Cross-Linguistic Similarity in Foreign Language Learning*. Clevedon, Buffalo, Toronto: Multilingual Matters LTD.

Rochet, Bernard L. (1995). Perception and production of second-language speech sounds by adults. W. Strange (Editor), *Speech Perception and Linguistic Experience: Issues in Cross-Language research*, Timonium, MD: York Press, 379–410.

Rognoni, Luca and Maria Grazia Busà (2013). Testing the effects of segmental and suprasegmental phonetic cues in foreign accent rating: An experiment using prosody transplantation. *Proc. New Sounds 2014, Montreal*, 547–560.

Rose Medical Solutions (2016). Pronunciation Coach. Last access: 04/10/2016, URL `http://www.rose-medical.com/pronunciation-coach.html`.

Rosetta Stone (2016). Rosetta Stone. Last access: 04/10/2016, URL `http://www.rosettastone.com`.

Russel, Jane and Nina Spada (2006). Synthesizing research on language learning and teaching. J. Norris and L. Ortega (Editors), *The Effectiveness of Corrective Feedback for the Acquisition of L2 Grammar*, Amsterdam: John Benjamins, 133–162.

Sadakataa, Makiko and James M. McQueen (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of the Acoustic Society of America*, 134: 1324–1335.

Sadler, D. Royce (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18: 119–144.

Schmidt, Richard W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11: 129–158.

Setter, Jane and Jennifer Jenkins (2005). State-of-the-art review article. *Language Teaching*, 38: 1–17.

Skinner, Burrhus Frederic (1957). *Verbal Behavior.* New York: Appleton-Century-Crofts.

Strange, Winifred, Andrea Weber, Erika S. Levy, Valeriy Shafiro, Miwako Hisagi and Kanae Nishi (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America*, 122: 1111–1129.

Stuart, Andrew and Joseph Kalinowski (2015). Effect of delayed auditory feedback, speech rate, and sex on speech production. *Perceptual & Motor Skills: Learning & Memory*, 120: 747–765.

Tateishi, Miwako and Stephen Winters (2013). Does ultrasound training lead to improved perception of a non-native sound contrast: Evidence from Japanese leaners of English. *Proc. Annual Conference of the Canadian Linguistic Association 2013, Victoria*, 1–15.

Thorndike, Edward Lee (1932). *The Fundamentals of Learning.* New York: Columbia Teachers College.

Tranel, Bernard (1987). *The Sounds of French.* Cambridge, New York: Cambridge University Press.

Trouvain, Jürgen, Anne Bonneau, Vincent Colotte, Camille Fauth, Dominique Fohr, Denis Jouvet, Jeanin Jügler, Yves Laprie, Odile Mella, Bernd Möbius and Frank Zimmerer (2016a). The IFCASL corpus of French and German non-native and native read speech. *Proc. Language Resources and Evaluation Conference 2016 , Portorož*, 1333–1338.

Trouvain, Jürgen, Camille Fauth and Bernd Möbius (2016b). Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 read speech. *Proc. Speech Prosody 2016, Boston*, 31–35.

Trouvain, Jürgen, Jeanin Jügler and Yves Laprie (2016c). On the assessment of computer-assisted pronunciation training tools. *Proc. Konferenz Elektronische Sprachsignalverarbeitung 2016, Leipzig*, 175–182.

Trouvain, Jürgen, Yves Laprie, Bernd Möbius, Bistra Andreeva, Anne Bonneau, Vincent Colotte, Camille Fauth, Denis Fohr, Dominique Jouvet, Odile Mella, Jeanin Jügler and Frank Zimmerer (2013). Designing a bilingual speech corpus for French and German language learners. *Proc. Corpus et Outils en Linguistique, Langues et Parole: Status, Usages et Mésuages, Strasbourg, France*, 32–34.

Trouvain, Jürgen and Bernd Möbius (2014a). Individuelle Ausprägung von Atmungspausen in der Mutter- und in der Fremdsprache als Anzeichen kognitiver Belastung. *Proc. Elektronische Sprachsignalverarbeitung 2014, Dresden*, 177–184.

Trouvain, Jürgen and Bernd Möbius (2014b). Sources of variation of articulation rate in native and non-native speech: Comparisons of French and German. *Proc. Speech Prosody 2014, Dublin*, 275–279.

Truscott, John (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46: 327–369.

Truscott, John (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review*, 55: 437–456.

Tsurutani, Chiharu (2012). Evaluation of speakers with foreign-accented speech in Japan: The effect of accent produced by English native speakers. *Journal of Multilingual and Multicultural Development*, 33: 589–603.

Ulbrich, Christiane and Ineke Mennen (2015). When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilinguism*, 1–28.

Ullakonoja, Riikka (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). *Proc. International Congress of Phonetic Sciences 2007, Saarbrücken*, 1701–1704.

Ur, Penny (2012). *A Course in English Language Teaching*. Cambridge: Cambridge University Press.

van Doremalen, Joost, Helmer Strik and Catia Cucchiarini (2009). Optimizing non-native speech recognition for CALL applications. *Proc. Interspeech 2009, Brighton*, 592–595.

van Leussen, Jan-Willem and Paola Escudero (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, 6: article 1000.

Walker, Douglas C. (2001). *French Sound Structure*. Calgary, Alberta: University of Calgary Press.

Wang, Ruili and Jingli Lu (2008). Investigation of golden speakers for second language learners from imitation preference perspective by voice modification. *Speech Communication*, 53: 175–184.

Wang, Yue, Michelle M. Spence, Allard Jongman and Joan. A. Sereno (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustic Society of America*, 106: 3649–3658.

Watson, John Broadus (1926). *Behavioursim*. New York: Norton.

Wiese, Richard (1996). *The Phonology of German*. Oxford: Oxford University Press.

Wik, Preben and Anna Hjalmarsson (2009). Embodied conversational agents in computer assisted language learning. *Speech Communication* , 51: 1024–1037.

Winters, Stephen and Mary Grantham O'Brien (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication*, 55: 486–507.

Wong, Jance Wing Sze (2012). Training the perception and production of English /e/ and /æ/ of Cantonese ESL learners: A comparison of low vs. high variability phonetic training. *Proc. Australasian International Conference on Speech Science and Technology 2012, Sydney*, 37–40.

Wong, Jance Wing Sze (2014). The effects of high and low variability phonetic training on the perception an production of English vowels /e/-/æ/ by Cantonese ESL learners with high and low L2 proficiency levels. *Proc. Interspeech 2014, Singapore*, 524–528.

Wong, Jance Wing Sze (2015). The effects of high-variability phonetic training on Cantonese ESL learners' production of English /ɪ/-/iː/ contrast - an acoustic analysis. *Phonetics Teaching and Learning Conference (PTLC) 2015, London*, 107–111.

Yates, Aubrey J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60: 213–232.

Young, Dolly Jesusita (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest?. *The Modern Language Journal*, 75: 426–439.

Zanini, Sergio, Andrea Clarici, Franco Fabbro and Antonio Bava (1999). Speaking speech effects on delayed auditory feedback disruption on speech fluency. *Perceptual & Motor Skills: Learning & Memory*, 89: 1095–1109.

Zhao, Sixuan, Soo Ngee Koh and Kang-Kwong Luke (2012). Accent reduction for computer-aided language learning. *Proc. European Signal Processing Conference 2012, Bucharest*, 335–339.

Zimmerer, Frank, Bistra Andreeva, Jeanin Jügler and Bernd Möbius (2015). Comparison of pitch profiles of German and French speakers speaking French and German. *Proc. 18th International Congress of Phonetic Sciences (Glasgow).*

Zimmerer, Frank, Anne Bonneau and Bistra Andreeva (2016). Influence of L1 prominence on L2 production - French and German speakers. *Proc. Speech Prosody 2016, Boston*, 370–374.

Zimmerer, Frank, Jeanin Jügler, Bistra Andreeva, Bernd Möbius and Jürgen Trouvain (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. *Proc. Speech Prosody 2014, Dublin*, 1037–1041.

Zimmerer, Frank and Jürgen Trouvain (2015a). "Das Haus" or "das Aus"? - How French learners produced word-initial /h/ in German. Adrian Leemann, Marie-José Kolly, Stephan Schmid and Volker Dellwo (Editors), *Trends in Phonetics and Phonology. Studies from German-speaking Europe*, Frankfurt (Main)/Bern: Lang, 198–206.

Zimmerer, Frank and Jürgen Trouvain (2015b). Perception of French speakers' German vowels. *Proc. Interspeech 2015, Dresden*, 1720–1724.

Zimmerer, Frank and Jürgen Trouvain (2015c). Productions of /h/ in German: French vs. German speakers. *Proc. Interspeech 2015, Dresden*, 1922–1926.

Zimmerman, Barry J. and Dale. H. Schunk (2001). Reflections on theories of self-regulated learning and academic achievement. Barry J. Zimmerman and Dale. H. Schunk (Editors), *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, Mahwah, NJ: Lawrence Erlbaum Associates, 289–307.

Issues of PHONUS published between 1995 and 2017

B. Möbius & J. Trouvain (eds.) (2015). PHONUS 18: Fabian Brackhane: "*Kann was natürlicher, als Vox humana, klingen?" - Ein Beitrag zur Geschichte der mechanischen Sprachsynthese.* Saarbrücken: Insitute of Phonetics, Saarland University.

Barry, W.J., B. Möbius & J. Trouvain (eds.) (2014). PHONUS 17: Eva Lasarcyk: *Empirical evaluation of the articulatory synthesizer VocalTractLab as a discovery tool for phonetic research: Articulatory-acoustic investigations of paralinguistic speech phenomena.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Trouvain (eds.) (2011). PHONUS 16. *In memoriam Wolfgang von Kempelen.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. (ed.) (2009). PHONUS 15: Manfred Pützer: *Die Rolle kortikaler und subkortikaler Strukturen bei der Initiierung und Produktion differenzierter CV-Silbenwiederholungen. Eine fMRT-Studie.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Trouvain (eds.) (2008). PHONUS 14: Dominika Oliver: *Modelling Polish Intonation for Speech Synthesis.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. (ed.) (2008). PHONUS 13: Manfred Pützer: *Stimmqualität und Artikulation bei Dysarthrophonien.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. (ed.) (2007). PHONUS 12: Bistra Andreeva: *Zur Phonetik und Phonologie der Intonation der Sofioter-Varietät des Bulgarischen.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Trouvain (eds.) (2007). PHONUS 11: Roland Marti: *ó w dolnoserbšćinje (ó in Lower Sorbian / ó im Niedersorbischen).* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Trouvain (eds.) (2006). PHONUS 10: Caren Brinckmann: *Improving Prosody Prediction for Speech Synthesis – With and Without Symbolic Prosody Features.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Trouvain (eds.) (2005). PHONUS 9. Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. (ed.) (2004). PHONUS 8: Jürgen Trouvain: *Tempo Variation in Speech Production. Implications for Speech Synthesis.* Saarbrücken: Institute of Phonetics, Saarland University of the Saarland.

Barry, W.J. (ed.) (2004). PHONUS 7: Marc Schröder: *Speech and Emotion Research: An overview of research frameworks and a dimensional approach to* *emotional speech synthesis.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & M. Pützer (eds.) (2002). PHONUS 6. *Festschrift für Max Mangold zum 80. Geburtstag.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Koreman, with K. Kirchhoff (eds.) (2000). PHONUS 5. *Proceedings of the Workshop on Phonetics and Phonology in ASR. Parameters and Features, and their Implications.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Koreman (eds.) (1999). PHONUS 4. Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Koreman (eds.) (1997). PHONUS 3. Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & A. Addison (eds.) (1996). PHONUS 2. *Proceedings of the Workshop on Pronunciation Training at German Universities, Colleges of Education and Polytechnics.* Saarbrücken: Institute of Phonetics, Saarland University.

Barry, W.J. & J. Koreman (eds.) (1995). PHONUS 1. Saarbrücken: Institute of Phonetics, Saarland University.

Electronic versions (PDFs) are available online at:
http://www.coli.uni-saarland.de > Research Groups > Phonetics > Phonus