

Daniel Pape

**Microprosodic differences  
in a cross-linguistic vowel comparison  
of speech production and speech perception**

**Bibliografische Information Der Deutschen Bibliothek**

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Diese Publikation ist als Dissertation ausgewiesen: Humboldt-Universität zu Berlin, Philosophische Fakultät II, Institut für Deutsche Sprache und Linguistik im Jahre 2008 (Disputation am 23. Januar 2008).

Gedruckt auf holz- und säurefreiem Papier, 100 % chlorfrei gebleicht.

© Weißensee Verlag, Berlin 2009  
Simplonstraße 59, 10245 Berlin  
Tel. 030/29 04 91 92  
[mail@weissensee-verlag.de](mailto:mail@weissensee-verlag.de)  
[www.weissensee-verlag.de](http://www.weissensee-verlag.de)

Alle Rechte vorbehalten

Printed in Germany

ISBN 978-3-89998-173-5

# Table of content

<i>Abstract</i>	9
<i>Zusammenfassung</i>	13
<i>Structure and nomenclature</i>	19
<i>General introduction</i>	21
<b>Part 1: Speech Production</b>	<b>25</b>
<b>1 Introduction</b>	<b>25</b>
<b>2 Literature</b>	<b>27</b>
2.1 Passive movement hypotheses – Coupling of articulatory structures	28
2.1.1 Universality	28
2.1.2 Influencing effects	29
2.1.3 Responsible mechanisms	30
2.1.4 Discrepancies	31
2.2 Intentional movement theory – Perceptual enhancement, active control	32
2.2.1 Compensation for psychoacoustic phenomena	33
2.2.2 Enhancement of vowel contrasts	33
2.2.3 Compensation for intensity differences	35
2.3 The special case: IF0 comparing tense and lax vowels	36
2.4 Exceptions: Non-Existence of IF0	38
2.5 IF0 in the singing voice	38
2.6 Summary of the main theories	39
2.6.1 Automatic consequence of vowel production	39
2.6.2 Enhancement of vowel contrasts	40
<b>3 Production experiment: Intrinsic F0 comparing German tense and lax vowels – EMMA, EGG and acoustic data</b>	<b>42</b>
3.1 Introduction to the experiment	42
3.2 Methods	44
3.2.1 Recording and technical parameters	44
3.2.2 Pre-processing	46
3.2.3 Speech corpus and acoustical segmentation	48
3.2.4 Statistic analyses and tools	50
3.2.5 Summary of the methods	53
3.3 Results and Discussion	54
3.3.1 Landmarks onset, mid and offset of the vowel	54
3.3.2 Functional Data Analyses (FDA) of the complete curve sets	60
3.4 Conclusions	67

<b>Part 2: Speech Perception</b>	<b>71</b>
<b>1 Introduction</b>	<b>71</b>
<b>2 Literature</b>	<b>74</b>
2.1 Perception of the source pitch (non-speech signals)	74
2.1.1 Pitch perception theories	75
2.1.2 Pitch perception accuracy of complex sounds	76
2.2 Biasing factors for the pitch perception of the source signal	76
2.2.1 Musical education	77
2.2.2 Pitch memory	78
2.2.3 Temporal aspects (duration of the stimulus)	79
2.3 Influence of the spectrum on the pitch perception of the source pitch	79
2.3.1 Timbre measurement for non-speech signals	80
2.3.2 Spectral influence on source pitch: non-speech	81
2.3.3 Spectral influence on source pitch: speech	83
2.4 Neuroscience evidence for differences in pitch processing	95
2.4.1 Differences in the processing of musical and prosodic pitch	96
2.4.2 Neural landmarks for prosody identification	98
2.4.3 Musicians versus non-musicians	99
<b>3 Perception experiment 1: Intrinsic pitch differences for vowels differing in tenseness and roundedness</b>	<b>100</b>
3.1 Introduction and aims of the experiment	100
3.2 Methods	102
3.2.1 Stimulus preparation	102
3.2.2 Acoustic properties of the stimuli	102
3.2.3 Listeners	106
3.2.4 Psychophysical method	107
3.2.5 Equipment and Procedure	108
3.2.6 Dependence on other factors: Loudness adjustment	109
3.2.7 Statistics	109
3.3 Results and Discussion	112
3.3.1 F0 differences between vowels for equal pitch impression	112
3.3.2 Accuracy of the response functions with respect to F0 changes	114
3.3.3 Effect of musical education	115
3.3.4 General intrinsic pitch sensitivity: German vs. Catalan	116
3.4 Conclusion and Outlook	118
<b>4 Perception experiment 2: Cross-linguistic intrinsic pitch perception for tenseness and vowel height differences</b>	<b>121</b>
4.1 Introduction	121
4.1.1 Aims of the experiment – research hypotheses	121

4.1.1	Structure of the experiment	123
4.2	Methods	124
4.2.1	The stimuli	124
4.2.2	Further processing of the stimuli (speech and music)	131
4.2.3	The procedure	131
4.2.4	The listeners	134
4.3	Results and Discussion	137
4.3.1	Threshold differences for musical tone and vowel	137
4.3.2	Intrinsic pitch for the vowel height contrast / i a /	143
4.3.3	Intrinsic pitch for the tenseness contrast / i I /	153
4.3.4	Outlook: Systematical description of the subjects' response pattern	155
<b>General conclusion</b>		<b>157</b>
<b>1</b>	<b>Linking speech production and perception</b>	<b>157</b>
1.1	Tenseness differences	157
1.2	Intrinsic F0 versus intrinsic pitch in German	158
<b>2</b>	<b>Cross-linguistic perception differences</b>	<b>160</b>
<b>3</b>	<b>Perception differences due to musical education</b>	<b>162</b>
<b>4</b>	<b>Music versus language mode for pitch processing</b>	<b>164</b>
<b>5</b>	<b>Classification of the perceptual results with regard to intrinsic pitch theories</b>	<b>165</b>
<b>6</b>	<b>Critical remarks/Evaluation</b>	<b>166</b>
<b>Glossary</b>		<b>179</b>
<b>Appendices</b>		<b>185</b>
<b>1</b>	<b>Functional data analysis: Explanation and Methods</b>	<b>185</b>
<b>2</b>	<b>Pitch perception theories – complete review and evaluation</b>	<b>190</b>
<b>3</b>	<b>Loudness measurement for perception experiment 2</b>	<b>194</b>
<b>4</b>	<b>Instructions for the listeners of the perception experiment 1 - German listeners</b>	<b>197</b>
<b>5</b>	<b>Instructions for the listeners of the perception experiment 2</b>	<b>198</b>
<b>6</b>	<b>Numerical results of the perception experiment 2 for all listeners and all languages</b>	<b>201</b>

---

# Abstract

The PhD thesis deals with the examination of both speech production and perception, with special regard to the phenomena of “intrinsic fundamental frequency (IF0) – intrinsic pitch (IP)”. With regard to speech production, IF0 describes the dependence of the fundamental frequency on vowel height: Other things being equal, close vowels show a higher F0 compared to open vowels. With regard to speech perception, intrinsic pitch describes the dependence of the perceived vowel pitch on the height of the presented vowel: Close vowels are perceived with a lower pitch impression compared to open vowels. The main aim of the thesis is to investigate similarities and differences by conducting a combined examination of the phenomenon “IF0-IP”, comparing tenseness differences in German on the one hand and cross-linguistic differences on the other hand.

## Speech Production

Reasons for production IF0 differences are discussed controversially in the literature: On the one hand researchers argue that IF0 is an automatic consequence of the connection between the tongue and the laryngeal structures, with the effect of increasing F0 when articulating high vowels. In contrast, enhancement theories claim an increase in the robustness of the vowel identification: IF0 differences may serve as a vowel quality cue to robustly enhance the discriminability between different vowels and/or vowel classes. Strong evidence is presented in defence of both theories, but the majority of researchers claim the automatic consequence theory to be more reliable. However, challenging for the automatic consequence theory are the results for contrasting tense and lax vowels in English, Danish and German: Here it was found that these different vowel classes show a similar acoustic F0, although their articulatory vowel height is significantly different. However, according to the automatic consequence theory, these vowels should show a differing F0 due to the different tongue height.

Thus, to examine this discrepancy is the aim of the speech production part of the thesis. Articulatory and acoustic properties of tense vowels in comparison to lax vowels are examined, with special regard to the temporal behaviour of both vowel systems. For this reason, combined acoustic and articulatory data were recorded by means of synchronized audio, EGG (Electro-Glottography) and EMMA (Electromagnetic Midsagittal Articulography). The corpus consisted of the German tense and lax vowels / i u a/ in nasal context and was provided by four native German speakers.

The results for the examination of the acoustic F0 at vowel onset, mid, and offset gave evidence that indeed no significant difference is found comparing the tense and lax vowel pairs. In contrast, the articulatory positions of the tongue height sensors significantly differed with respect to tenseness. Accordingly, correlation and regression analyses were carried out to obtain an objective measure for the tenseness differences: Here it was found that the link between F0 and articulatory

positions remained weak and non-significant. Hence, the articulatory data cannot explain the discrepancy between F0 and tongue height for the tenseness contrast in German. It is concluded that the missing link between F0 and tongue height persists with regard to tenseness, so other mechanisms seem to be at work here.

Since tense and lax vowels show a different temporal pattern with respect to F0, the question arose whether the non-existing link between F0 and articulatory positions is a by-product of this difference in temporal alignment. Therefore, to examine whether the inclusion of the complete time dimension would reveal interesting results, the three-dimensional set of vowel contours was examined with Functional Data Analysis (FDA). This statistical method takes into account the time series as a separate dimension and is used in this thesis to reveal differences in the temporal structure. It was found that these differences were not significant with regard to tenseness. Independent of the IF0 timing, the factors F0 and tongue height explain most of the variance within the F0 and articulatory dataset. Furthermore, these factors are highly correlated. It has to be concluded that for tenseness the differences in timing are not important. Thus, the classical analyses of fixed vowel landmarks seem completely sufficient.

To conclude, no biomechanical reasons explaining the discrepancy between German tense and lax vowels could be found, so other causes to raise the F0 of lax vowels seem plausible. Hoole et al. (2006) could show that additional enhancement mechanisms seem to take place. These actively raise the F0 of lax vowels in comparison to tense vowels. However, the reasons for this active increase are not yet clear; it could be hypothesized that they serve perceptual enhancement, since productional links were not able to explain the phenomenon.

### **Speech Perception**

The literature argues in favour of contrasting streams: It is argued that intrinsic pitch either (1) serves as a compensation mechanism for the existing IF0 differences in speech production or (2) is caused by psychoacoustic phenomena of the auditory system or (3) is used to actively enhance or change the vowel quality impression for the listener. All these streams are faced with the drawback that empirical evidence is presented for German and English - thus only concentrating on Germanic languages.

For the perception part of this thesis, two research aims are defined: (1) to examine possible perceptual consequences of the found tenseness IF0 discrepancy in speech production, and (2) to study intrinsic pitch for languages other than Germanic ones. Since Germanic languages show a more crowded vowel system, it is argued that in these languages a special intrinsic pitch effect might take place. However, this possibility has not been examined until now, thus evidence is still needed to discern whether intrinsic pitch differences exist at all in other than Germanic languages. To pursue these research questions, two subsequent perceptual experiments are conducted.

The first experiment was conducted with German subjects. It examined intrinsic pitch for three isolated German vowels / i:ɪy: ɪɪ/ with equal phonological vowel height, but tenseness and roundedness differences. The vowels showed a rising F0

and varying vowel duration. To examine whether the intrinsic pitch perception is influenced by the native language inventory, an identical experiment was also cross-linguistically carried out with Catalan listeners in Spain. The results give evidence that (1) the tenseness difference /i: i/ is not significant for the examined conditions, neither for German nor Catalan subjects. In contrast, (2) differences in roundedness /i: y/ significantly affected intrinsic pitch for both languages. Surprisingly, (3) the (screened) factor musical education influenced listener sensitivity only for the German listeners. That means that Catalan listeners, independent of their musical education, are found to be insensitive to even large F0 differences, and judged the vowel pitch differences mainly based on vowel quality: For example, they judged /i:/ always to be higher than /i:/, independent of the between-vowel F0 difference. Thus, for the examined conditions it is given evidence that intrinsic pitch and thus pitch discrimination is language-dependent, and therefore not as universal as seen in the classical literature. It seems that the categorical behaviour of the Catalan listeners signifies that intrinsic pitch is either absent or unexpected high, dependent on the definition of intrinsic pitch.

To examine this effect further, a second experiment was designed as a cross-linguistically tenable study. The main research aims of the second experiment are (1) to study the occurrence of intrinsic pitch when comparing German with non-Germanic languages and to (2) eliminate problems in the stimulus design of the first experiment. Since the first experiment showed clear differences for varying musical education, the second study is examining two contrasting populations: professional musicians as one extreme population and listeners without any musical education (non-musicians) as the other extreme population, with the population of amateur musicians in between. The test is comparably conducted with German listeners on the one hand and Italian, Portuguese, and Spanish listeners on the other hand. The latter languages are chosen with the aim of examining intrinsic pitch differences for the Romance language family.

In a pre-experiment, the subjects' difference threshold (dl) is screened for pitch discrimination tasks of musical and speech stimuli. The reason was to obtain knowledge about the listeners' ability to successfully manage the following tasks. When the tests prove the subject's suitability, two intrinsic pitch experiments are carried out. The first experimental part tests for German vowel pairs differing in height /i: a:/ and tenseness /i: i/. The second part examines vowel pairs for the listeners' native language (thus e.g. Spanish vowel stimuli for Spanish subjects). Results for the first part of the experiment (dl) indicate that strong cross-linguistic and inter-population differences for both the musical tone and vowel: First, the threshold values for these two conditions decrease with increasing musical education – which you would expect according to the psychoacoustic literature. Second, for the German listeners a lower threshold compared to the Romance languages is found - with the largest differences between the languages German and Italian. Third, comparing the musical tone and vowel, it is given evidence that the musical tone evokes lower thresholds than does the vowel. This result is consistent for all examined populations and languages in this experiment. Thus, it is shown that the processing of musical and vowel pitch significantly differs.

For the intrinsic pitch parts, the main finding is that the number of listeners insensitive to the given pitch task is significantly higher for all Romance

languages than for German. Spanish and Portuguese have the highest amount with nearly 50% of the non-musicians showing *pitch-insensitivity*. Italian holds an intermediate position compared to a low 6% in German. This gives clear evidence that intrinsic pitch has to be categorized as language-dependent. The insensitivity found is independent of the stimulus identity, i.e. also the native stimulus condition did not increase sensitivity. Thus, here the categorical behaviour already seen in experiment 1 is substantiated, since for nearly half of the Romance language listeners intrinsic pitch differences cannot be shown in the examined F0 region.

When analyzing the valid intrinsic pitch values for the remaining *pitch-sensitive* listeners, the following results are obtained: Only when pooled over all German listeners, for the vowel height comparison / i: α:[]/ it is found that the intrinsic pitch value is significant, although with a lower value than one would find in the classical literature. However, when separately examining the different populations, it is seen that no statistical difference could be extracted, neither for the professional musicians nor non-musicians. Thus, intrinsic pitch is generally present in German, but the significance of the phenomenon is missed when examining populations contrasting in the important factor musical education. In the classical studies this difference is not found since the musical education was not seen as important for the intrinsic pitch tasks.

With regard to cross-linguistic differences, it is found that for all languages intrinsic pitch was absent for the group of professional musicians. When comparing differences in musical education, it could be shown that across all languages the intrinsic pitch values significantly differ when comparing professional musicians with non-musicians. Thus, important evidence is given here that increased musical education significantly influences the occurrence and distribution of intrinsic pitch. When examining only non-musicians, inconsistent patterns are found, partly because for Spanish and Portuguese the small number of valid (pitch-sensitive) listeners limits the strength of interpretation. Since Italian has a larger number of valid pitch-sensitive listeners, the results are more interpretable: Here, intrinsic pitch is present for the native vowel condition. Thus, Italian presents a special case: It seems that intrinsic pitch is present in this language, but only for the population of non-musicians and the native vowels. In summary, for the vowel height dimension it is given evidence that intrinsic pitch definitely has to be classified in a language specific manner: For the Romance languages Spanish and Portuguese, intrinsic pitch was mostly not verifiable, whereas the production counterpart IF0 is shown to be present with typical values.

Concerning the tenseness distinction / i i[]/, significant intrinsic pitch differences can be stated for German non-musicians, but not for the professional musicians. Thus, in contrast to the first experiment intrinsic pitch differences are in fact found. Since the results for the musically educated listeners are non-significant, it is assumed that non-musicians rather judge linguistic pitch differences, whereas musicians concentrate on judging the signal-present pitch. In future experiments, the actual cause for the tenseness discrepancy of the results of experiments has to be further examined. Furthermore, future experiments have to clarify whether the subjects' vowel identification would benefit from F0 differences, which would be the apparent consequence of the intrinsic pitch differences found.

# Zusammenfassung

Diese Dissertation untersucht die Interaktion zwischen Sprachproduktion und Sprachwahrnehmung mit speziellem Augenmerk auf die Phänomene *intrinsische Grundfrequenz* (IF0) und *intrinsische Tonhöhe* (IP). In der Sprachproduktion beschreibt die IF0 die Abhängigkeit der Grundfrequenz von der Vokalhöhe, d.h. geschlossene Vokale wie / i: u: / weisen eine höhere Grundfrequenz auf als offene Vokale wie / a:/, wenn alle anderen Faktoren konstant gehalten werden. In der Sprachwahrnehmung definiert *intrinsische Tonhöhe* die Abhängigkeit der wahrgenommenen Tonhöhe von der Vokalhöhe des dargebotenen Vokals, d.h. / a:/ wird höher wahrgenommen als / i: u:/, sofern sie mit gleicher Grundfrequenz präsentiert werden.

Das Hauptziel dieser Dissertation ist die Untersuchung von Gemeinsamkeiten und Unterschieden von kombinierten IF0/IP-Experimenten. Dazu werden auf der einen Seite Unterschiede in deutschen gespannt-ungespannt Vokalpaaren und auf der anderen Seite Unterschiede zwischen verschiedenen Sprachen untersucht.

## Sprachproduktion

Die Gründe für intrinsische Grundfrequenzunterschiede werden in der Literatur kontrovers diskutiert. Auf der einen Seite geht man davon aus, dass diese Unterschiede eine automatische Konsequenz aus der existierenden biomechanischen Verbindung von Zunge und Larynx ist, welche sich dann in einer Grundfrequenzerhöhung bei der Produktion von hohen Vokalen äußert (biomechanische Theorien). Andererseits gehen die *Verstärkungstheorien* davon aus, dass die Grundfrequenzunterschiede aktiv erzeugt werden, um eine bessere Diskriminierung zwischen verschiedenen Vokalen und/oder Vokalklassen zu ermöglichen. Beweise für die Validität der jeweiligen Richtungen gibt es auf beiden Seiten, wobei meist die biomechanischen Theorien bevorzugt werden. Eine Herausforderung für diese ergibt sich allerdings für die Gespanntheitsopposition des Deutschen: Hier wurde gezeigt, dass gespannte und ungespannte Vokalpaare eine ähnliche Grundfrequenz, jedoch unterschiedliche artikulatorische Zungenhöhenpositionen aufweisen. Damit werden die biomechanischen Theorien in Frage gestellt, da aufgrund dieser die Differenz in der Zungenhöhe auch eine unterschiedliche intrinsische F0 fordert.

Das Ziel des Produktionsteils definiert sich über die Untersuchung der Gründe für die Diskrepanz des beschriebenen Gespanntheitsphänomens. Aus diesem Grund werden kombinierte akustische und artikulatorische Daten mithilfe von synchronen akustischen, elektroglottographischen (EGG) und artikulatorischen (EMMA) Aufnahmeverfahren erhoben. Das Korpus besteht aus deutschen gespannten und ungespannten Vokalen / [i u a] /, welche von vier männlichen Versuchspersonen in nasalem Kontext gesprochen wurden.

Die Ergebnisse für die Untersuchung der intrinsischen F0 an Vokalbeginn, Mitte und Ende zeigen, dass keine signifikanten Unterschiede zwischen den jeweiligen gespannten und ungespannten Vokalpaaren zu finden sind, während sich die artikulatorischen Positionen signifikant unterscheiden. Daraufhin wurden Korrelationen und Regressionen für unterschiedliche Vokalkonfigurationen berechnet, um die dargestellten Unterschiede zu quantifizieren. Wie aufgrund der

## Zusammenfassung

biomechanischen Theorie vorhergesagt sind die Korrelationen für den Vergleich zwischen hohen und tiefen Vokalen stark signifikant. Dagegen zeigen sich die Korrelationen für den Vergleich zwischen gespannten und ungespannten Vokalpaaren schwach und nicht signifikant, was auf die fehlende biomechanische Verbindung bezogen auf die Gespanntheitsopposition hindeutet.

Da gespannte und ungespannte Vokale eine unterschiedliche zeitliche Entwicklung der Grundfrequenzkontur aufweisen, sollte im folgenden die Frage untersucht werden, inwieweit der zeitabhängige Verlauf die Unterschiede hervorruft. Aus diesem Grund wurden die kompletten zeitabhängigen akustischen und artikulatorischen Datensätze einer *functional data analysis* unterzogen. Diese statistische Methode bezieht die zeitliche Achse als eigenständige Dimension ein, um so Unterschiede in der temporären Struktur zu suchen. Jedoch konnten als Resultat nur die *zeitunabhängige* Grundfrequenzhöhe (F0-Konturen) und Zungenhöhe (artikulatorische Verläufe) als einzige Faktoren extrahiert werden, und erklärten damit den Hauptteil der gefundenen Varianz. Die beiden Hauptfaktoren der F0 und Artikulation sind miteinander hoch und signifikant korreliert, was die biomechanischen Theorien stützt. Bezogen auf die Gespanntheit waren jedoch keine gekoppelten Unterschiede und keine zeitabhängige Variable nachweisbar. Daher kann in weiteren Analysen in der Tat die traditionelle Statistik von einzelnen Zeitpunkten benutzt werden, ohne die Gefahr unzulässiger Datenreduktion und damit des Verlusts von wichtiger Information.

Zusammenfassend kann gesagt werden, dass keine biomechanischen Gründe für die Diskrepanz der Gespanntheit in Bezug auf IF0 gefunden wurden. Es scheint, dass zusätzliche Verstärkungsmechanismen in Erscheinung treten, um die Grundfrequenz von ungespannten Vokalen additiv zu erhöhen. Hoole et al. (2006) konnte anhand der Analysen von Muskelaktivitäten zeigen, dass eine additive Erhöhung der Grundfrequenz von ungespannten gegenüber gespannten Vokalen im Deutschen auftritt. Jedoch sind die genauen Gründe für eine solche aktive Erhöhung nicht klar. Es kann nun jedoch davon ausgegangen werden, dass sie für eine perzeptive Verstärkung benutzt werden, da wie beschrieben für die Gespanntheitsunterschiede eine Lösung der Diskrepanz von F0 und artikulatorischen Strukturen in der Sprachproduktion nicht aufgezeigt werden konnte.

### Sprachwahrnehmung

Die Sprachwahrnehmung betreffend gibt es in der Literatur drei Hauptströmungen: In der ersten wird davon ausgegangen, dass *intrinsic pitch* als Kompensation für die intrinsischen Grundfrequenzunterschiede dient. In der zweiten beruht *intrinsic pitch* auf den generellen Eigenschaften des auditiven Apparates. In der dritten wird *intrinsic pitch* benutzt, um eine aktive Verstärkung oder Veränderung der Vokalqualität zu ermöglichen. Alle diese Theorien haben das Problem, dass die sie stützenden empirischen Beweise nur für germanische Sprachen nachgewiesen wurden (Deutsch oder Englisch).

Damit definieren sich die folgenden Ziele für den Wahrnehmungsteil: Erstens werden die im Produktionsteil beschriebenen Ergebnisse in Bezug auf die Verbindung zur Wahrnehmung untersucht. Zweitens soll erstmalig die intrinsische Tonhöhe für nichtgermanische Sprachen untersucht werden: Da germanische Sprachen eine höhere Anzahl von Vokalen aufweisen als viele

andere Sprachen, kann davon ausgegangen werden, dass in germanischen Sprachen intrinsische Tonhöhe einen anderen Stellenwert hat als in anderen Sprachen. Diese Möglichkeit wurde bisher noch nicht untersucht, es wird im Gegenteil IP als universell angesehen. Daher muss erst gezeigt werden, ob und in welchem Umfang intrinsische Tonhöhe in anderen Sprachen existent ist. Zur Untersuchung der dargestellten Forschungsthemen wurden zwei aufeinander aufbauende Experimente durchgeführt.

Das erste Experiment untersucht die Vokaltonhöhenunterschiede für deutsche Hörer für die Vokale / i y i̯/. Diese weisen die gleiche phonologische Vokalhöhe auf, unterscheiden sich aber in ihrer Gespanntheit beziehungsweise Rundung. Für die Parametrisierung der Vokale wurden eine steigende Grundfrequenzkontur und die natürlichen Längenunterschiede benutzt. Um herauszufinden, ob die IP-Wahrnehmung von der jeweiligen Sprache des Hörers abhängt, wurde das gleiche Experiment im Sprachvergleich mit katalanischen Hörern in Spanien durchgeführt. Die Ergebnisse zeigen, dass der Gespanntheitsunterschied / i i̯/ in keiner der beiden Sprachen statistisch signifikant ist. Im Gegensatz dazu war der Unterschied zwischen gerundetem und ungerundetem gespannten Vokal (/ i:̯/ versus / y:̯ /) in beiden Sprachen hochsignifikant. Überraschenderweise beeinflusste die Variable *musikalische Bildung* die Tonhöhenwahrnehmung nur für die deutschen Hörer. Die katalanischen Hörer beurteilten die Vokale eher bezogen auf die Vokalqualität - unabhängig von der musikalischen Bildung. Beispielsweise hörten sie den Vokal / i:̯/ immer höher als / i̯/, unabhängig vom tatsächlichen Grundfrequenzunterschied zwischen beiden Vokalen. Die Ergebnisse des ersten Experimentes deuten also auf eine Sprachabhängigkeit von IP hin. Das zeigt, dass IP nicht so universell ist wie bisher angenommen. Die Ergebnisse für die Katalanen deuten an, dass IP hier entweder nicht vorhanden oder aber unerwartet hoch ist, abhängig von der Betrachtungsweise der IP Definition.

Zur Untersuchung dieser Hypothese wurde ein zweites, crosslinguistisch valides Wahrnehmungsexperiment entwickelt. Die zu untersuchenden Hauptziele waren erstens die Untersuchung von IP im Vergleich von Deutsch und nichtgermanischen Sprachen und zweitens die Eliminierung der möglichen Probleme im Stimulusdesign des vorherigen Experiments. Da dieses einen klaren Einfluss von musikalischer Ausbildung zeigte, wurden jetzt zwei gegensätzliche Populationen untersucht: professionelle Musiker und Nichtmusiker. Das Experiment wurde einerseits für deutsche Hörer und andererseits für italienische, portugiesische und spanische Hörer durchgeführt. Die letzteren wurden mit dem Ziel der vergleichenden Untersuchung von IP für die romanische Sprachfamilie ausgewählt.

Um festzustellen, ob die Hörer überhaupt für die nachfolgenden IP-Studien geeignet sind, wurden in einem Vorexperiment die Unterschiedsschwellen (dl) der Tonhöhe für musikalische Töne und Vokale bestimmt. Die nachfolgenden IP Experimente bestanden aus Vokalpaaren, die sich einerseits in der Vokalhöhe (/ i:̯ α:̯/) und andererseits in der Gespanntheit (/ i:̯ i̯/) maximal unterschieden. Der erste Teil des Experiments verwendete deutsche Stimuli, während der zweite Teil Stimuli der jeweiligen nativen Sprache des Hörers benutzte (d.h. beispielsweise italienische Vokale für die italienischen Hörer).

## Zusammenfassung

Die Ergebnisse des Schwellenexperimentes (dl) zeigen, dass starke cross-linguistische und Gruppenunterschiede sowohl für den musikalischen Ton als auch für den Vokal zu finden sind: Erstens werden die Schwellenwerte geringer mit steigender musikalischer Ausbildung, was aufgrund vorheriger Studien auch zu erwarten ist. Zweitens sind die Werte für die deutschen Hörer generell geringer als für die Hörer der romanischen Sprachen, der höchste Unterschied tritt zum Italienischen auf. Drittens kann im Vergleich von musikalischem Ton und Vokal gezeigt werden, dass der musikalische Ton eine geringere Schwelle evoziert als der Vokalstimulus. Dies gilt für alle Sprachen und untersuchten Gruppen. Daher kann davon ausgegangen werden, dass Unterschiede in der Verarbeitung von musikalischer und Vokaltonhöhe auftreten, was sich dann in den unterschiedlichen Schwellwerten widerspiegelt.

Für die intrinsischen Tonhöhenexperimente zeigen sich folgende Hauptergebnisse: Der Prozentsatz der Hörer, die unempfindlich für die präsentierten Grundfrequenzunterschiede sind, ist in den romanischen Sprachen um ein Vielfaches höher als im Deutschen: Nahezu die Hälfte aller spanischen und portugiesischen Nichtmusiker zeigt diese Unempfindlichkeit, d.h. einer der Vokale wird *immer* höher wahrgenommen als der andere Vokal, auch wenn der andere mit der maximal höheren Grundfrequenz dargeboten wird. Im Gegensatz dazu zeigen nur 6% aller deutschen Hörer diese Eigenschaft, während das Italienische eine Zwischenposition einnimmt. Dieses Ergebnis zeigt, dass die intrinsische Tonhöhe klar als ein sprachabhängiges Phänomen kategorisiert werden muss. Weiterhin wird in den Ergebnissen deutlich, dass die gefundene Unempfindlichkeit nicht abhängig vom Sprachmodus ist, d.h. die nativen Stimuli evozierten keine höhere Tonhöhenempfindlichkeit als die deutschen Stimuli. Die eher kategorische Verhaltensweise der katalanischen Hörer des ersten Experimentes konnte für alle romanischsprachigen Hörer des zweiten Experiments bestätigt werden. Demnach ist für die Mehrzahl der romanischen Hörer IP also nicht nachweisbar.

Für die verbleibenden tonhöhenempfindlichen Hörer kann folgendes gezeigt werden: Über alle Gruppen gerechnet, konnten für die deutschen Hörer ähnlich signifikante Unterschiede wie in der Literatur extrahiert werden, wenn auch mit einem geringeren IP-Wert. Wurden jedoch die Gruppen einzeln analysiert, dann waren die Ergebnisse nicht signifikant, weder für die professionellen Musiker noch für die Nichtmusiker. Daher kann davon ausgegangen werden, dass intrinsische Tonhöhenunterschiede im Deutschen zwar vorhanden sind, das Phänomen aber eher schwach ausgeprägt ist. Das wichtige Ergebnis der Nichtsignifikanz in den Gruppen wurde in den klassischen Studien jedoch nicht gefunden, da der Faktor musikalische Ausbildung nicht erhoben wurde.

In Bezug auf cross-linguistische Unterschiede konnte gezeigt werden, dass für die professionellen Musiker in keiner Sprache signifikante Unterschiede nachweisbar sind. Nur wenn die Ergebnisse von professionellen Musikern mit denen von Nichtmusikern verglichen werden, zeigen sich Signifikanzen: Hier kann gezeigt werden, dass eine höhere musikalischen Ausbildung auch in einer höheren Empfindlichkeit und stabileren Verteilung von IP Unterschieden mündet.

Wird nur die Population der Nichtmusiker betrachtet, dann kann für das Spanische und Portugiesische aufgrund des kleinen Prozentsatzes tonhöhenempfindlicher Hörer keine Aussage getroffen werden. Im Gegensatz dazu sind für das Italienische die Ergebnisse klarer interpretierbar: Hier können intrinsische Tonhöhenunterschiede für die nativen Vokale nachgewiesen werden. Damit tritt

für das Italienische eine Besonderheit auf: Intrinsische Tonhöhenunterschiede existieren, aber nur für Nichtmusiker und nur für den Test auf native Vokale. Zusammenfassend kann gesagt werden, dass als Hauptergebnis eine sprachspezifische Einordnung von IP zu fordern ist, im Gegensatz zur bisher vertretenen universellen Sichtweise. Weiterhin konnten für das Spanische und Portugiesische keine perzeptiven IP-Unterschiede gezeigt werden. Da aber produktive F0-Unterschiede nachgewiesen wurden, müssen die klassischen Kompensationstheorien auch überarbeitet oder zumindest um eine sprachspezifische Komponente erweitert werden, da sich Produktion und Perzeption nicht kompensatorisch verhalten.

Für die Untersuchung des Gespanntheitsunterschiedes / i i̯ / konnte im zweiten Experiment gezeigt werden, dass signifikante Unterschiede im Deutschen auftreten, allerdings nur für die Nichtmusiker. Somit sind in diesem Experiment Unterschiede nachweisbar, die sich in der ersten Studie nicht zeigten. Sollten die nichtsignifikanten Ergebnisse tatsächlich aufgrund der ungenügenden Parameter des ersten Experiments entstehen, eignet sich folgende Hypothese zur Erklärung: Da die Ergebnisse für alle musikalisch ausgebildeten Versuchspersonen nichtsignifikant sind, kann man davon ausgehen, dass Nichtmusiker eher die linguistische Vokaltonhöhe beurteilen, während sich Musiker auf die (akustische) signalpräsente Tonhöhe verlassen.

In weiteren Experimenten ist zu prüfen, ob sich die Vokalidentifikation tatsächlich aufgrund der nachgewiesenen intrinsischen Tonhöhenunterschiede verbessert, was die Ergebnisse des zweiten Experimentes und die Ergebnisse von Hoole et al. nahelegen.

---

# Structure and nomenclature

This thesis examines the speech research areas production, perception, and the interaction of production and perception. According to these areas, the work is divided into a production and a perception part. Specifically, both areas examine the phenomena of microprosodic differences. In the following, the structure is presented:

The general introduction gives a general overview and generally explains the speech phenomena to be examined.

In the **speech production part**, intrinsic F0 differences in German for both tense/lax and high/low vowels are examined. This part consists of the chapters *Introduction*, *Literature review*, *Results*, and *Discussion-Conclusion*. In these chapters a production experiment is presented that examines intrinsic F0 differences for an EMMA/EGG experiment with German speakers.

Based on the results of the production part, the **speech perception part** studies the perception of intrinsic pitch for German tense/lax vowels. Further, by contrasting different languages the cross-linguistic presence and importance of intrinsic pitch is examined, both for the tenseness and vowel height dimension. To examine these questions, two successive perceptual experiments are conducted: Experiment 1 examines the described differences for vowels in natural-speech conditions. Experiment 2, carried out for different languages, uses both F0 and time normalized stimuli to examine general trends for intrinsic pitch.

The structure of the perception part follows the classical composition: Following the *Introduction* and *Literature review* **the first perceptual experiment** is presented in the order *Introduction*, *Methods*, *Results* and *Discussion/Conclusion*. Subsequently, **the second experiment** resumes the discussion and extends the hypotheses from the first experiment: Succeeding the *introduction* and *methods* chapter, the languages German, Italian, Spanish, and Portuguese are used to set up a systematic cross-linguistic comparison. Subsequently, the obtained results are *discussed*.

Subsequently, in the **general discussion** the obtained results of the production and perception part are resumed and discussed. The main focus lies here on tenseness differences for German and cross-linguistic differences. Finally, possible critical remarks concerning the design of the perceptual studies and methods are discussed.

The **references** give the cited literature. Internet sources used in this work are given in the text footnotes.

The **glossary** presents a short description of the main technical items and terms. Please note that all terms are explained during the first use in the main text, but for fluent reading and fast cross-referencing they are also given in the glossary.

The **appendices** present tables and further explanations important for the understanding of the results of the thesis (but not felt important to be included in the main text). Also, the appendices include the complete tables of the perceptual experiments, additional explanations of algorithms, and analysis scripts.