

# Hörstrom Segregation von Vokalen basierend auf konkurrierenden F0 und Timbre Merkmalen

Verena Kies<sup>1</sup>, Marion David<sup>1</sup>, Steven van de Par<sup>1</sup>

<sup>1</sup> Medizinische Physik und Akustik, Exzellenzcluster 'Hearing4all', Universität Oldenburg

## Einleitung

Die sequentielle Segregation wurde bereits in einigen Studien untersucht, dabei wurden jedoch meist Töne oder komplexe Töne verwendet und als exakte Wiederholungen voneinander präsentiert (beispielsweise Singh in 1987). Im Gegensatz zu solchen Stimuli verändern sich die Eigenschaften von natürlichen Geräuschen mit der Zeit. Sprache beispielsweise weist eine natürliche spektral-zeitliche Variabilität auf und ist nicht vollständig vorhersagbar. Zwei der wenigen Studien, in denen die sequentielle Gruppierung von Sprache untersucht wurde, sind von Gaudrain et al. (2007) und von Nootboom et al. (1978). Aus diesen Studien resultierte, dass F0 trotz der spektral-zeitlichen Variabilität zwischen den Vokalen ein robustes Merkmal für eine Vokal-Segregation ist. Darüber hinaus fanden Clarke et al. (2014) heraus, dass bei einer Veränderung der Vokaltraktlänge (VTL) zwei Sprecher wahrgenommen werden können.

## Begründung

Timbre und F0 sind Eigenschaften, welche eine Stimme einzigartig machen. Da der Einfluss des Timbres allein und in Kombination mit der F0 auf die Segregation von Sprache aber noch wenig erforscht ist, soll in diesem Experiment untersucht werden, wie F0 und Timbre die Wahrnehmungsorganisation von Vokalen beeinflussen. Darüber hinaus wird auch untersucht, ob eines der beiden Merkmale dominanter ist und ob es einen Punkt gibt, an dem sowohl F0 als auch Timbre die Segregation in ähnlicher Weise beeinflussen. Des Weiteren soll herausgefunden werden, ob es sich auf die sequentielle Segregation auswirkt, wenn sich die Probanden speziell auf ein Merkmal (F0 oder Timbre) konzentrieren. Um diese Forschungsfragen zu beantworten, wird die Segregation von Vokalen, bei denen sowohl F0 als auch die mit dem Timbre verbundenen spektralen Unterschiede variieren, untersucht. Um die Bildung verschiedener Hörströme durch Vokalsequenzen herbeizuführen, wird eine objektive Methode verwendet.

## Methoden

### Probanden

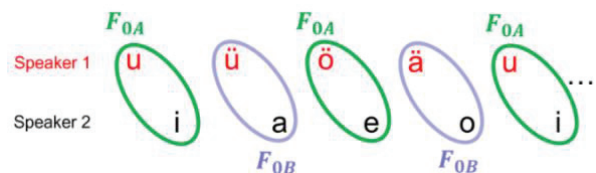
An diesem Experiment nahmen zwölf Probanden teil, sechs Männer und sechs Frauen. Ihr Alter lag zwischen 20 und 29 Jahren (Median = 26,3) und sie waren alle deutsche Muttersprachler. Keiner von ihnen hatte zuvor an ähnlichen Experimenten teilgenommen und alle waren normalhörend. Sie erhielten eine Vergütung von 10 Euro pro Stunde. Die Probanden nahmen an vier Terminen à zwei Stunden teil.

### Stimuli

Die Sequenzen bestehen aus Vokalen, die von zwei verschiedenen weiblichen Sprechern aufgenommen wurden und die durch eine nachträgliche Manipulation

unterschiedliche Grundfrequenzen hatten (sieben DF0-Bedingungen: 0, 2, 3, 5, 7, 9 und 12 Halbtöne). Für die Aufnahme wurden ein rauscharmes Messmikrofon (Gras, Typ 40HL-N, Seriennummer: 180883), ein Behringer-Vorverstärker (Powerplay Pro Modell HA 4400), eine RME Fireface UCX-Soundkarte und ein PC verwendet. Die aufgenommenen Vokale wurden mit einem Butterworth-Filter zweiter Ordnung und Grenzfrequenzen von 100 und 5000 Hz bandpassgefiltert und ein stationärer Teil von 200 ms wurde herausgeschnitten. Somit waren die Stimuli kurz genug, um eine eventuelle obligatorische Segregation zu beobachten (van Noorden, 1975), und lang genug, um die Natürlichkeit der Vokale zu erhalten. Alle Stimuli wurden mit 48000 Hz abgetastet. Die Frequenzmanipulation wurde mit dem Praat-Softwarepaket mit der Pitch-Synchronous Overlap-and-Add (PSOLA)-Methode durchgeführt. Somit variieren sowohl die Grundfrequenz der Vokale als auch die mit dem Timbre verbundenen spektralen Unterschiede innerhalb der Sequenzen.

In einer Sequenz waren alle acht deutschen Vokale in randomisierter Reihenfolge enthalten. Zwischen den Vokalen wurden keine Pausen eingefügt, sodass die stationären Teile der Vokale nur durch zwei Kosinus-Rampen getrennt waren (jeweils 30 ms). Die Sequenzen wurden so aufgebaut, dass jeder zweite Vokal von der gleichen Sprecherin stammte und zwei Vokale hintereinander die gleiche Grundfrequenz  $F_{0A}$  oder  $F_{0B}$  hatten (Abbildung 1). Die Grundfrequenzdifferenz (DF0) war innerhalb einer Sequenz immer konstant. Die Hälfte der generierten Sequenzen begann mit  $F_{0A}$ , die andere Hälfte mit  $F_{0B}$ . Auch die Reihenfolge der Sprecherinnen war über die Sequenzen ausgeglichen. Die 16-Bit-Sequenzen wurden mit Matlab erzeugt.



**Abbildung 1:** Schematische Darstellung der Vokal Sequenzen. Jeder zweite Vokal stammte von dem gleichen Sprecher und zwei Vokale hintereinander hatten die gleiche Grundfrequenz.

## Durchführung

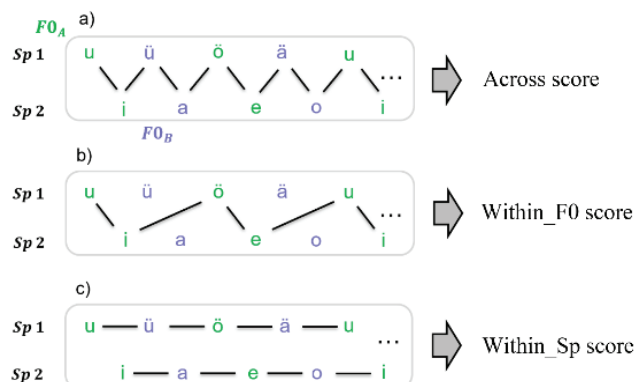
Die Aufgabe der Probanden in einem ersten Verifikationsversuch bestand zunächst darin, isolierte Vokale zu identifizieren. Dadurch wurde sichergestellt, dass die Vokale als Sprache erkannt wurden. Da auch in der Realität die Menschen darauf trainiert sind, auf verschiedene Merkmale zu hören, wurde vor dem Experiment ebenfalls ein Training durchgeführt. Die Probanden wurden darauf trainiert, sich gleichermaßen auf F0 und Timbre zu konzentrieren. Dieses Training bestand aus zwei Teilen: einem ‚Stimmen Training‘ und einem ‚F0 Training‘. Bei dem ‚F0 Training‘ wurden die Probanden eine halbe Stunde lang darauf trainiert die Sequenz basierend auf der F0 zu segregieren (Instruktion: Bitte teilen Sie die Sequenz in zwei Untersequenzen mit der gleichen Grundfrequenz auf. Bitte versuchen Sie, sich auf eine Grundfrequenz zu konzentrieren. Verwenden Sie einen der Vokale als Startpunkt und geben Sie die ersten vier Vokale, die Sie hören, wieder). Bei dem ‚Stimmen Training‘ sollten sich die Probanden eine halbe Stunde lang auf die Stimme fokussieren (Instruktion: Bitte teilen Sie die Sequenz in zwei Untersequenzen der gleichen Stimme auf. Bitte versuchen Sie, sich auf eine Stimme zu konzentrieren. Verwenden Sie einen der Vokale als Startpunkt und geben Sie die ersten vier Vokale, die Sie hören, wieder. Während beim ‚Stimmen Training‘ zwei verschiedene Stimmenpaare getestet wurden und der Parameter F0 konstant gehalten wurde (DF0 = 0 Halbtöne), erfolgte beim ‚F0 Training‘ die Testung von zwei DF0s (DF0 = 3 und 9 Halbtöne) und hierbei wurde der Parameter Stimme konstant gehalten.

Nach der Durchführung des Trainings wurden die Probanden in drei Tests (‚Free listening segregation test‘; ‚Segregation test based on F0‘ und ‚Segregation test based on voice‘) gebeten, die Reihenfolge der Vokale wiederzugeben. Beim ‚Free listening segregation test‘ hatten die Probanden die Aufgabe, die Sequenz in zwei Untersequenzen aufzuteilen und die Reihenfolge der Vokale innerhalb einer Untersequenz anzugeben. Die Aufmerksamkeit der Probanden wurde bei diesem Test in noch keine Richtung gelenkt. Beim ‚Segregation test based‘ on F0 war die Instruktion die gleiche wie beim ‚F0 Training‘ und beim ‚Segregation test based on voice‘ wurden die Probanden genauso instruiert wie beim ‚Stimmen Training‘. Die Aufmerksamkeit der Probanden wurde also in den letzten beiden Tests in Richtung einer Segregation basierend auf F0 oder auf Timbre gelenkt. Die Probanden gaben ihre Antwort in ein graphical user interface ein.

## Bewertung der Ergebnisse

Die Reihenfolge der wahrgenommenen Vokale hängt von der Wahrnehmungsorganisation ab, die Sequenz kann integriert oder segregiert werden. Im Falle einer Integration hören die Probanden alle Vokale in der zeitlich korrekten Reihenfolge hintereinander und nehmen nur einen einzigen Strom (also eine Sequenz) wahr, unabhängig vom Sprecher und der DF0. Die Integration der Sequenz führt zu der Erhöhung des across scores (Abbildung 3a). Eine Segregation kann auf der Grundlage der F0 Differenz oder der Timbre Differenz erfolgen. Ist eine dieser beiden Differenzen zwischen den Vokalen in der Sequenz groß genug, werden zwei Ströme (zwei Untersequenzen) wahrgenommen. Wird die Sequenz segregiert, also die Vokale innerhalb eines Stroms

wiedergegeben, erhöht sich der within score. Werden die Vokale mit der gleichen Grundfrequenz wiedergegeben, basiert die Segregation auf den F0 Merkmalen und es findet eine Erhöhung des within\_F0 scores statt (Abbildung 3b). Im Gegensatz dazu erhöht sich der within\_Sp score, wenn nur die Vokale vom gleichen Sprecher in der richtigen Reihenfolge wiedergegeben werden und die Segregation somit basierend auf dem Timbre Merkmal stattfindet (Abbildung 3c).

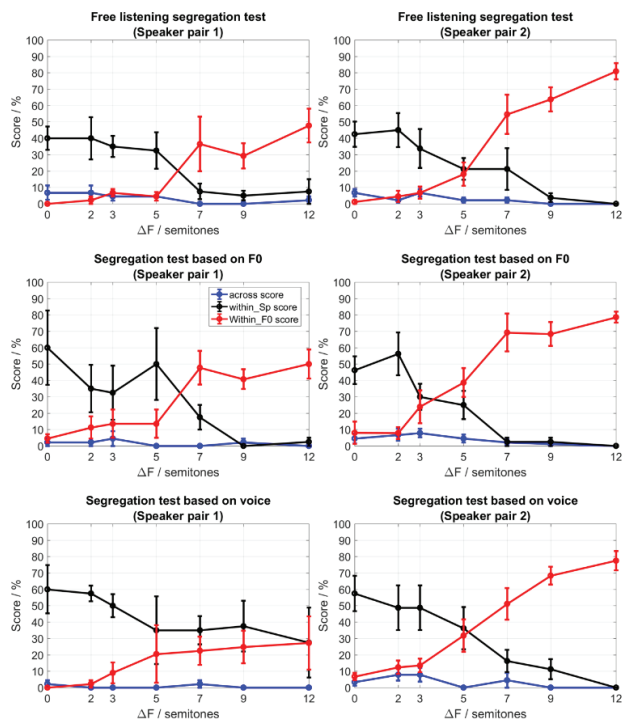


**Abbildung 3:** Schematische Darstellung der Bewertung der Ergebnisse. a) Integration der Sequenz, b) Gruppierung nach F0 und c) Gruppierung nach der Stimme. Die zwei Farben, grün und blau, repräsentieren unterschiedliche F0 Werte.

## Ergebnisse

Die Ergebnisse des Vokal-Identifikationstests zeigten, dass alle Probanden perfekt in der Lage waren, alle dargebotenen Vokale zu identifizieren (100%ige Genauigkeit). Bei jedem Probanden wurde individuell entschieden, mit welchem der beiden Stimmenpaare die drei Tests durchgeführt werden. Die Sequenzen des Stimmenpaares, bei dem der jeweilige Proband bei dem ‚Stimmen Training‘ das höhere Ergebnis erzielte, wurden bei den drei folgenden Tests verwendet. Somit ergaben sich zwei verschiedene Probandengruppen: die 1. Probandengruppe, die aus vier Probanden besteht und die die drei folgenden Tests mit den Sequenzen des 1. Stimmenpaares durchführte und die 2. Probandengruppe, die aus acht Probanden besteht und die die drei folgenden Tests mit den Sequenzen des 2. Stimmenpaares durchführte.

Die Ergebnisse der drei Tests sind in Abbildung 4 dargestellt, auf der linken Seite die Mittelwerte der 1. Probandengruppe (4 Probanden) und auf der rechten Seite die Mittelwerte der 2. Probandengruppe (8 Probanden). Der across score ist in blau dargestellt, der within\_F0 score in rot und der within\_Sp score in schwarz. Den Ergebnissen ist zu entnehmen, dass bei allen drei Tests die F0 basierte Segregation mit zunehmender DF0 zunimmt und die Timbre basierte Segregation mit zunehmender DF0 abnimmt. Darüber hinaus ist zu erkennen, dass der across score bei allen DF0 Konditionen unter 10 % liegt.



**Abbildung 4:** Ergebnisse der Segregationstests. Dargestellt ist der across score in blau, der within\_F0 score in rot und der within\_Sp score in schwarz. Die linken Abbildungen zeigen die Mittelwerte der vier Probanden, die das erste Stimmenpaar hörten und die rechten Abbildungen die Mittelwerte der acht Probanden, die das zweite Stimmenpaar hörten. Die oberen Abbildungen zeigen die Ergebnisse des ‚Free listening segregation tests‘, die mittleren die des ‚Segregation tests based on F0‘ und die unteren die des ‚Segregation tests based on voice‘.

Eine ANOVA mit Messwiederholungen wurde bei allen drei Tests für den within\_Sp score und den within\_F0 score durchgeführt, mit der Probandengruppe als zwischen-Subjekt-Faktor und den DF0s als inner-Subjekt-Faktor. Beim within\_F0 score ist der Haupteffekt von F0 bei allen drei Tests signifikant (‚Free listening segregation test‘:  $F(6,82) = 25,3$ ,  $p < 0,001$ ]; ‚Segregation test based on F0‘:  $F(6,82) = 19,95$ ,  $p < 0,001$ ]; ‚Segregation test based on voice‘:  $F(6,82) = 13,49$ ,  $p < 0,001$ ) und auch beim within\_Sp score ist der Haupteffekt von F0 bei allen drei Tests signifikant (‚Free listening segregation test‘:  $F(6,82) = 8,05$ ,  $p < 0,001$ ]; ‚Segregation test based on F0‘:  $F(6,82) = 13,76$ ,  $p < 0,001$ ]; ‚Segregation test based on voice‘:  $F(6,82) = 5,15$ ,  $p < 0,001$ ). Darüber hinaus ist beim within\_F0 Score der Effekt der Probandengruppe beim ‚Free listening segregation test‘ und beim ‚Segregation test based on voice‘ signifikant ( $F(1,82) = 7,59$ ,  $p < 0,05$ ];  $F(1,82) = 11,83$ ,  $p < 0,05$ ). Zusätzlich wurde eine ANOVA mit wiederholten Messungen durchgeführt, mit den DF0s und dem Test als inner-Subjekt-Faktoren und der Probandengruppe als zwischen-Subjekt-Faktor. Diese ANOVA zeigte, dass der Haupteffekt des Tests nicht signifikant ist ( $p = 0,065$ ).

Während der Punkt an dem F0 und Timbre die Segregation auf ähnliche Weise beeinflussen, bei der 2. Probandengruppe

bei allen Tests zwischen 3,6 und 5,1 Halbtönen liegt, variiert dieser bei der 1. Probandengruppe zwischen den Tests stärker und liegt bei ca. 6 Halbtönen beim ‚Free listening segregation test‘ und beim ‚Segregation test based on F0‘ und bei 12 Halbtönen beim ‚Segregation test based on voice‘.

## Diskussion

Der Verlauf der beiden within Scores war bei allen Tests wie erwartet: ein steigender within\_F0 score und ein fallender within\_Sp score mit steigender DF0. Diese Ergebnisse legen nahe, dass beide Merkmale (F0 und Timbre) die Wahrnehmung der zeitlichen Beziehungen zwischen den Vokalen innerhalb einer Sequenz beeinflussen und zu einer sequentiellen Segregation von Sprache führen. Bei niedrigen DF0 dominiert die Segregation basierend auf den Timbre Unterschieden, bei hohen DF0 dominiert die Segregation basierend auf den F0 Unterschieden und die Timbre Unterschiede werden unterdrückt.

Der within\_F0 score der 2. Probandengruppe stimmt gut mit den Ergebnissen von Gaudrain et al. (2007) überein, besonders bei hohen DF0. Beim Vergleich muss jedoch beachtet werden, dass Gaudrain et al. die primitive Segregation testeten, während in diesem Experiment die schemabasierte Segregation getestet wurde. Ein zusätzlicher Unterschied ist die Länge der Sequenzen, die in diesem Experiment verwendeten Sequenzen bestehen aus acht Vokalen, während die in Gaudrains Experiment genutzten Sequenzen aus sechs Vokalen bestehen. Des Weiteren ist der Rhythmus des F0 Merkmals innerhalb der Sequenz in beiden Experimenten nicht der gleiche, in diesem Experiment haben immer zwei Vokale hintereinander die gleiche F0, während in Gaudrains Experiment die Vokale abwechselnd die gleiche F0 haben. Auch der geringe across score trifft die Erwartungen, da es die Aufgabe der Probanden war zu segregieren und nicht zu integrieren.

Insgesamt zeigt sich für die 1. Probandengruppe ein ähnlicher Einfluss von Timbre und F0 Merkmalen auf die sequentielle Segregation von Vokalen. Auch Singh (1987) zeigte diese Konkurrenz zwischen den beiden Merkmalen bei der sequentiellen Segregation komplexer Töne, obwohl die Timbre Unterschiede in Singhs Experiment nicht durch verschiedene Sprecher verursacht wurden. Bei der 2. Probandengruppe konnte gezeigt werden, dass die F0 Merkmale einen größeren Einfluss auf die sequentielle Trennung haben als die Timbre Merkmale. Dieser große Einfluss von F0 Merkmalen auf die sequentielle Trennung von Vokalen konnte auch von Nooteboom et al. (1978) nachgewiesen werden.

Der Vorteil dieses Experiments gegenüber anderen früheren Experimenten, die ebenfalls die sequentielle Segregation von Vokalen untersuchten, besteht darin, dass der Vokal Identifikationstest sicherstellte, dass Sprache auch als Sprache erkannt wird. Es wurden mehr Probanden getestet und eine objektive Messmethode verwendet. Der Vorteil gegenüber dem Experiment von Gaudrain et al. (2007) ist, dass auch der Einfluss verschiedener Stimmen ausgewertet wurde und der Schwerpunkt bei diesem Experiment auf der sequentiellen Segregation lag.

Da sich weder der within\_F0 Score noch der within\_Sp Score der drei Tests signifikant voneinander unterscheidet, kann gefolgert werden, dass es keinen Einfluss auf die Segregation

hat, wenn sich die Probanden auf F0 oder auf die Stimme konzentrieren, und dass das stärkere Merkmal automatisch für die sequenzielle Segregation der Vokale verwendet wird.

### Fazit

Aus diesem Experiment kann geschlussfolgert werden, dass bei niedrigen DF0 die Timbre Merkmale überwiegen und somit eine Segregation basierend auf den Timbre Unterschieden stattfindet. Im Gegensatz dazu überwiegt bei hohen DF0 die Segregation basierend auf den F0 Merkmalen und die Timbre Unterschiede werden unterdrückt. Bei der 1. Probandengruppe scheint die Segregation basierend auf der Stimme und der F0 bei dem ‚Free listening segregation test‘ und bei dem ‚Segregation test based on F0‘ fast ausgeglichen zu sein, während für die 2. Probandengruppe die Grundfrequenz ein stärkeres Merkmal für die Segregation zu sein scheint als das Timbre. Der Punkt, an dem Timbre und F0 die sequenzielle Segregation von Vokalen in ähnlicher Weise beeinflussen liegt fast immer zwischen 3 und 6 Halbtönen. Darüber hinaus konnte festgestellt werden, dass es die Segregation von Vokalen nicht beeinflusst, wenn sich die Probanden speziell auf ein Merkmal (F0 oder Timbre) fokussieren. Daraus kann geschlussfolgert werden, dass das stärkere Merkmal für die sequenzielle Segregation verwendet wird und das beide F0 und Timbre Merkmale für die Hörstrom Segregation wichtig sind.

### Literatur

- [1] Clarke, J., Gaudrain, E., Chatterjee, M., and Başkent, D.: “T’ain’t the way you say it, it’s what you say— perceptual continuity of voice and top-down restoration of speech,” *Hear. Res.* 315 (2014), 80–87.
- [2] Gaudrain, E., Grimault, N., Healy, E., Béra, J.: Effect of spectral smearing on the perceptual segregation of vowel sequences. *Hear. Res.* 231 (2007), 32–41
- [3] Nooteboom, S. G., Brokx, J. P. L. & de Rooij, J. J.: Contributions of prosody to speech perception. *IPO Annual Progress Report* 11 (1976), 34–54.
- [4] Praat, URL: <http://www.praat.org>
- [5] Singh, P.: Perceptual organization of complex-tone sequence: A tradeoff between pitch and timbre? *Journal of The Acoustical Society of America* 82 (1987), 886–899.