Vokalnormalisierung

Jonathan Harrington

Johnson, K. (2004). Speaker normalization. In R. Remez, & D. B. Pisoni (Eds.), The Handbook of Speech Perception. Blackwell **johnson.pdf**

Adank, P., Smits, R., and van Hout, R. (2004): A comparison of vowel normalization procedures for language variation research. Journal of the Acoustical Society of America, 116, 3099–3107. **adank04jasa.pdf**

Vorperian, H. & Kent, R. (2007). Journal of Speech, Language, and Hearing Research, 50, 1510 –1545. **vorperian07.jshlr.pdf**

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. The Journal of the Acoustical Society of America, 109, 2988–2998. **perry01.pdf**

Vokalnormalisierung

Das Problem: wie trennt man akustisch und auditiv den phonetischen von dem sprecherbedingten (anatomischen) Beitrag in einem Vokal?

- Phonetisch: um zB /e, i/ zu differenzieren

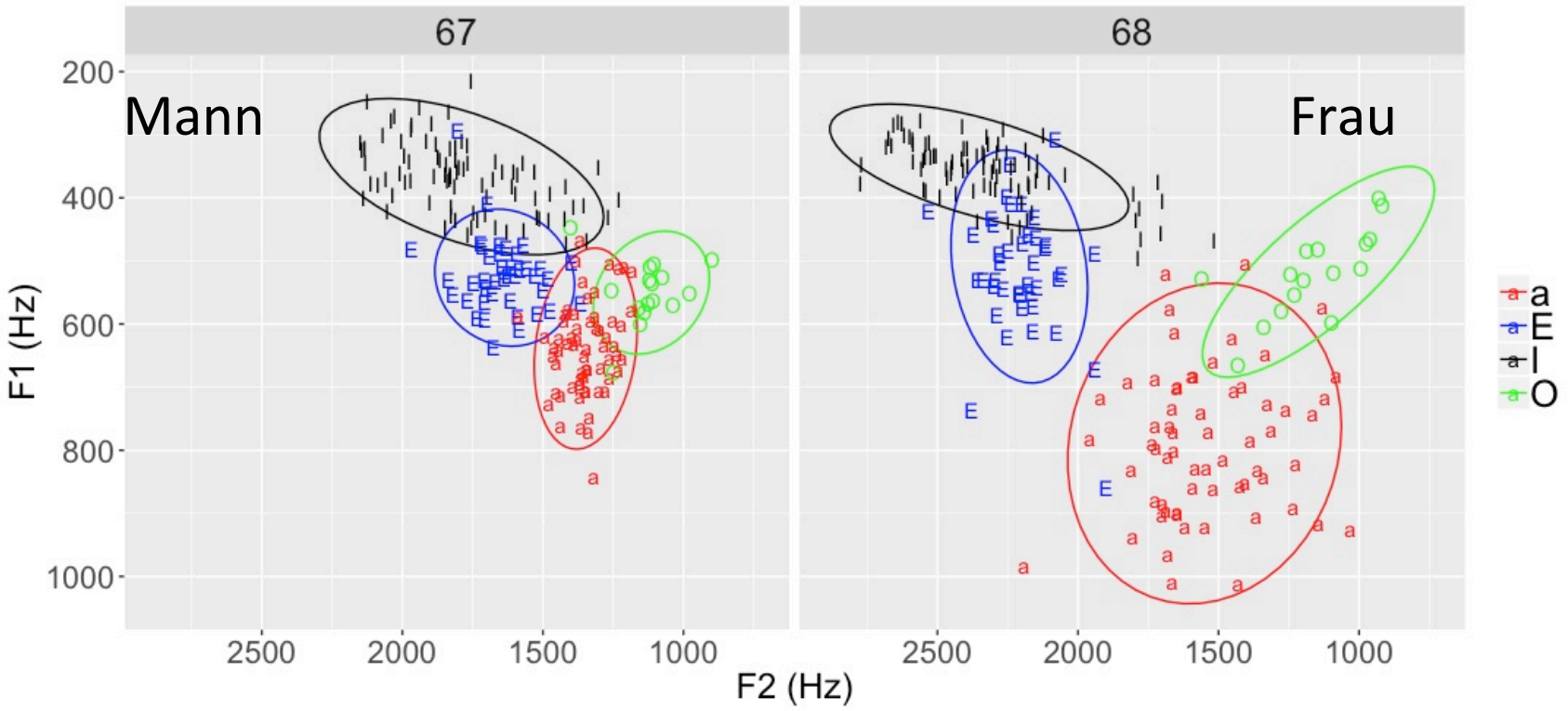
- Sprecherbedingt: Die Länge des Vokaltrakts hat einen großen Einfluss auf Formanten. Die Formanten von Kindern > Frauen > Männern

(Daher z.B. F1 von [e] von einem Mann kann einen ähnlichen Wert haben wie F1 von [i] von einer Frau usw.)

Formantwerte: Männer und Frauen

Die Größe der geschlechtsspezifischen Unterschiede variieren zwischen Vokalen. (Non-uniform scaling[1])

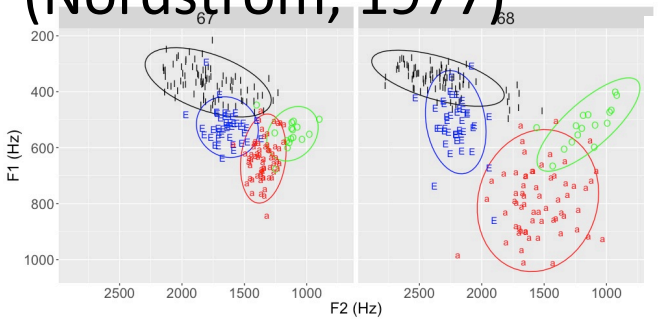
Gelesene Sätze Kiel-Corpus



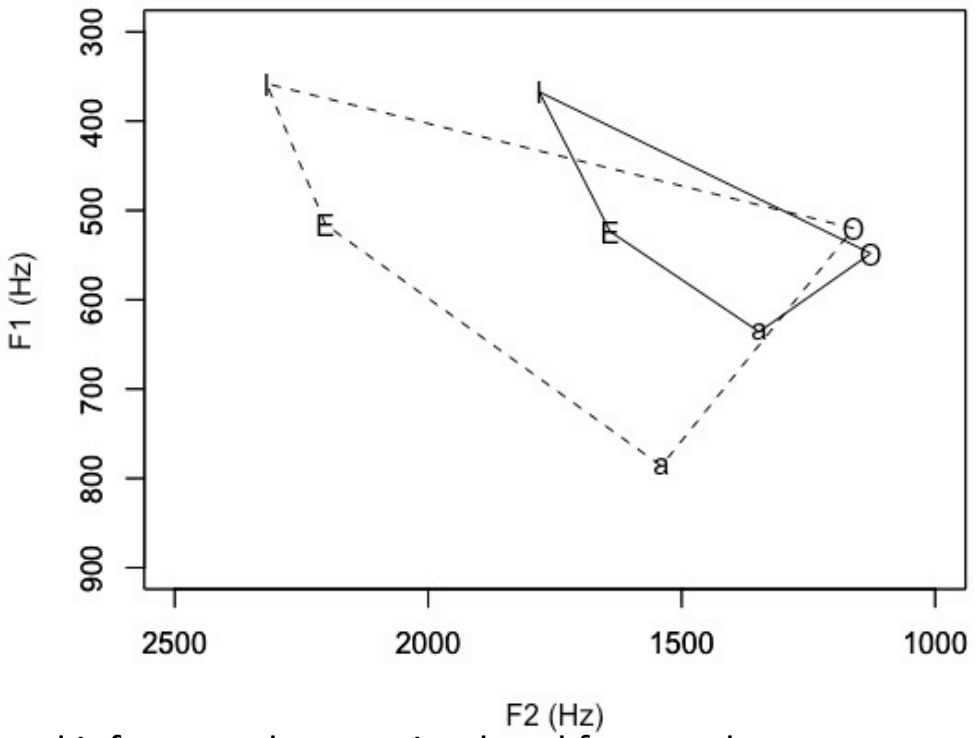
1. Fant (1975): Non-uniform vowel normalisation. *STL-QPSR*, 16, 1-19. http://www.speech.kth.se/prod/publications/files/qpsr/1975/1975_16_2-3_001-019.pdf **fant75.pdf**

Formantwerte: Männer und Frauen

F1 in offenen Vokalen und F2 in den vorderen Vokalen sind für Frauen deutlich höher als für Männer – das ist weil sie eher von der Länge des Rachenraums abhängen, die in Frauen deutlich kleiner ist (Nordström, 1977)[1]



Mittelwert davon



Mann

Frau

Nordström, P. E. (1977). Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics 5*, 81–92.

Perzeption

Intrinsische Normalisierung

Hörer normalisieren aufgrund der Information im Vokal selbst

Extrinsische Normalisierung

Hörer benötigen eine Stichprobe von Vokalen um für den Sprecher normalisieren zu können

Perzeption und die intrinsische Vokalnormalisierung

Derselbe phonetische Vokal (ob gesprochen von einem Mann oder Frau) soll ein ähnliches Muster entlang dem Basilarmembran verursachen (Potter & Steinberg, 1950[1]).

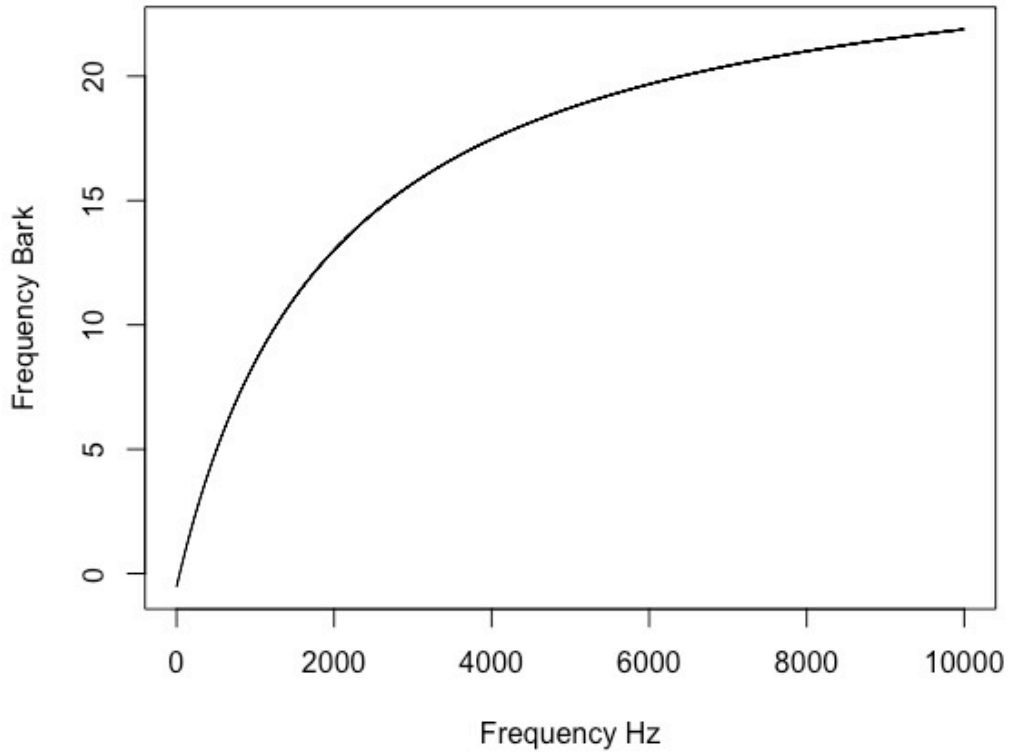
Der Pfad zur intrinsischen Normalisierung

1. Der Basilarmembran erzeugt eine quasi-logarithmische Transformation der Frequenzen (nächste Folie). Daher müsste allein eine Bark-Skala-Transformation einiges zur Sprechernormalisierung beitragen (Syrdal & Gopal, 1986[2]).

2. Perzeption hängt ab von dem **Verhältnis zwischen Formanten** nicht deren absolute Werte. Daher F2/F1, F3/F2 in Bark müsste zur Sprechernormalisierung beitragen (Miller, 1989[3];.Peterson, 1962[4])

1. Potter, R. & Steinberg, J. (1950) *Journal of the Acoustical Society of America* 22, 807-820. 2. Syrdal, A. K., & Gopal, H. S. (1986*) Journal of the Acoustical Society of America*, 79, 1086-1100. 3. Miller, J. D. (1989). *Journal of the Acoustical Society of America*, 85, 2114-2134. 4. Peterson, G. (1961). *Journal of Speech and Hearing Research*, 4, 10-28.

Perzeption und die intrinsische Vokalnormalisierung



Ein 1-Bark[1] Intervall entspricht einer Entfernung von 1.2 mm entlang des Basilmembrans

```
library(emuR)
x = 0:10000
plot(x, bark(x) ,type="l")
```

1. Genannt nach H. Barkhausen und zuerst von Zwicker (1961) vorgeschlagen *J. Acoustical Society of America*, 33, 248.

Perzeption und die extrinsische Vokalnormalisierung

Verschiedene Untersuchungen zeigen aber, dass Sprechernormalisierung eher **extrinsisch** sein könnte (abhängig von einer Stichprobe derselben Person).

Ladefoged & Broadbent (1957)[1]

Hörer mussten am Ende vom Satz ein ambiges Wort zwischen *bit*/*bet* identifizieren. Die Formanten im Trägersatz davor wurden entweder nach oben (= kleinerer Vokaltrakt) oder nach unten (=größerer Vokaltrakt) geschoben. Dasselbe akustische *bit*/*bet* Stimulus wurde mit einer höheren Wahrscheinlichkeit als *bet* bei tiefen Formanten im Trägersatz identifiziert.

1. Ladefoged, P. & Broadbent, D.E. (1957) Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98-104.

Perzeption und die extrinsische Vokalnormalisierung

Joos (1948)[1]

Hörer kalibrieren die Vokale eines Sprechers
im Verhältnis zu dessen Eckvokale.

Verbrugge et al (1976)[2]

1. Hörer identifizieren Vokale genauer aus einer Reihenfolge von Silben gesprochen (a) von derselben im Vgl. zu (b) verschiedenen Personen.

2. Die Hinzufügung der Eckvokale hat kaum eine Wirkung auf die Identifizierung.

1. Joos, M.A. (1948) Acoustic Phonetics, *Language 24, Suppl. 2*, 1-136. 2. Verbrugge, R., Strange, W., Shankweiler, D.. & Edman, T.R. (1976) What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America. 60*, 198-212

Vokalnormalisierung und akustische Klassifizierungen

Die besten Algorithmen für die Sprechernormalisierung von Vokalen sollen (a) die anatomisch bedingten Unterschiede zwischen Sprechern entfernen aber (b) ohne den phonetischen Inhalt zu zerstören.

Adank et al (2004)[1] wie Disner (1980)[2] davor testeten verschiedene bekannte Normalisierungs-Algorithmen für (a, b) auf eine Datenbank der niederländischen Sprache gesprochen von 80 Männern und 80 Frauen.

1. adank04.jasa.pdf

2. Disner (1980) *J. Acoustical Society of America*, 67, 253-161.

Vokalnormalisierung und akustische Klassifizierungen

Syrdal & Gopal (1986)[1] (intrinsisch)

Die Sprechernormalisierung soll stattfinden in einem Raum F1 – f0 (Bark) x F3 – F2 (Bark)

1. Bark Skalierung
2. Vordere Vokale wie [i] haben F3 und F2 eng zusammen; für hintere Vokale wie [u] sind F3 und F2 weit auseinander (Daher F3 – F2 als akustisches Merkmal für die Frontierung).
3. Tiefe Vokale wie [a] haben eine tiefere Grundfrequenz als hohe Vokale wie [i]. Daher F1 – f0 hoch für [a], klein für [i].
4. Die weitere perzeptive Basis von 2, 3: Zwei Frequenzgipfel mit einem Abstand von 3.5 Bark oder weniger werden perzeptiv integriert[2] (nicht differenziert). Integration von F3-F2 in vorderen abe nicht hinteren, und in F1 – f0 in tiefen aber nicht hohen Vokalen.

1. Syrdal, A. & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory represenation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086-1100.

2. Chistovich (1985*) J. Acoustical Soc. America*, 77, 789-805

Vokalnormalisierung und akustische Klassifizierungen

Lobanov (1971)[1] - extrinsisch

Standardnormalisierung.

Fn.mean, Fn.sd: Mittelwert und Standardabweichung von Fn (zB F1 n = 1) über eine Stichprobe von Vokalen desselben Sprechers.

$$Fn.norm = (Fn - Fn.mean)/Fn.sd$$

Gerstman (1968)[2] - extrinsisch

ähnlich

$$Fn.norm = (Fn - Fn.min)/Fn.Bereich$$

1. Lobanov, B. (1971). Classification of Russian *vowels spoken by different speakers.* Journal of the Acoustical Society of America, 49, 606–608. 2. Gerstman, L. (1968). Classification of self-normalized vowels. *IEEE Transactions of Audio and Electroacoustics*, AU-16, 78–80.

Vokalnormalisierung und akustische Klassifizierungen

Nearey (1978).

Die Formanten eines Vokals sind skaliert durch einen sprecherbedingten Konstanten (unterschiedliche Sprecher variieren in k)

$Fnorm = Fn/k$

$\log(Fnorm) = \log(Fn) - \log(k)$

log(k) wird auf verschiedene Weisen eingeschätzt.

In einer Version:

log(k) = Mittelwert von G1mean + G2mean

G1mean = Mittelwert von log(F1)

G2mean = Mittelwert von log(F2)

berechnet über alle Vokale desselben Sprechers.

1. Nearey, T.(1978). *Phonetic Feature Systems for Vowels* Indiana Uni versity Linguistics Club, Indiana.

Vokalnormalisierung und akustische Klassifizierungen

Miller (1989)[2]

Einflüsse von der intrinsischen Normalisierung aber extrinsisch

F1norm = log(F1/SR)

F2norm = log(F2/F1)

F3norm = log(F3/F1)

SR = sensory reference = der geometrische Mittelwert[1] aller Grundfrequenzwerte desselben Sprechers

1. geometrischer Mittelwert von 2, 4, 8 = $(2 \times 4 \times 8)^{(1/3)}$

2. Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, 85, 2114-2134.

Vokalnormalisierung und akustische Klassifizierungen

Nordström and Lindblom (1975)[1]

Vokale von Frauen werden auf der Basis der eingeschätzen Unterschiede in der Gesamtvokaltraktlänge zwischen Männern und Frauen herunterskaliert. Uniform-scaling (weil dieselbe Skalierung auf alle Vokale angewandt wird).

Das Algorithmus beruht auf die Idee, dass F3 in offenen Vokalen im Verhältnis zur gesamten Vokaltraktlänge ist.

Nur die Vokale von Frauen werden normalisiert

$F_{norm} = k \; F_n$ $\quad k = F3(\text{Männer})/F3(\text{Frauen})$

F3(Männer): Der F3-Mittelwert in allen Vokalen produziert von Männern.

F3(Frauen): Der F3-Mittelwert in allen Vokalen produziert von Frauen.

1. Nordström, P. & Lindblom, B. (1975) A normalization procedure for vowel formant data, *International Congress of Phonetic Sciences* in Leeds. Siehe auch: Nordström, P. (1977) Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics* 5, 81–92

Vokalnormalisierung und akustische Klassifizierungen

Nordström and Lindblom (1975)[1]

Kritik von Fant (1975)[1]. Der 'scale-factor' ist non-uniform variiert also zwischen Vokalen (siehe S. 3 und 4). d.h. man braucht eigentlich einen unterschiedlichen *k* pro Vokal.

Fant (1975): Non-uniform vowel normalisation. *STL-QPSR*, 16, 1-19. http://www.speech.kth.se/prod/publications/files/qpsr/1975/1975_16_2-3_001-019.pdf
fant75.pdf

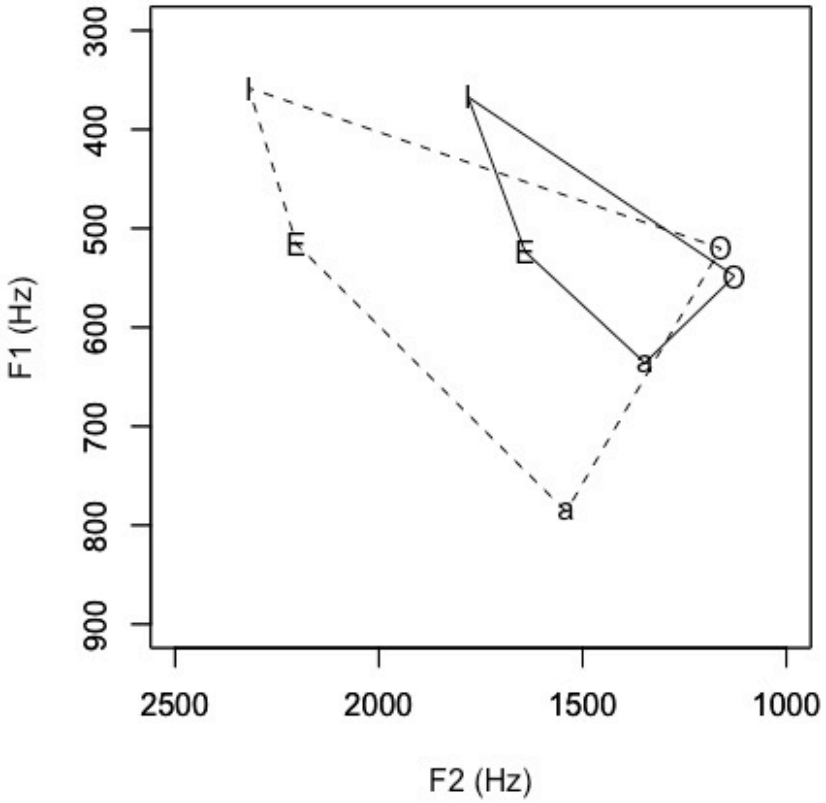
Akustische Klassifizierungen in Adank et al (2004): Ergebnisse für Vokal

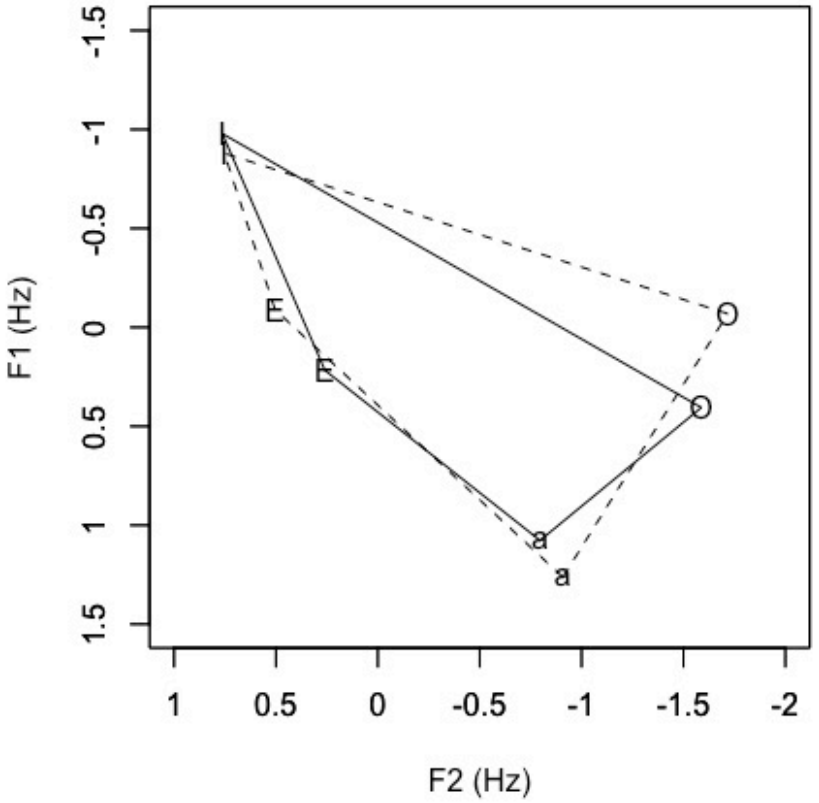
Hier wird trainiert und klassifiziert auf Vokale
Vokale sollen differenziert bleiben. Lobanov (1971) schneidet sehr gut ab.

Hz: Klassifizierung nicht normalisierter Daten.

		LDA 1	QDA 1
	HZ	79	81
	LOG	80	81
Vowel-intrinsic	BARK	80	82
	ERB	80	82
	MEL	80	82
	S & G	69↓	70
	LOBANOV	92↑	93↑
	NEAREY1	90	91↑
Vowel-extrinsic	NEAREY2	82↑	83
	GERSTMAN	84↑	86↑
	NORDSTRÖM	82↑	84↑
	MILLER	76↓	77

Lobanov Normalisierung





Akustische Klassifizierungen: Ergebnisse für Vokal

Hier wird auf Gender trainiert (also inwiefern kann akustisch zwischen m-W Sprechern differenziert werden).

Die besten Algorithmen (Lobanov, Nearey, Gerstman) schneiden mit 50% ab (= Differenzierung zwischen Gender nicht mehr möglich).

	Predictor variables	LDA 2 *F*0, *F*1, *F*2, *F*3	LDA 3 *F*0	LDA 4 *F*1, *F*2, *F*3
	HZ	93	89	80
	LOG	93	89	80
Vowel-intrinsic	BARK	93	89	80
	ERB	93	89	80
	MEL	92	89	80
	S & G	53*	...	...
	LOBANOV	50*	51*	51*
	NEAREY1	50*	51*	49*
Vowel-extrinsic	NEAREY2	81	78	69
	GERSTMAN	53*	53*	51*
	NORDSTRÖM	83	82	52*
	MILLER	79	...	...

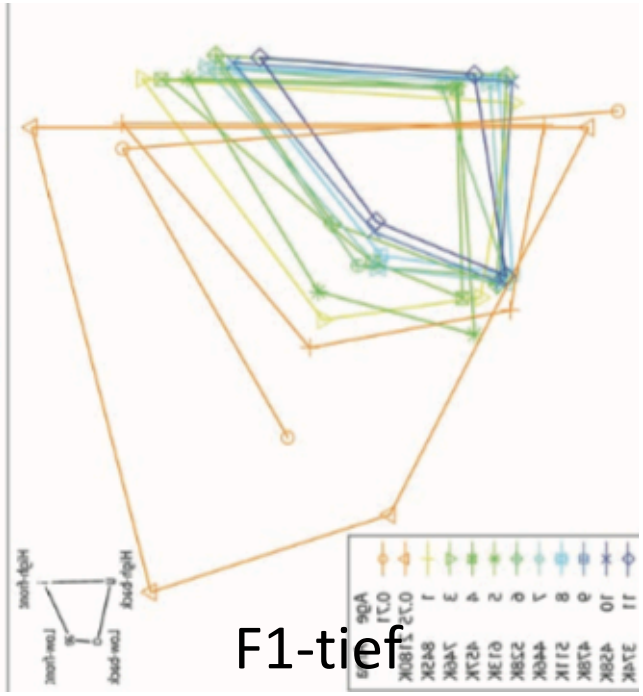
Vokale und M-W Unterschiede in Kindern.

Kinder haben insbesondere viel höhere F1-Werte in offenen und höhere F2-Werte in vorderen Vokalen als Erwachsene.

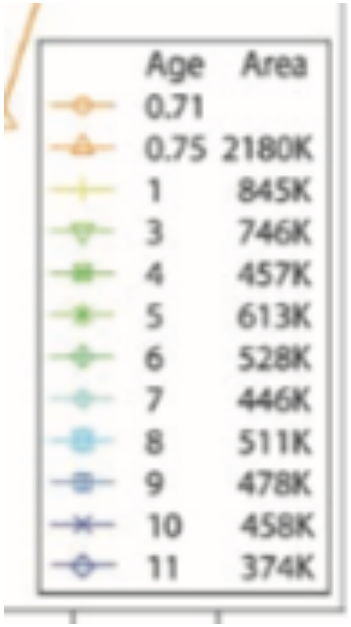
Es gibt eine plötzliche Änderung im Alter 1-4 Jahren – weil der Kehlkopf senkt und dadurch der Rachenraum länger wird

F2-hoch

F2-tief



F1-tief

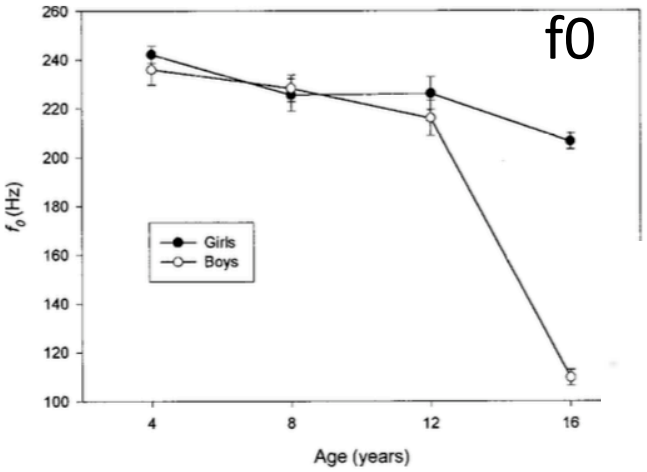


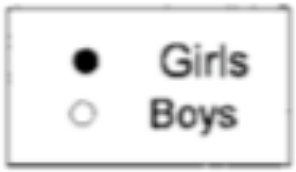
Vokaldaten von Kindern im Alter 8 Monate bis 11 Jahre

Aus vorperian07.pdf

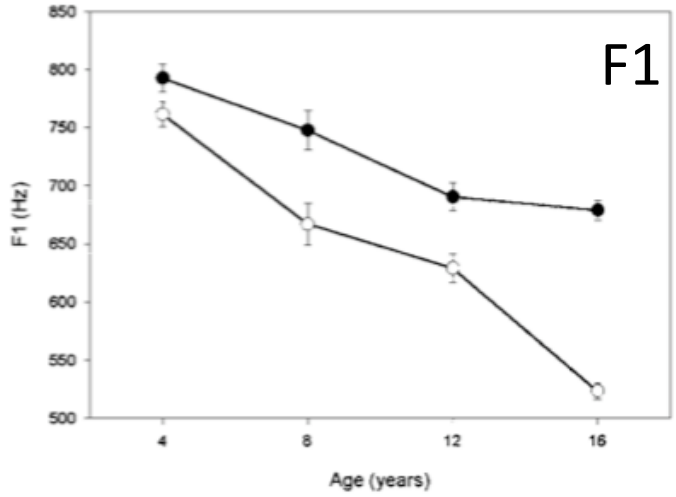
Vokale und M-W Unterschiede in Kindern.

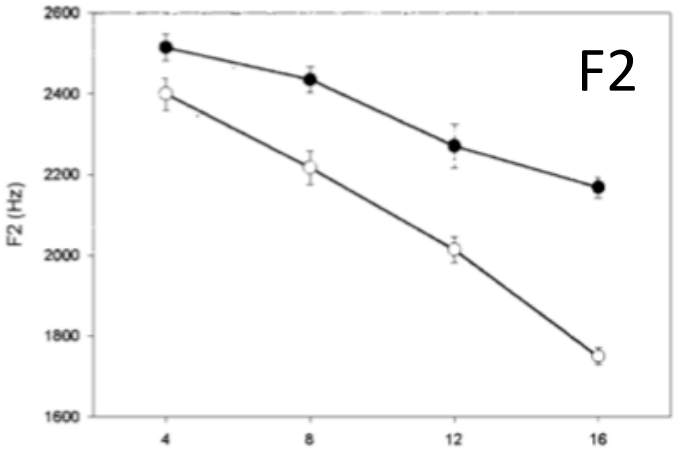
Änderungen in f0, F1, F2 zwischen 4 und 16 Jahren in Mädchen und Jungen.





Es gibt bereits geschlechtsspezifische Unterschiede in F1 und F2 ab 4 Jahren.



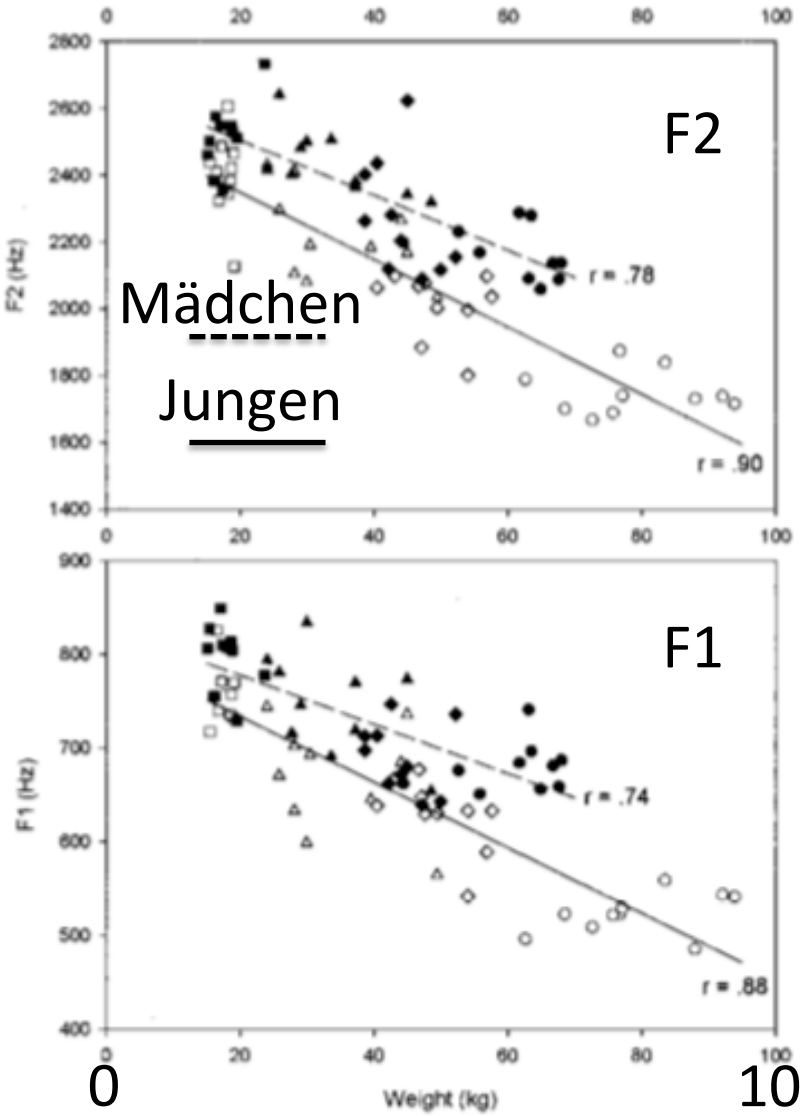


4 16

Daten aus **perry01.pdf**

Vokale und M-W Unterschiede in Kindern.

Formanten in den Kindern wurden mit Körpergröße korreliert.
Geschlechtsspezifische Formantunterschiede sind nicht nur anatomisch sondern eventuell auch soziophonetisch (da die Regressionslinie für Mädchen höher liegt).



0 100 kg

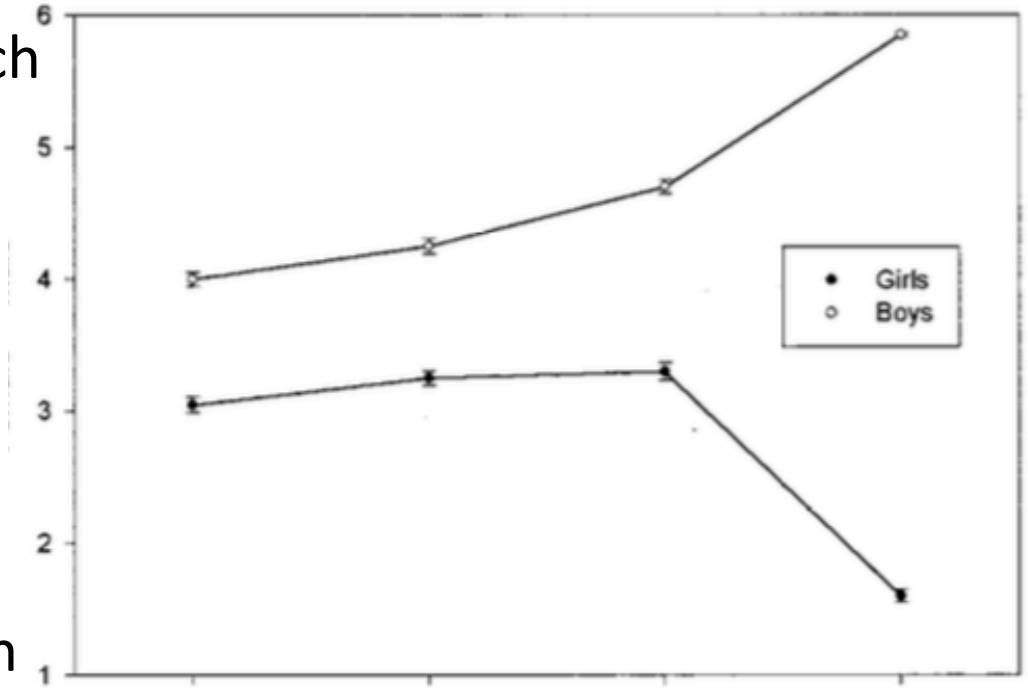
Gewicht

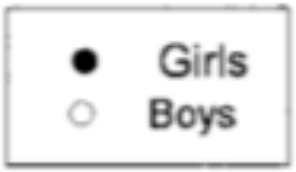
Daten aus perry01.pdf

Vokale und M-W Unterschiede in Kindern.

CVC Silben von Kindern. 20 (erwachsene) Hörer mussten antworten auf einem 6 Punkt Skala wie männlich/weiblich die Vokale klingen.

männlich



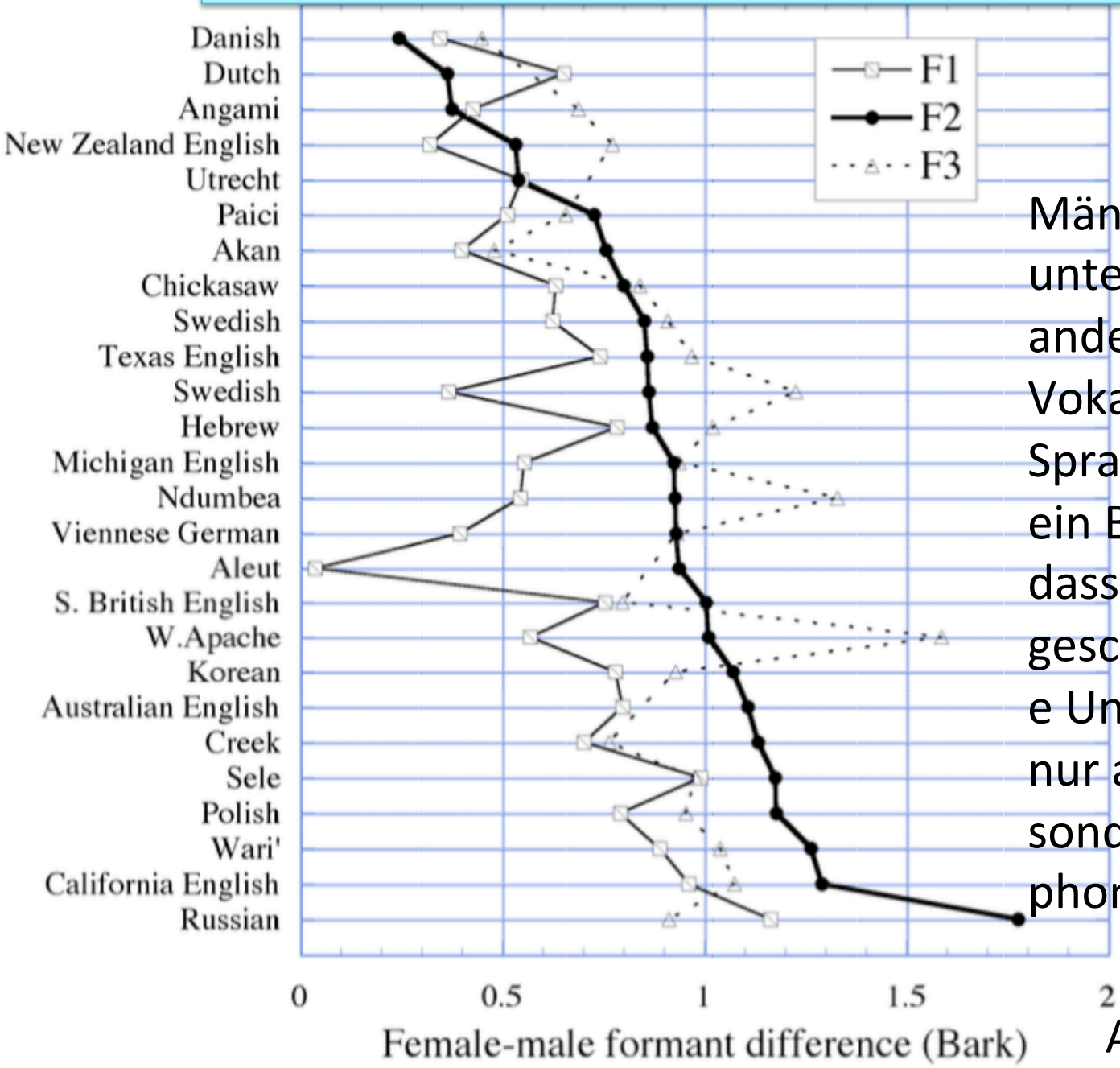


weiblich

4 Alter 16 Jahre

Daten aus perry01.pdf

Soziophonetische M-W Unterschiede.



Männer und Frauen unterscheiden sich anders in Vokalformanten von Sprache zu Sprache – ein Beweis daher, dass geschlechtsspezifische Unterschiede nicht nur anatomisch sondern auch phonetisch sind

Aus johnson04.pdf