

Speaker normalization in speech perception

Iveta Patáková

IVETA PATÁKOVÁ

11.1.2017

SPRACHE



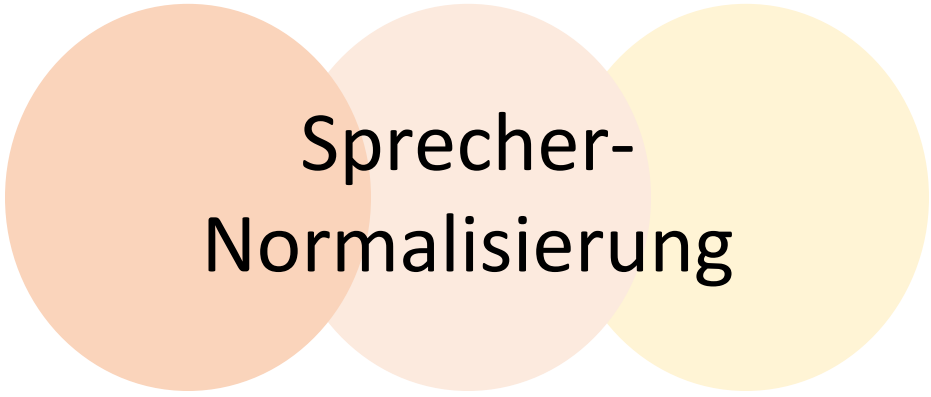
variable



INNERHALB



ZWISCHEN



Sprecher- Normalisierung



phonologisch identische Aussagen zeigen
akustische Variationen zwischen
Sprechern



Hörer können trotz diesen
Variationen Wörter erkennen

Starke Variation der
Vokale

Sprecher



Sprecher

Männer



Frauen

Erwachsene



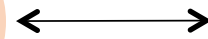
Kinder

Solche sprecherspezifischen Merkmale ändern oft dieselben Cues, die für phonetische Unterschiede benötigt werden

ein hohes F2

trennt

[i]



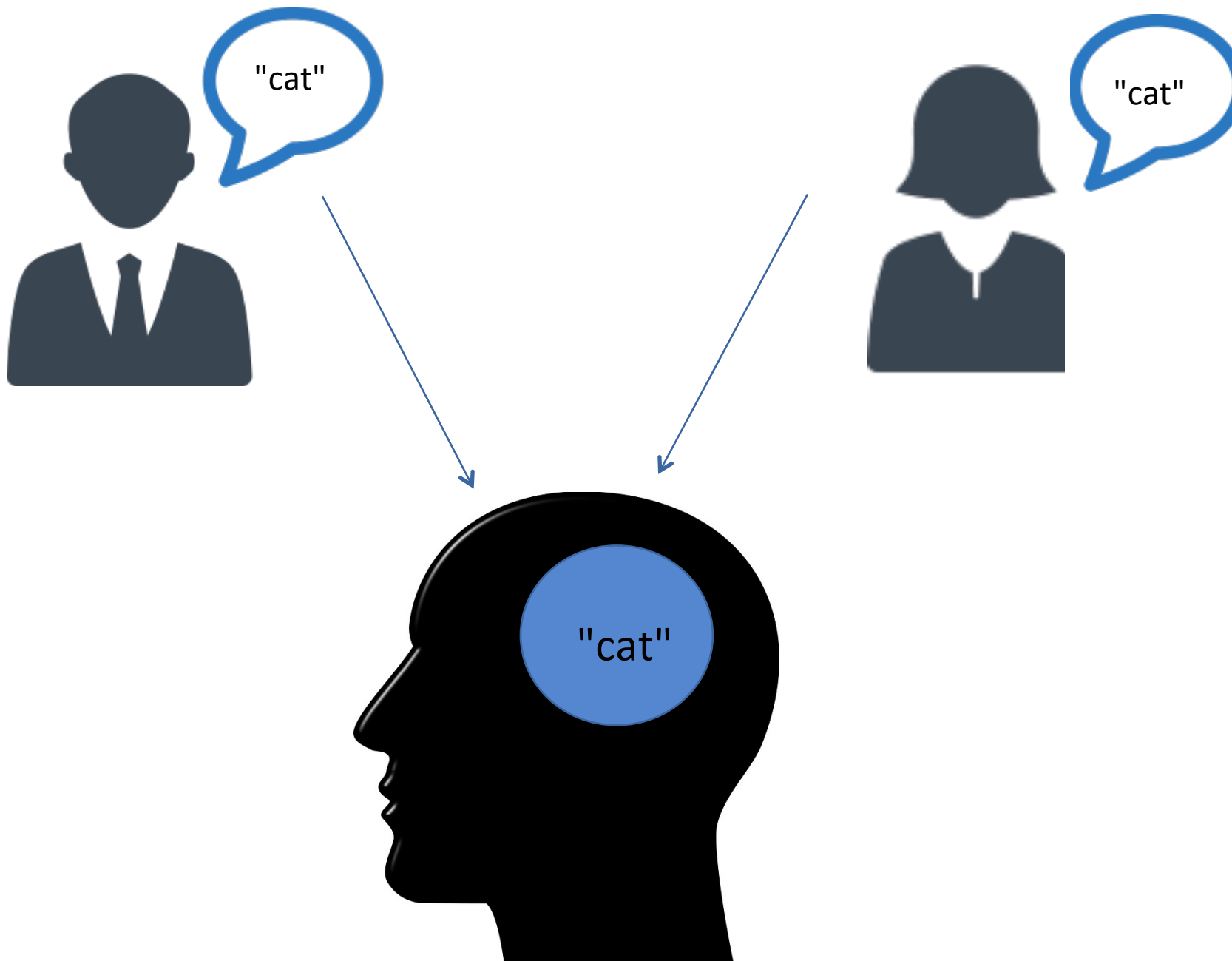
[e]

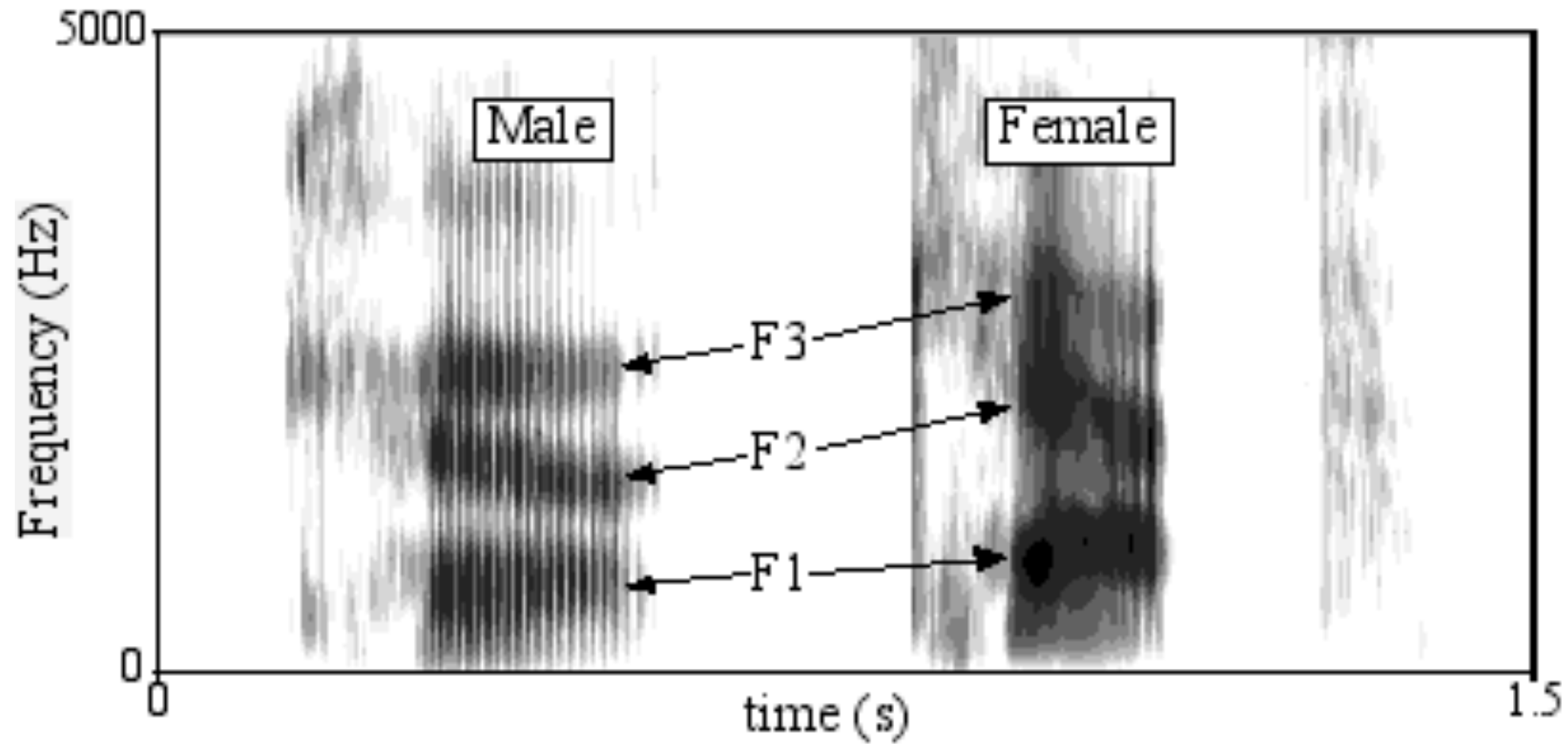


Frauen haben höhere F2 als Männer



F2-Mann-[i] kann ähnlich F2-Frau-[e] sein

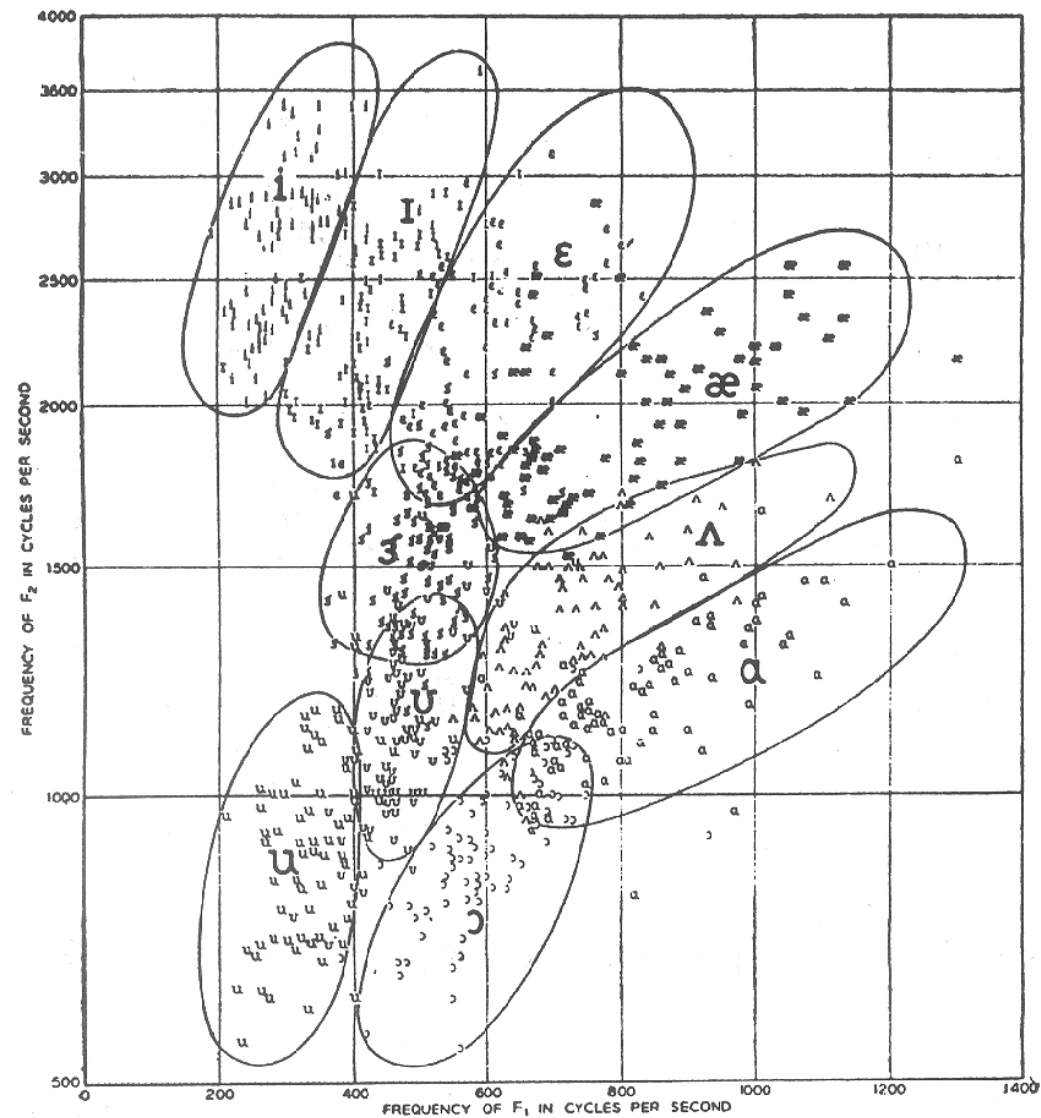




Spectrogram of a man and a woman saying "cat". The three lowest vowel formants (vocal tract resonant frequencies are marked as F1, F2 and F3) (Johnson 2004)

Peterson & Barney (1952)

- Es wurden Vokale von 76 Sprechern aufgenommen- Frauen, Männer und Kinder
- Signifikante Unterschiede zwischen Vokalrealisationen
- Große Variabilität in Vokalproduktion
- Vokalformanten bei Frauen – um 20 % höher als bei Männern
- Vokalformanten bei Kindern noch höher als bei Frauen



Scatter plot of first and second formant values of American English vowels. From Peterson Barney 1952

FORMANT-RATIO THEORY

IDEE :

~~Absolute
Formantwerte~~

Verhältnis (Ratio)
zwischen Formanten

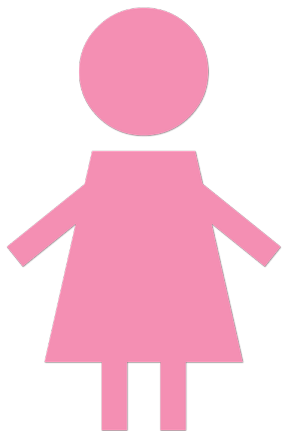


Vokale sind relative Muster, nicht absolute Frequenzen der Formanten

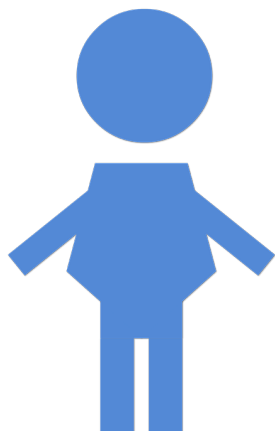


Durch die Berechnung von Formant-Ratios entsteht eine gewisse Normalisierung.

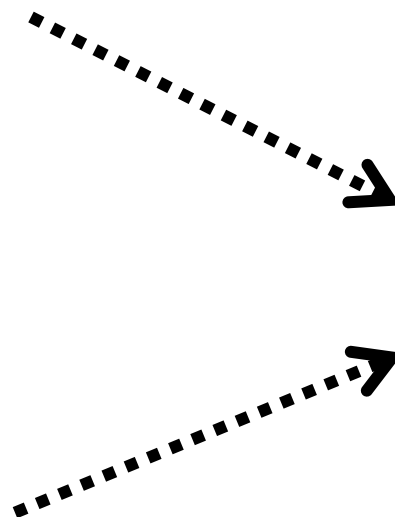
[i]



$F1 = 250 \text{ Hz}$
 $F2 = 3000 \text{ Hz}$



$F1 = 200 \text{ Hz}$
 $F2 = 2400 \text{ Hz}$



Verhältnis $F2/F1$
ist für beide
gleich

Problem mit allen
Formant-Ratio-
Theorien



Unterschied zwischen
Männern und Frauen **ist
nicht linear**

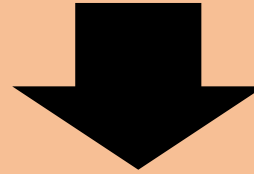
- d.h. der männlich-weiblich Unterschied in den Formanten für [a] ist größer als für [u]
- der Formant-Unterschied zwischen zwei Sprechern A, B ist immer ein Verhältnis A/B
- das Verhältnis wird sehr stark von der Identität des Vokals beeinflusst

Die Änderungen in
den Formanten



NICHT linear

F1 (Sprecher A) um 20 % $\uparrow$ als F1 (Sprecher B)



F2 (Sprecher A) um 20% $\uparrow$ als F2 (Sprecher B)

Sonst würde das Verhältnis F1/F2 nicht linear

Formant-ratio
Normalisierung



Intrinsisches
Verfahren

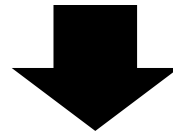
- Man benötigt nur den einen Vokal vom Sprecher, um die Normalisierung durchzuführen

Bestandteil von F-R-T



Hörer nicht mehr als den
aktuell wahrgenommenen
Vokal benötigt, um die
Normalisierung durchzuführen

Extrinsische Normalisierung



- man benötigt als eine Stichprobe der Vokale von einem Sprecher um zu normalisieren



Gibt es irgendwelche kognitiven Beweise, dass Normalisierung extrinsisch ist?

Ladefoged and Broadbent (1957)

- Hörer haben einen Satz wahrgenommen mit einem final 'bit' oder 'bet'
- also Trägersatz folgt von bVt , V = /ɪ/ oder /ɛ/
- Höhere Formanten im Trägersatz führten zu mehr 'bet' Urteile.
- Wenn daher die Formanten generell höher sind, dann gibt es eine größere Wahrscheinlichkeit, dass das gleiche bVt Stimulus eher als /bet/ (im Gegensatz zu /bit/) wahrgenommen wird.

Sprechernormalisierung scheint also
extrinsisch zu sein!

Summerfield and Haggard (1975)

- Die Hörer reagieren langsamer, Vokale zu identifizieren, wenn sie Wörter in einer Liste hören, gesprochen von unterschiedlichen im Vergleich zum selben Sprecher

Verbrugge et al. (1976)

- Die Vokalidentifikation der Wörter in einer Liste war akkurater, wenn die Wörter von einem Sprecher (9,5% Error) im Vergleich zu unterschiedlichen Sprechern (17% Error) produziert wurden.

Nussbaum und Morin (1992)

- Nussbaum und Morin haben einen beschleunigten „Phoneme Monitoring Task“ benutzt, um die Ungewissheit des Sprechers auszuwerten.
- Es wurde herausgefunden, dass die Hörer die Target-Silbe langsamer in „Mixed-Talker“ Liste berichtet haben.

Hörer-Erwartungen

Johnson 1999

- Kontinuum zwischen „HOOD“ [hʊd] und „HUD“ [hʌd]
- die Stimme war synthetisch und ambig ob männlich oder weiblich
- die Grenzen zwischen Vokalkategorien unterscheiden sich bei männlichen und weiblichen Stimmen
- die Vokalgrenzen ändern sich in Abhängigkeit von Geschlecht des Sprechers

Geschlechtsspezifische Unterschiede und Sprache

Bladon, Henton und Pickering (1984)

- Dänisch : der kleinste geschlechtsp. Unterschied
- Russisch : der größte geschlechtsp. Unterschied

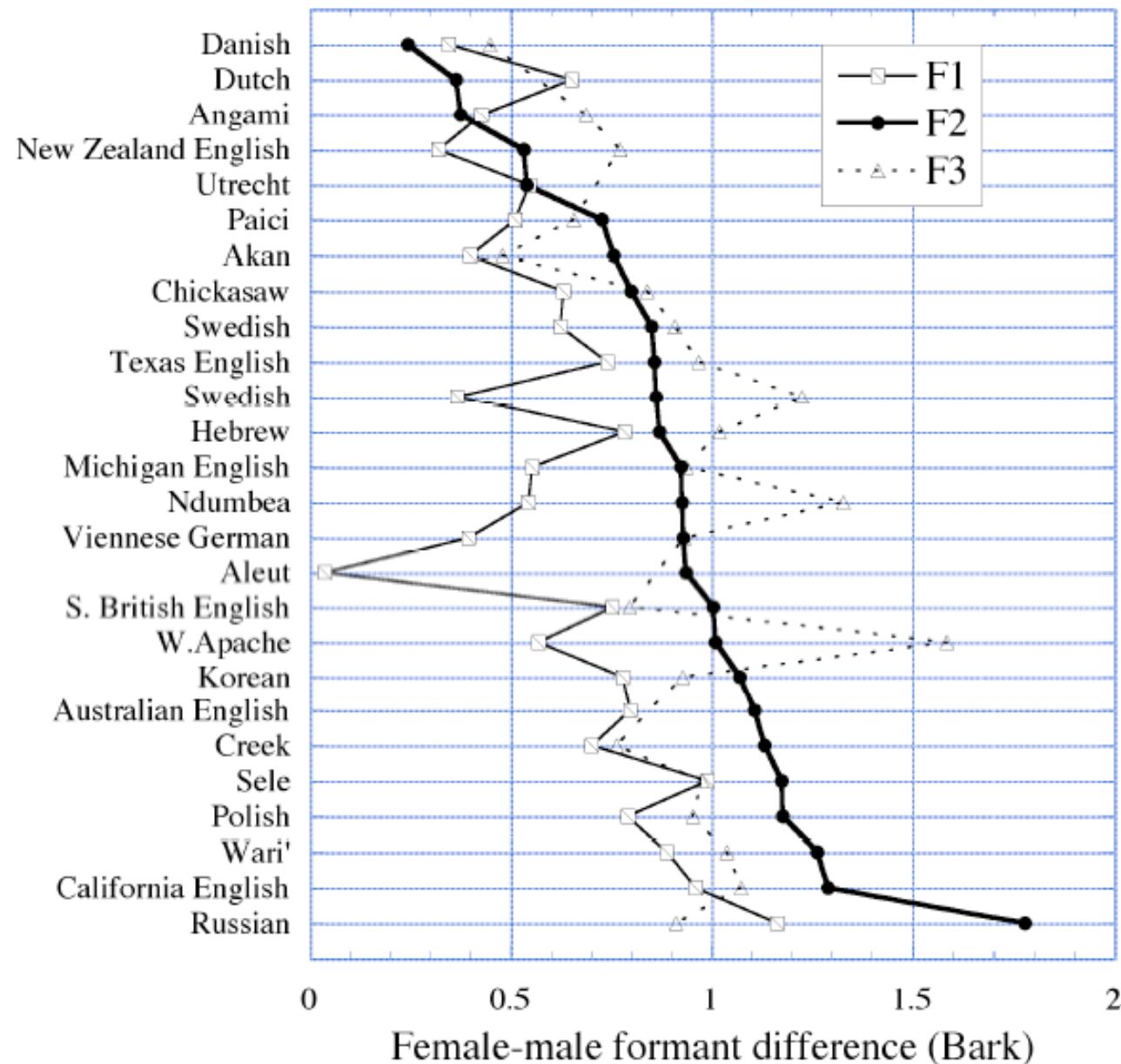


Figure 5. The difference between men's and women's average formant frequencies for 26 groups of speakers. F1 is plotted with a crossed open box, F2 is plotted with a filled circle, and F3 is plotted with an open triangle.

Quellen

- Johnson, K. (2004). Speaker normalization. In R. Remez, & D. B. Pisoni (Eds.), The Handbook of Speech Perception. Blackwell
- Harrington, J. & Cassidy, S. (1999) Techniques in Acoustic Phonetics. Kluwer.
- http://www.utdallas.edu/~assmann/hcs6367/johnson_strand_dimperio99.pdf
- http://www.ling.upenn.edu/courses/Fall_2013/ling520/PetersonBarney52.pdf