Speaker Normalization of Stressed and Unstressed Vowels in Articulatory and Formant Spaces

Christian Geng, Christine Mooshammer

Research Centre for General Linguistics, Typology and Universals (ZAS), GWZ Berlin e.V., Jaegerstr. 10/11, 10117 Berlin, GERMANY

geng@zas.gwz-berlin.de; timo@zas.gwz-berlin.de

ABSTRACT

One of the goals of phonetic investigations is to find strategies for vowel production independent of speaker-specific vocal-tract anatomies and individual biomechanic properties. In this study we apply a technique for speaker normalization to formant spaces, lingual vowel target positions which is often referred to as Procrustes Analysis. Procrustes Analysis comes in two guises: In the orthogonal version, which is based on Euclidean geometry, only transformations are allowed which preserve the angles between the measured landmarks or objects. It is more apt for the analysis and normalization of formant spaces. In contrast, the second version of the model is based on affine geometry, which means that the angles between corresponding landmarks do not have to be preserved. It has its strengths in the articulatory domain.

We can show the relevance of these models by

a) analysing the remaining variances after application of these procedures to articulatory and acoustical vowel targets and by

b) relating the transformations applied to the articulatory target configurations to anatomical landmarks (EPG midline traces).

AIMS OF THIS STUDY

- Extraction of speaker-independent strategies for stress production in vowels in the articulatory and the acoustical domain
- Investigation of the relationship between articulatory tongue positions and formant values during the acoustically defined mid-vowel
- Relation between speaker-dependent strategies to anatomical landmarks (palate shape)

DATA ACQUISITION

- Tongue tip movements of 7 German speakers (5 male, 2 female) by EMMA (AG 100, Carstens Medizintechnik)
- Nonsense syllables: /tVt/ with /i-ɪ, y-ʏ, e-ɛ, ɛː, ø-œ, aː-a, o-ɔ, u-ʊ/
- Stress alternations fixed by morphologically conditioned word stress and contrastive stress, e. g.: "Ich habe /'tʰoːtʰɐ/, nicht /tʰo'tʰaːl/ gesagt." ("I said /'tʰoːtʰɐ/, not / tʰo'tʰaːl/.")
- 15 sentences repeated 6 (4 speakers) or 10 times (3 speakers)
- 4 sensors placed on the tongue (1 to 5 cm from tongue tip), one on lower incisors and lower lip, nasion and upper incisors (references sensors)
- Speech signal recorded simultaneously on DAT
- Palate midlines measured by sliding caliphers on EPG palates (6 speakers)
  Palate shape measures: palate length, doming
- At the acoustically defined mid-vowel:
  ➔ Formant frequencies F1 F2
  ➔ Articulatory target positions of 4 tongue sensors

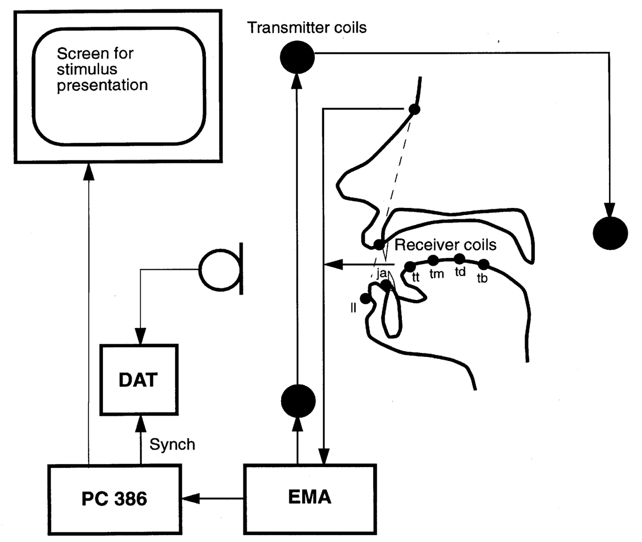


Figure 1: Experimental set-up

METHOD: PROCRUSTES ANALYSIS

Based on Gower (1975), Rohlf & Slice (1990)

Translation, rotation and scaling of landmarks until the least-squares fit no longer improves

Input landmarks:
Formant values (F1, F2) — Tongue configurations (4 x/y)

↓

Generalized Procrustes Analysis:
Translation — Rotation (affine=stretching and shrinking) — Scaling

↙ ↘

Speaker-independent outputs: Consensus Objects
= speaker-independent landmarks
• consensus formant space
• consensus articulatory space visualization by 2D Principal Component Analysis factorial plots

Speaker-dependent outputs:
1. transformation eigenvalues: straincrosses (amount of affine deformation)
2. speaker-dependent modelled data (raw data transformed to Procrustes space)

RESULTS

1. Formant Spaces

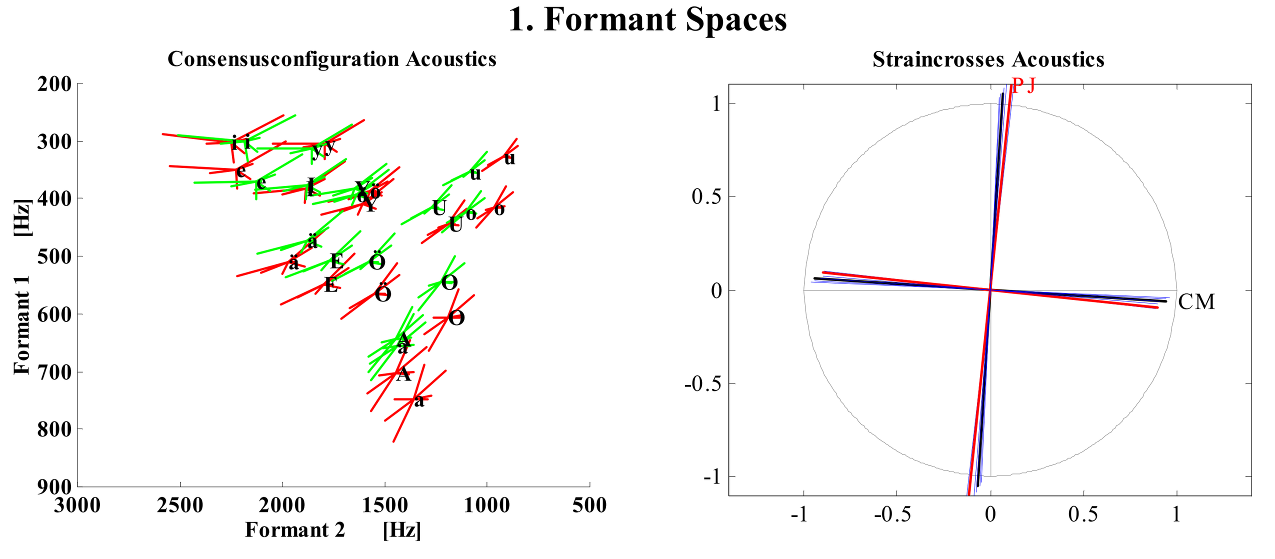


Figure 2, left panel: formant consensus object of stressed and unstressed vowels. Lines correspond to speaker-dependent models. Right panel: Amount of necessary speaker-dependent affine transformation compared to the consensus object.

Very little deviation of speaker-dependent straincrosses from consensus object (unit circle)
➔ only rigid rotation and scaling necessary for formant spaces

2. Articulatory Spaces

Consensus object: Reduction 8 dimensions to 2 by Principal Component Analysis

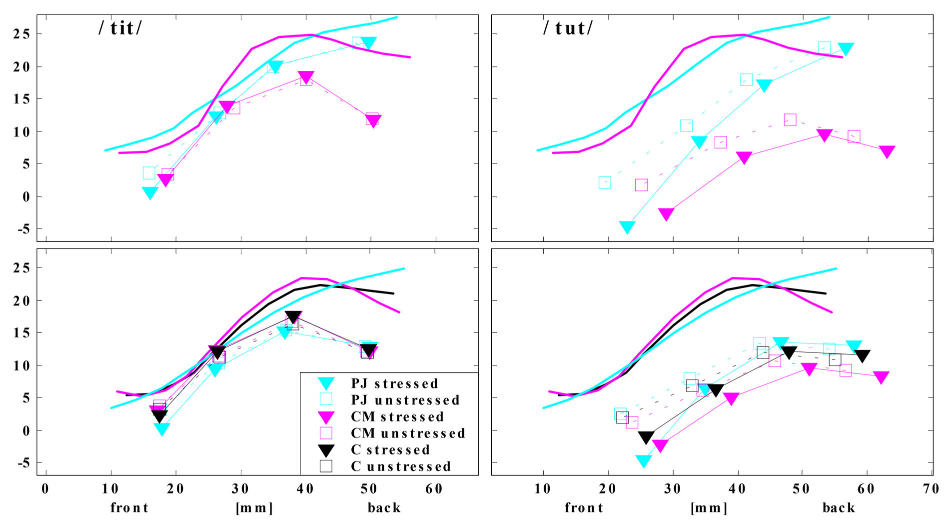
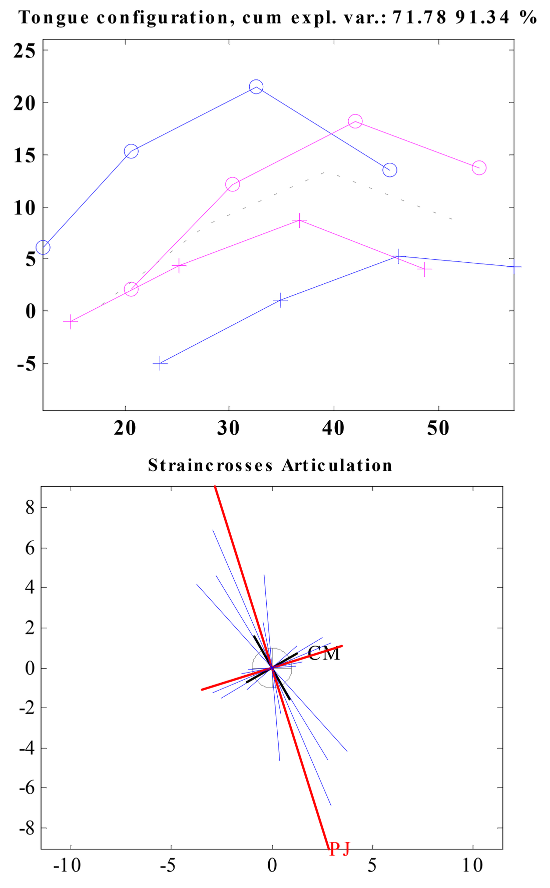


*Figure 4: Comparison between **raw articulatory configurations of two speakers (upper panels)** and modeled data (lower panels) for /tit/ (left) und /tut/ (right). Male PJ, female CM, consensus of seven speakers.*

Raw data: tongue configurations for /i/ simply differ for the two speakers because of different palate shapes although in both cases a close constriction is formed.

Modeled data: Data in consensus space primarily reflect differential use of articulatory strategies.



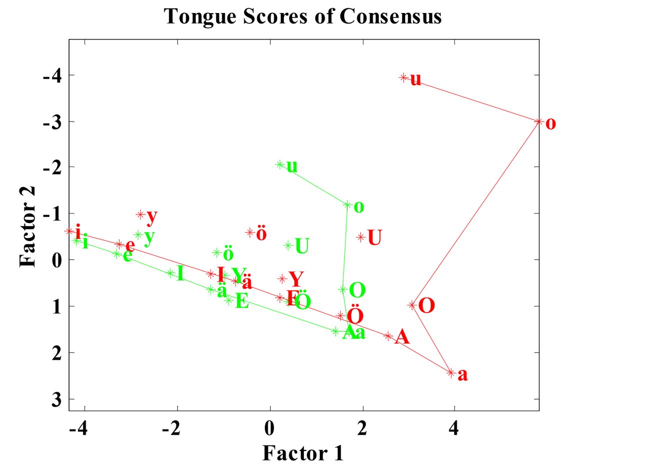


Figure 5: Upper left panel: 2 factorial solution of modelled data, Factor 1: front raising, Factor 2: back raising. Upper right panel: Factor space for stressed and unstressed vowels. Lower left panel: Speaker-specific straincrosses

Unstressed vowels show an elevated tongue tip (/i, e, y/) or more fronted articulations. ➔ greater degree of coarticulation

Straincrosses indicate a high amount of speaker-dependent affine transformation for articulatory spaces compared to acoustical spaces

3. Relationsship Anatomy and Speaker-dependent Models

- ❑ Anatomical Measure: Doming index calculated as ratio between the total midsagittal length of individual EPG palates and the distance of the first and the last point on the palate. Higher values indicate a palate with a higher degree of doming.

Subject	CG	CM	DF	PJ	JD	RW
Trace length (mm)	50.17	52.86	41.09	50.31	45.22	44.51
Trace distance (mm)	47.35	47.18	38.24	49.49	41.43	41.30
Doming ratio	1.06	1.12	1.07	1.02	1.09	1.08

- ❑ The bootstrap canonical correlation between the first 2 deformation eigenvalues and doming ratio accounts for 91% of the total variance
- ❑ One outlier (subject RW) decreases the canonical correlation for the whole sample to 74%
- ❑ This suggests that the Procrustes method primarily removes anatomical variance from the data. Inter-individual differences are touched, but to a lesser degree.

SUMMARY

- ❑ Speaker-specific **formant** values are mapped to the consensus object by **rigid** scaling and rotation.
- ❑ **Lingual** tongue positions during vowels have to be **affine**ly stretched and shrunk to a much greater degree than formant values.
- ❑ For five of six speakers speaker-dependent differences in stress production can be related to their palate shape.
- ❑ Stressed vowels are produced by less coarticulation.

CONCLUSIONS

- Generalized Procrustes Analysis provides a useful means for modeling speaker-independent strategies of stress produciton.
- Not all speaker-dependent differences are related to different palate shapes.
  ➔ Differential behaviour independent of biomechanical or anatomical properties
  OR
  ➔ other structures (not captured with palate shape parameters) might also play an important role (e.g. pharynx length or shape, jaw contribution etc.)
- Therefore Generalized Procrustes Analysis is sensitive to heterogeneous samples.

Acknowledgements

Special thanks to Suse Fuchs, Dirk Fischer and Daniel Pape for help with analyzing the data. Work supported by German Research Council (DFG) grant GWZ 4/5-2(A1).

References

Disner, S. (1980) Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America* 67, 253-261.

Gower, J. (1975). Generalized Procrustes analysis. *Psychmetrika* 40, 33-51.

Rohlf, F. & D. Slice. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* 39, 40-59.