

The effect of dynamic acoustic cues on age classification

Luciana Albuquerque^{1,2,3,4}, António Teixeira^{1,3}, Catarina Oliveira^{1,5}, Daniela Figueiredo^{2,5}



¹ Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal;
² Center of Health technology and Services Research, University of Aveiro, Aveiro, Portugal
³ Department of Electronics Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal
⁴ Department of Education and Psychology, University of Aveiro, Aveiro, Portugal
⁵ School of Health Science, University of Aveiro, Aveiro, Portugal

lucianapereira@ua.pt

Introduction

- European Portuguese (EP) vowels have been traditionally characterized by static cues
 - e.g., F1 and F2 obtained at vowel midpoint
- **But Speech is inherently dynamic**
- Some studies use dynamic cues in a classification task
 - e.g., to predict dialect, stress and vowel

Objective

- Determination of the **potential usefulness of the dynamic proprieties** of EP vowels in two classification tasks:
 - ☐ classification into age groups
 - binary classification: cut-off at 65 years
 - ☐ classification of vowel for two age groups (senior/non-senior)

Method

- **Speech corpus:**
 - **36** real **disyllabic words** (mostly CV.CV) (4 by vowel)
 - **9 vowels:** [i], [e], [ɛ], [a], [ɔ], [o], [u] in stressed position; [i] and [e] in unstressed position
 - Produced by **112 healthy native** Portuguese **speakers** aged between 35 and 97 (56 ♂ and 56 ♀)
 - 112 participants x 36 words x 3 repetitions=12096 recordings
- **Acoustic features:**
 - Vowel duration
 - F1, F2, and F3:
 - “optimal ceiling” of the vowel mid point was used
 - 35 time points within the central 60% of each token
 - z-score normalization
 - F1, F2, and F3 trajectories:
 - Discrete Cosine Transform (DCT) applied to the 35 formant values keeping 1st 4 coefficients (C0-C3)
 - Formant Trajectory Length (TL)
- **Classifiers:**
 - Linear Discriminant Analysis (LDA)
 - Classification and Regression Trees (CART)
 - 10-fold cross-validation
 - Deep Neural Networks (DNN)
 - 80% for training and 20% for test

Results

- **LDA:**
 - Static cues for F1, F2 and F3: sensitivity **0.22**
 - Gradually adding the other DCT coefficients (C1 to C3) as well as TL: sensitivity **0.26**
 - Static and dynamic cues and duration: sensitivity **0.45**
 - All experiments: specificity was above 0.7
- **CART:**
 - **Age classification:** (1) **higher error rate when using only static cues**; (2) error reduction by including additional DCT coefficients; (3) error reduction when adding TL; (4) great duration impact in error reduction

	Parameters	DCT coefficients	Error	1 st predictor	2 nd predictor	3 rd predictor
Static	F1, F2	C0	0.483	F1_C0	F2_C0	—
	F1, F2, F3	C0	0.478	F1_C0	F3_C0	F2_C0
Dynamic	F1, F2, F3	C0, C1	0.460	F1_C0	F2_C1	F1_C1
	F1, F2, F3	C0, C1, C2	0.448	F1_C0	F2_C0	F2_C2
	F1, F2, F3	C0, C1, C2, C3	0.440	F1_C0	F1_C1	F2_C0
	F1, F2, F3; TL	C0, C1, C2, C3	0.438	F1_C0	F2_C1	F3_C1
	F1, F2, F3; TL duration	C0, C1, C2, C3	0.406	duration	F1_C0	F2_C0

- **Vowel classification:** (1) error values approx. 0.8 with only static versus approx. 0.40 with all DCT coefficients and no duration; (2) with TL and duration the error attains approx. 0.35; (3) **the best 3 predictors when using a mix of cues only include the dynamic ones**
- **DNN:**
 - Simple topology: 2 hidden layers of 40 units each

Input features	Sensitivity (train)	Sensitivity (test)
All static + dynamic	74.2	33.6
Only static	5.3	5.5

- **Only static features as input gave much worst results**

Conclusion

- Features related to the dynamics of formants F1-F3 result in better age classification performance, somehow supporting the hypothesis of relevant information existence regarding age in the acoustical dynamic cues
- The importance of duration, identified in our previous related work (Albuquerque et al., InterSpeech 2019), was reconfirmed as an important predictor of age

This research was financially supported by the project Vox Senes POCI-01-0145-FEDER-03082 (funded by FEDER, through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI), and by national funds (OE), through FCT/MCTES), by the grant SFRH/BD/115381/2016, and by IEETA (UID/CEC/00127/2019).