Effects of acoustic characteristics on dysarthric speech intelligibility

Wei Xue, Catia Cucchiarini, Roeland Van Hout, Helmer Strik
Center for Language Studies, Radboud University, the Netherlands

Background

- Subjective ratings of speech intelligibility are time consuming and the methods applied often differ between studies.
- Need for objective acoustic measurements of speech intelligibility

-> Exploring interpretable objective measurements that are correlated to speech intelligibility ratings

Databases

Table 1: The distributions of speakers in three databases.

Number of speakers in database	Dysarthric speaker			Reference speaker			In total
	Male	Female	Total	Male	Female	Total	
TORGO [1]	5	3	8	4	3	7	15
IS2016 [2]	7	0	7	4	1	5	12
COPAS [3]	29	20	49	33	48	81	130

▪Intelligibility ratings of speakers

- TORGO - rated on sentence level for each dysarthric speaker using Frenchay Dysarthria Assessment (FDA, a-e has been transformed to 1-9); reference speakers were set as fully intelligible.
- IS2016 – rated on sentence level using Visual Analog Scale (VAS) and averaged for each dysarthric speaker; reference speakers were set as fully intelligible.
- COPAS – calculated by percentage of correct target phonemes using wordlists in Dutch Intelligibility Assessment (DIA) for each speaker (dysarthric and reference).

Table 2: Advantages and disadvantages of each database.

Database	Advantage	Disadvantage
TORGO / IS2016	Global subjective intelligibility ratings	Limited number of speakers
COPAS	The large number of speakers	Intelligibility ratings were only on phoneme level

Method

▪ Acoustic measures

- Vowel features extraction using VowelTriangle [4]
  - Vowel Space Area (VSA)
  - Distance between corner vowels and the centroid of the VSA
  - Mean frequency of formant 1 (f1) and 2 (f2) of corner vowels -> used to calculate Vowel Articulation Index (VAI) and Formant Centralization Ratio (FCR)
- Global features related to fundamental frequency, intensity and formants.

▪ Correlations

Table 3: Details of acoustic features.

VowelTriangle	Global features
Area2 (mean + 2 standard deviation)	pitchMin (minimal pitch)
Area1 (mean + 1 standard deviation)	pitchMax (maximal pitch)
i.dist (distance between /i/ and centroid)	pitchMean (mean of pitch)
a.dist (distance between /a/ and centroid)	pitchStd (standard deviation of pitch)
u.dist (distance between /u/ and centroid)	pitchVar (variability of pitch)
VTL (Vocal Tract Length)	IntMin (minimal intensity)
Intensity	IntMax (maximal intensity)
Slope (of F0)	IntMean (mean of intensity)
meanf1_a	IntStd (standard deviation of intensity)
meanf1_i	f1
meanf1_u	f2
meanf2_a	f3
meanf2_i	f4
meanf2_u	Center_gravity
VAI / FCR (based on f1/f2 of vowels)	

Results

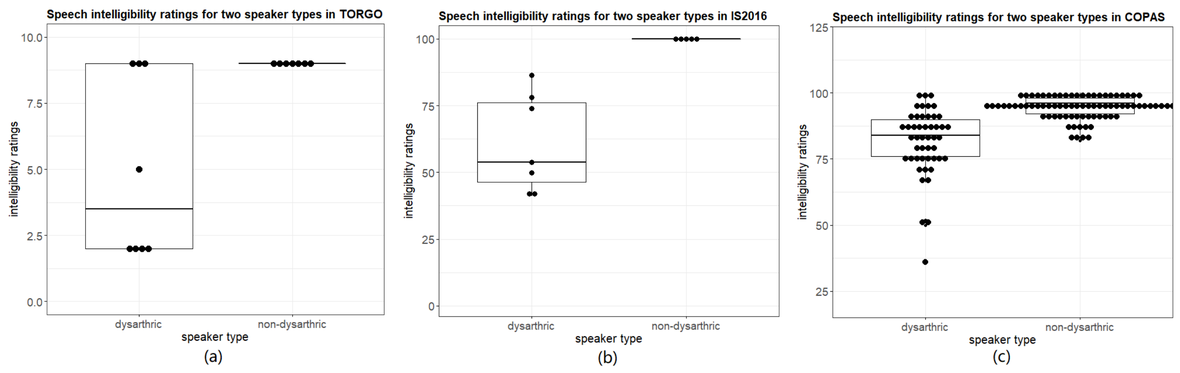

Figure 1: Boxplots of intelligibility ratings for two speaker types in (a) TORGO, (b) IS2016 and (c) COPAS.

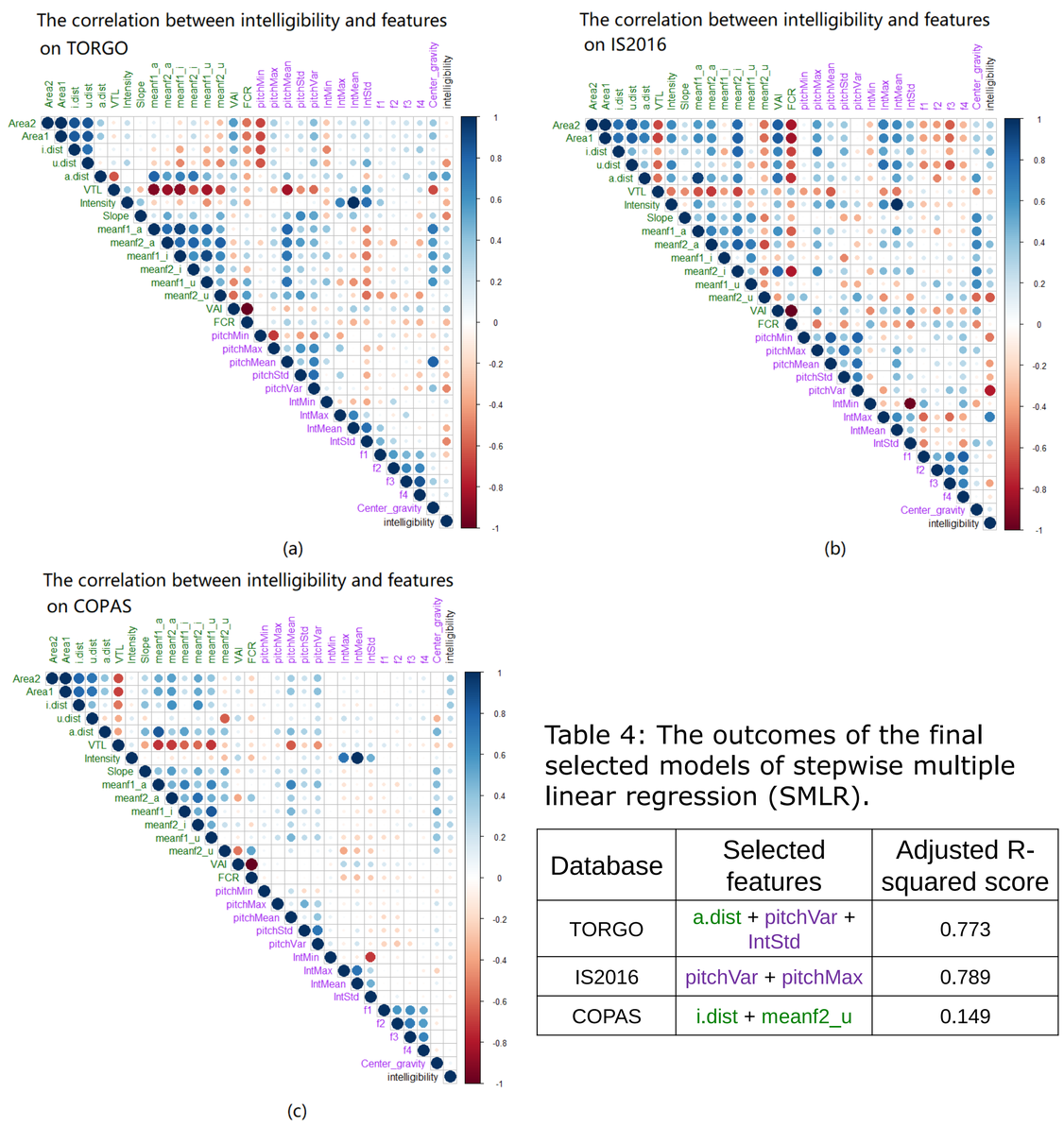

Table 4: The outcomes of the final selected models of stepwise multiple linear regression (SMLR).

Database	Selected features	Adjusted R-squared score
TORGO	a.dist + pitchVar + IntStd	0.773
IS2016	pitchVar + pitchMax	0.789
COPAS	i.dist + meanf2_u	0.149

Figure 2: Correlation plots between intelligibility ratings (black) and features (vowel features – green and global features - purple) on (a) TORGO, (b) IS2016 and (c) COPAS.

Discussion and Conclusion

▪ Discussion

- The correlations between global features and intelligibility in TORGO and IS2016 were generally stronger than those in COPAS. One of the explanations could be that the first two corpora adopted global measures of intelligibility, while in COPAS ratings on the phoneme level were used.
- The distance between the corner vowel /u/ and the centroid of VSA is positively correlated with speech intelligibility in IS2016 and COPAS, but the opposite pattern is found in TORGO. This may be due to the different languages used in the databases. Dutch and Flemish (a variety of Dutch) were used in IS2016 and COPAS, respectively, while English was used in TORGO.
- Vowel Space Area is positively correlated with intelligibility in IS2016 and COPAS, but the distance between the corner vowels and the centroid of VSA plays a more important role than the size of VSA, as shown in the selected features of SMLR on TORGO and COPAS.
- Pitch variability is negatively correlated with intelligibility in TORGO and IS2016, which means increasing pitch variability leads to decreased intelligibility. This could be explained as a deficit in the mechanisms of pitch control. The reason for the lack of a correlation in COPAS is similar to that in 1.

• Conclusion

- Vowel space and pitch variability appear to be related to intelligibility ratings on different levels.
- The positions of corner vowels could be potential predictors of intelligibility ratings.
- This kind of research is made very difficult by the fact that different corpora and different studies apply different ratings. It would be a good idea to agree on a number of standard subjective ratings, maybe at different levels, so as to facilitate future research.
- The important insights that derive from this comprehensive research may also help analyze speech intelligibility in the field of second language learning, with potential overlaps and generalizations.

Reference

[1] F. Rudzicz, A. K. Namasivayam, T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria," *Language Resources and Evaluation*, vol. 46, no. 4, pp. 523-541, 2012.
[2] M. Ganzeboom, M. Bakker, C. Cucchiarini, and H. Strik, "Intelligibility of disordered speech: Global and detailed scores," in *Proc. INTERSPEECH*, pp. 2503–2507, 2016.
[3] C. Middag, "Automatic analysis of pathological speech," Dissertation. Ghent University, 2012.
[4] R. van Son, C. Middag, and K. Demuynck, "Vowel Space as a Tool to Evaluate Articulation Problems," In *INTERSPEECH*, pp. 357-361, 2018.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 766287