The relation between acoustic-phonetic properties and speech intelligibility in noise across languages and talkers

Sabine Hochmuth[1,2], Birger Kollmeier[2], Barbara Shinn-Cunningham[1]

[1] *Auditory Neuroscience Lab, Boston University, Boston/MA, USA, E-Mail: sabineh@edu.bu*

[2] *Medical Physics and Cluster of Excellence Hearing4all, Universität Oldenburg, Oldenburg, Germany*

Introduction

This study explored the role of acoustic-phonetic properties on talker- and language-specific differences in the intelligibility of speech in noise. Matrix sentence speech was recorded by bilingual talkers of German/Spanish, Spanish/English, and English/German, as well as by monolingual talkers of each language. We analyzed how acoustic-phonetic speech features varied across languages, looking both at inter-individual (across talker) as well as intra-individual (within talker) variability. We then measured intelligibility of the matrix sentence speech in speech-shaped noise for normal-hearing, native listeners in each language. Matrix speech has the advantage of being similar in intelligibility across languages, as it uses the same size, closed stimulus set and has the same semantic and syntactic information in all languages. Various acoustic-phonetic parameters, previously identified as affecting speech intelligibility (including speaking rate, vowel space area, and energy in the mid-frequency region), were determined for each individual talker, in each language.

Methods

Speech material: Sentences of the American English, German, and Spanish matrix sentence tests (see [1]) spoken by accent-free bilingual and monolingual talkers of the respective languages were used. Besides the already existing recordings of the monolingual original test talkers and bilingual German/Spanish talkers presented in [2] further recordings were made with accent-free bilingual German/English, English/Spanish talkers, and monolingual talkers in each language. Each group of talkers (monolingual and bilinguals) consisted of four different talkers (2 male, 2 female each) so that each language was finally represented by 12 different talkers.

Average long-term frequency spectra were very similar across languages within the same talker. Across talkers, larger differences were observed. Amplitude-modulation spectra were very similar within the same talker speaking German and English. In Spanish amplitude modulation spectra showed higher regularity, i.e., higher modulation depths, at modulation frequencies reflecting syllable rates compared to English and German.

Speech intelligibility measurements: Speech recognition curves (SRT at 50% intelligibility and slope at this point) were determined for each talker in a stationary speech-shaped noise (ICRA1-noise, [3]) by adaptively measuring SRTs of 20% and 80% intelligibility. Five German, ten English and 7 Spanish listeners participated in these pilot intelligibility measurements. All listeners were native in the respective testing language and had normal hearing abilities (pure-tone thresholds not exceeding 20 dB HL).

Acoustic-phonetic properties: Five acoustic-phonetic properties described in the literature as probably being related to speech intelligibility (e.g. [4], [5] and [6]) were investigated in this study:

- *Speaking rate:* Number of syllables per second calculated over whole speech material per talker (90 – 100 sentences);
- *Pitch range:* Range in semitones between 0.1- and 0.9-quantile of the fundamental frequency within a sentence, averaged across all sentences per talker;
- *Speech-weighted SNR:* Calculated at an unweighted SNR of 0 dB using the band importance function of the Speech Intelligibility Index (SII) for speech in noise;
- *Vowel space area:* Euclidian area in the F1–F2 plane covered by the vowels /i/, /a/ and /u/. Formant frequencies were converted to mel;
- *Modulation depth:* Calculated as proposed in [7] for octave bands (250Hz–8kHz) and broadband signals. Mean and peak values across modulation frequencies between 1 and 32 Hz were takes as measures.

Results

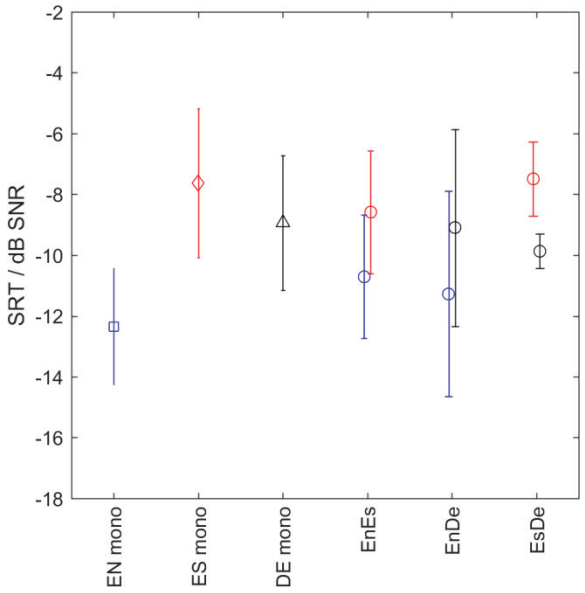

Figure 1: Mean SRTs across talkers within different groups of talkers (monolingual English (blue), German (black), Spanish (red); bilingual English-Spanish, English-German, Spanish-German).

Preliminary SRTs of 50% intelligibility averaged across talkers within each group are shown in Figure 1. On average Spanish was found to be least intelligible with an average SRT across all talkers of -7.9 dB SNR, followed by German with a mean SRT of -9.3 dB SNR, and English with a mean SRT -11.4 dB SNR. Variation in SRT across talkers was up to 9 dB in English, up to 8 dB in German and up to 7 dB in Spanish.

Spearman's rank correlations were calculated to assess rank-order relations between SRTs and each acoustic-phonetic property within each language (see Rho and p-values in table 1). Significant correlations are printed in bold face.

Table 1: Spearman's correlation coefficient (Rho) and probabilities of Acoustic-phonetic properties and speech recognition thresholds (SRTs). Mean and peak modulation depth for broadband.

Acoustic-phonetic property	English		German		Spanish	
	Rho	p	Rho	p	Rho	p
Speaking rate	0.59	**0.046**	0.12	0.712	0.54	0.075
Pitch range	-0.50	0.099	-0.30	0.341	-0.80	**0.003**
Speech-weighted SNR	-0.71	**0.013**	-0.85	**<0.001**	-0.87	**<0.001**
Vowel-space-area	-0.71	**0.013**	-0.72	**0.008**	-0.78	**0.004**
Mean modulation depth	-0.42	0.177	0.34	0.285	-0.17	0.604
Peak modulation depth	-0.34	0.287	-0.13	0.068	0.16	0.619

Speech-weighted SNR and vowel-space area were found to be significantly correlated with SRTs in each language. Higher amount of energy in the mid-frequency region of the long-term spectrum and a larger vowel-space area (larger jaw opening and extremer tongue movements) led to better intelligibility in all the investigated languages. Pitch range correlated significantly with SRTs only for Spanish, i.e., the larger the variation in fundamental frequency the better intelligible was a talker in Spanish. Speaking rate correlated significantly with intelligibility in English and a tendency could be observed in Spanish in that a lower speaking rate led to lower SRTs.

Conclusions

Talker-specific as well as language-specific effects were observed in intelligibility. Differences in intelligibility were larger across talkers (max. 9 dB) than across languages (max. 3.5 dB). Spanish showed consistently highest SRTs, while in English consistently lowest SRTs were observed. Energy in the mid-frequency regions of the long-term frequency spectrum and vowel space area correlated with talker intelligibility in each language. Possible insights into language specific effects might be gained via acoustic-phonetic cues like pitch range and speaking rate, although clear conclusions on the effect of language in intelligibility cannot be drawn from the preliminary results yet. Phonological differences across languages might also have an impact on differences in intelligibility across the investigated languages.

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (DFG) Cluster of Excellence Hearing4all and Grant No. HO 6080/1-1, and the National Institutes of Health–National Institute on Deafness and other Communication Disorders Grant No. R01 DC013825. We thank Erick Gallun for providing a tool for analyzing the modulation spectra.

References

[1] Kollmeier, B., et al: The multilingual matrix test: Principles, applications, and comparison across languages: A review. International Journal of Audiology 54 Suppl.2 (2015), 3-15.

[2] Hochmuth, S., Jürgens, T., Brand, T., and Kollmeier, B.: Talker- and language-specific effects on speech intelligibility in noise assessed with bilingual talkers: Which language is more robust against noise and reverberation?. International Journal of Audiology 54 Suppl.2 (2015), 23-34.

[3] Dreschler, W., Verschuure, H., Ludvigsen, C., and Westermann, S.: Icra noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International collegiums for rehabilitative audiology. Audiology 40 (2001), 148-157.

[4] Picheny, M. A., Durlach, N. I., and Braida, L. D: Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. Journal of Speech and Hearing Research 28 (1985), 96-103.

[5] Bradlow, A.R., Torretta, G.M., and Pisoni, D.B.: Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Communication 20 (1996), 255-272.

[6] Hazan, V., and Markham, D: Acoustic-phonetic correlates of talker intelligibility for adults and children. Journal of the Acoustical Society of America 16 (2004), 3108-3118.

[7] Gallun, F., and Souza, P: Exploring the role of the modulation spectrum in phoneme recognition. Ear and Hearing 29 (2008), 800-813.