A cross-linguistic study of speech modulation spectra

Léo Varnet, Maria Clemencia Ortiz-Barajas, Ramón Guevara Erra, Judit Gervain, Christian Lorenzi

Abstract ID: 3aSC1

A cross-linguistic study of speech modulation spectra

Léo Varnet [1], Maria Clemencia Ortiz-Barajas [2], Ramón Guevara Erra [2], Judit Gervain [2], and Christian Lorenzi [1]*

[1] Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, CNRS, Paris, France.
[2] Laboratoire Psychologie de la Perception, Université Paris-Descartes, CNRS, Paris, France.
*contact: leo.varnet@ens.fr

1. Introduction

• Languages have been classified into **linguistic categories** such as **stress-timed vs. syllable-timed** or **Head-Complement (HC) vs. Complement-Head (CH)**.

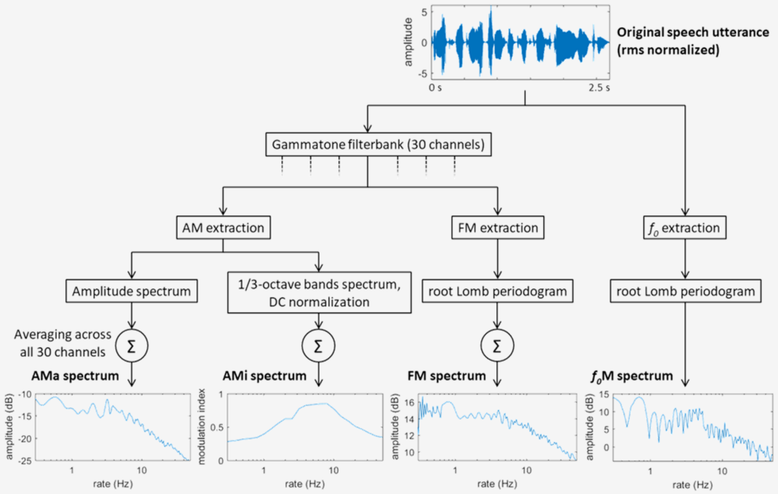
• It has been proposed that there may be correlations between these *linguistic* properties and some *acoustic* features of the speech signal, that young learners might use to break into language [2, 3].

• **Amplitude and Frequency Modulations (AM and FM)** have been shown to be of crucial importance for understanding speech. The modulation information contained in a given speech signal is typically characterized by the AM and FM spectra [4, 5].

• The aim of the present study was to determine whether different groups of languages can be distinguished on a purely acoustic basis in the modulation domain.

2. Methods

• **Calculating the modulation spectra**: for each utterance, 4 types of modulation spectra were computed



• **Statistical analysis**: conducted on relevant characteristics of the spectra (LF and HF slopes, maximum value, location of the peak). The comparison was done by means of a mixed model including a "language rhythm" factor (stress-timed vs. syllable- & mora-timed), a "basic word order" factor (HC languages vs. CH languages) and a random effect of speaker.

3. Read speech corpus

• 1797 utterances in 10 languages from 3 linguistic groups:
  - HC, stress-timed languages: **Dutch**, **English**
  - HC, syllable-timed languages: **French**, **Spanish**, **Polish**, **Zulu**
  - CH, syllable-timed languages: **Turkish**, **Basque**, **Marathi**, **Japanese**

The stimuli were **sentences read by 4 female native speakers of each language** (only 2 speakers for Marathi) [3].

• Overall, AMa, FM and f_0M spectra were **very similar** across languages and speakers. Their low-pass shape reflects the fact that speech signals mostly comprise slow temporal modulations.

• The AMi spectrum offers a more perceptually plausible representation of the AM information, emphasizing the medium- and high-rate regions compared to the low rates. All AMi spectra reach a maximum around 5 Hz ("syllable rate"), consistent with previous studies.

• The analysis the AMi spectra showed two significant effects of the linguistic factors: the **maximum value of the AMi peak distinguished between HC and CH languages**, while its exact **frequency position differed between stress-timed and syllable-timed languages**.

• No significant cross-linguistic differences in FM and f_0M spectra. Slight but significant differences were found in the 2-8 Hz region of the AMa spectra.

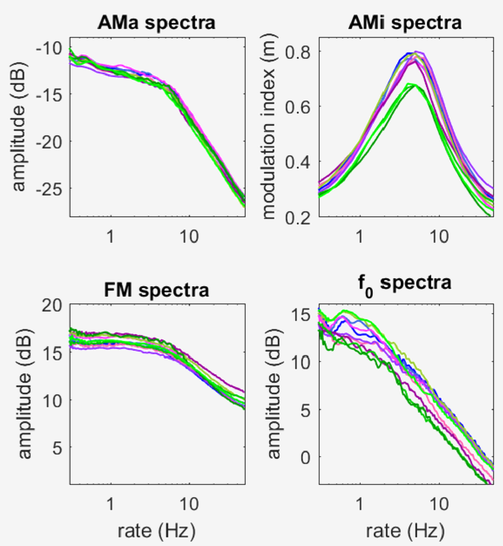


Fig. 2. Averaged modulation spectra for all languages of the read speech corpus. Blue lines: HC, stress-timed languages; indigo lines: HC, syllable-timed languages; green lines: CH, syllable-timed or mora-timed languages.

Acknowledgements:
This study was funded by the ANR grant "SpeechCode" (ANR-15-CE37-0009-01) to JG and ChL, the Human Frontiers Science Program Young Investigator Grant (RGY-0073-2014) to JG, ANR-11-0001-02 PSL and ANR-10-LABX-0087.

References:
[1] Cole, R., and Yeshwant M. (1994). "OGI Multilanguage Corpus LDC94S17," Philadelphia: Linguistic Data Consortium.
[2] Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). "A precursor of language acquisition in young infants," Cognition 29.
[3] Ramus, F., Nespor, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," Cognition 73.
[4] Sheft, S., Ardoint, M., and Lorenzi, C. (2008). "Speech identication based on temporal fine structure cues," JASA 124.

4. Semi-spontaneous speech corpus

• The initial analysis was based on a corpus using only 4 speakers per language and short, read sentences. In an attempt to generalize these results, we conducted a complementary analysis on a second corpus of **semi-spontaneous speech** [1] produced by a **large number of speakers** of **English**, **French**, **Spanish**, and **Japanese** (>100 per language). This corpus was split into a "strongly constrained" subset (short answers to closed questions) and a "weakly constrained" subset (longer answers to open questions).

• The cross-linguistic differences in AMi spectra were **successfully replicated on the strongly constrained subset**, indicating that the observed differences were not solely due to idiosyncratic differences such as speech rate. However, the comparison conducted on the weakly constrained subset showed that **with less controlled material, the cross-linguistic differences in AMi spectra disappeared**.

• The exploration of individual utterances suggests that the amplitude of the peak in the AMi spectrum is related to the rate of the most prominent envelope fluctuation in the speech signal, while the downward shift in peak rate for stress-timed languages originates from a greater occurrence of secondary peaks in the low-frequency region of the AMi spectrum.

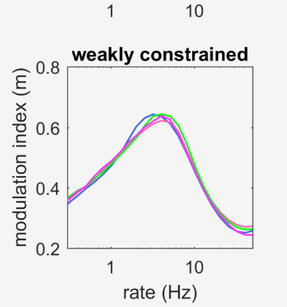


Fig. 3. Averaged AMi spectra for all languages of the semi-spontaneous speech corpus.

5. Conclusions

1. AM and FM spectra are highly similar across all investigated languages, when spectra are expressed in terms of absolute value.

2. When the AM spectrum is expressed in terms of "modulation index", a more perceptually-based metrics, 3 linguistic groups can be differentiated based on their AM content: CH languages, HC stress-timed languages and HC syllable-timed languages.

3. These findings persist for a larger number of speakers. Speaking style, however, has an influence on these acoustic differences that should be taken into account in future studies.