A cross-linguistic study of between-speaker variability in intensity dynamics in L1 and L2 spontaneous speech

Carolina Lins Machado
Leiden University
carolina@machado.eti.br

Introduction

Earlier research investigated dynamic aspects of the amplitude envelope in Zürich German [1], [2]. Intensity dynamics were characterized as the temporal displacement of acoustic energy associated to articulatory mouth opening (positive) and closing (negative) gestures. Results showed that negative dynamics explained more between-speaker variability than positive dynamics. The current study examined positive and negative intensity dynamics, in spontaneous speech produced by Dutch speakers in both their native language (L1) and their second language English (L2).

The underlying mechanisms of speech production in the L1 and in the L2 are both similar and diverging [3], [4]. While the similarities are found in the mechanical apparatus used during speech production [5], the differences are related to the complex mechanisms taking place before articulation, which are believed to be affected differently by the different languages [6]. Along with language-specific constraints [7], the anatomical characteristics of a speaker are believed to influence how she/he uses her/his speech apparatus [8]. Therefore, a combination of individual characteristics and language constraints may affect the acoustic signal differently between speakers.

Thus far, intensity dynamics were studied solely in one language. Therefore, this study set out to determine (i) whether intensity dynamics vary between speakers of another L1, namely Dutch; (ii) whether this variability is also persistent in L2 productions by the same speakers; and (iii) whether language has an overall effect on intensity dynamics.

Method

Informal monologues in the L1 and L2 of 51 female speakers were selected from the LUCEA corpus [9]. The speech was manually annotated orthographically by two annotators and checked by a third annotator. Next, duration of analysis chunks and intensity were normalized by language. For duration, the data was chunked to obtain uninterrupted speech segments of ca. 1.5 seconds, see [10]. For intensity, its curve was linearly normalized following the method in [2], to maintain only information related to the curve's trajectory, which can be associated to speaker-specific articulatory gestures. Prior to the normalization of the amplitude envelope, the data was prepared according to the steps in [1] to obtain an object containing intensity point values in time. The detection of amplitude minima and maxima was done semi-automatically.

Measures of positive and negative intensity dynamics (mean [MEAN], standard deviation [STDEV] and sequential variability [PVI]) were calculated and extracted following [1], where positive dynamics refers to the rate of intensity increase from a trough to the next peak and negative dynamics to the slope in time between a peak and the next trough. Subsequently, statistical analyses of the extracted measures were carried out. A factor analysis (FA) evaluated the orthogonality of positive and negative dynamics, i.e. whether they encode different information. Then, a multinomial logistic regression (MLR) assessed how much between-speaker variability is explained by each type of dynamics. Next, a linear discriminant analysis (LDA) evaluated how well speakers are discriminated based on positive and negative measures of intensity dynamics. Finally, the effect of language on intensity dynamics' acoustics was investigated employing linear mixed-effects (LME) models.

Results and discussion

First, there was inter-speaker variability in intensity dynamics in both languages. However, the results in [1] for the FA and the MLR were not fully replicated in the present investigation. A possible explanation lies in the fact that this study used spontaneous speech. In the L1, the positive measure of sequential variability had a strong positive correlation with its negative counterpart (r = .80). This was

interpreted as a greater degree of gestural overlap between the start and end of syllables in spontaneous speech.

Similarly, the MLR results indicated that for both languages positive and negative dynamics seemed almost equally able to explain inter-speaker variability (48-52%). Across languages, negative dynamics explained a slightly larger quantity of inter-speaker variability, following the previously proposed [1] reduced prosodic control over the mouth closing movement. Results of the LDA displayed a low speaker classification rate in the L1 and L2; negative measures of mean were better classifiers for both languages (L1 = 4.8%; L2 = 4.4%, chance level: 1.9%). The results of the LME (Table 1) revealed an effect of language on all measures of intensity dynamics, suggesting differences on the rhythmic aspects of Dutch and English.

Table 1. *Condensed results of the LME models' estimates (standard errors) and 95% confidence intervals explaining the effect of language on positive [+] and negative [–] measures of intensity dynamics (mean [MEAN], standard deviation [STDEV] and sequential variability [PVI]).*

	β_0 (Intercept)		β_1 (language = English)	
	Est. (SE)	*[95% CI]*	*Est. (SE)*	*[95% CI]*
MEAN_v_I[–]	3.23 (.05)	[3.14, 3.32]	– .17 (.03)***	[–.23, –.12]
STDEV_v_I[–]	1.61 (.03)	[1.55, 1.66]	– .07 (.02)*	[–.11, –.02]
PVI_v_I[–]	6.28 (.06)	[6.16, 6.40]	– .72 (.05)***	[–.83, –.62]
MEAN_v_I[+]	4.27 (.07)	[4.13, 4.41]	– .13 (.04)***	[–.21, –.05]
STDEV_v_I[+]	2.19 (.04)	[2.11, 2.28]	– .08 (.03)**	[–.14, –.02]
PVI_v_I[+]	6.34 (.06)	[6.23, 6.45]	– .68 (.05)***	[–.78, –.58]

Note: Significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, Bonferroni correction was applied.

Conclusion

Indexical characteristics in intensity dynamics were not restricted to the native language. Language had an effect on intensity dynamics' acoustics, indicating possible rhythmic differences between Dutch and English. Particularly, as in [1] and [2], negative intensity dynamics performed better than their positive counterparts in speaker discrimination methods and in explaining between-speaker variability.

Acknowledgment

This is the author's MA thesis, supervised by Dr. W.F.L. Heeren.

References

[1] He, L., & Dellwo, V. (2017). Between-speaker variability in temporal organizations of intensity contours. *The Journal of the Acoustical Society of America, 141*(5), 488-494.

[2] He, L., Zhang, Y., & Dellwo, V. (2019). Between-speaker variability and temporal organization of the first formant. *The Journal of the Acoustical Society of America*, *145*(3), 209-214.

[3] Levelt, W. J. M. (1999). Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 83-122). Oxford, England: Oxford University Press.

[4] Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah: Lawrence Erlbaum Associates, Inc.

[5] Hixon, T. J., Weismer, G., & Hoit, J. D. (2020). *Preclinical speech science: Anatomy, physiology, acoustics, and perception*. San Diego: Plural Publishing.

[6] Escudero, P. (2009). Linguistic Perception of ''similar'' L2 sounds. In P. Boersma, & S. Hamann (Eds.), *Phonology in perception* (pp. 151-190). Berlin: Mouton de Gruyter.

[7] Schwartz, G., & Kaźmierski, K. (2019). Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners. *Language Acquisition*, 1-28.

[8] Zuo, D., & Mok, P. P. K. (2015). Formant dynamics of bilingual identical twins. *Journal of Phonetics*, *52*, 1-12.

[9] Orr, R., & Quené, H. 2017. D-LUCEA: Curation of the UCU Accent Project Data. In Odijk, J., & van Hessen, A. (Eds.), *CLARIN in the Low Countries* (pp. 181-193). London: Ubiquity Press.

[10] Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628-639.