

Investigations on the speech spectrum of normal and loud speech

Corinna Harwardt

Fraunhofer FKIE, Wachtberg, Germany

`corinna.harwardt@fkie.fraunhofer.de`

Speaking up induces various modifications to acoustic characteristics of speech. In this paper we investigate the impact of raised vocal effort on the speech spectrum. In particular, we look at different spectral parameters and compare the changes we observe for each. The parameters we take into account are spectral tilt, spectral center of gravity, energy ratio and spectral moments. We carry out tests on the complete data set with all phonemes pooled into one distribution and tests with the data divided into the three phoneme classes vowels, sonorants and obstruents.

For our investigation we used the Oldenburger Logatome (OLLO) speech corpus (Wesker, 2005) which consists of data from 40 German speakers with four dialects (standard German, Bavarian, Eastfrisian, East Phalian). The available vowel phonemes are /a, a:, e, ε, i, ɪ, o, ɔ, u, ʊ, ə/. Furthermore /b, p, t, g, k, f, v, s, ʃ, z/ are the available obstruents and /l, m, n/.

To calculate spectral tilt (ST) we computed the regression line of the speech spectrum. The center of gravity (COG) is defined as in Son et al. (Son, 1999). The computation of the energy ratio was done as proposed by Wenndt et al. (Wenndt, 2002). Spectral moments (Mom1-Mom4) have been calculated according to Forrest et al. (Forrest, 1988). While examining the results, a special focus will be set on the comparison of spectral tilt and third moment as well as on center of gravity and first moment, because those parameters describe the same characteristics of the speech spectrum.

To compare spectral parameters of normal and loud speech we calculated the mean values of each parameter for both kinds of vocal effort (see Table 1). We marked those values with modifications of mean values greater than 10%. When we look at the complete distribution, which includes data from all phoneme classes, we see that all parameters, except the second spectral moment, are changed more than 10%. For sonorants and vowels we observe the same trend. The values of these two classes are all changed more than 10%. The obstruents are affected much lesser. For obstruents only spectral tilt as well as the third and forth spectral moment suffer from greater modifications.

To verify our observations we carried out Mann-Whitney tests (Hollander, 1999). Assuming $\alpha = 0.05$ ($z = 0 \pm 1.95996$) as significance level, we observe significant differences for all classes and parameters (see Table 2). Although we didn't find greater differences for the mean values of most spectral parameters of obstruents we monitor significant changes. Comparing the z-values of the different classes we see that the obstruents' values are much smaller than those of the other classes. Hence, we conclude that obstruents are less affected by changes of vocal effort than sonorants and vowels. Especially for vowels the z-values are great.

The parameters less affected are energy ratio and second spectral moment. Comparing the center of gravity and the first spectral moment we don't monitor great differences, whereas the comparison of spectral tilt and third moment shows great differences. A detailed analysis of results can be found in (Harwardt, 2011).

Table 1: Average mean values for different spectral parameters for normal (N) and loud (L) speech over the complete distribution as well as the vowel, sonorant and obstruent distributions. Changes of 10% or more are marked with an arrow downwards for decreasing values and upwards for increasing values.

		Complete Distribution	Vowels	Sonorants	Obstruents
ST	N	-40.915 ↑	-64.345 ↑	-59.051 ↑	-2.635 ↑
	L	-34.184	-50.272	-36.214	-1.883
COG	N	1123.806 ↑	916.068 ↑	733.178 ↑	1470.844
	L	1260.368	1203.671	1026.592	1381.986
ER	N	0.705 ↑	0.401 ↑	0.258 ↑	1.210
	L	0.794	0.576	0.448	1.243
Mom1	N	1131.599 ↑	923.876 ↑	740.988 ↑	1478.613
	L	1268.147	1211.479	1034.401	1389.707
Mom2	N	821856.848	748854.9 ↑	688175.7 ↑	943691.6
	L	863812.073	839627.5	950578.3	909285.8
Mom3	N	1.482 ↓	1.934 ↓	1.875 ↓	0.740 ↑
	L	1.182	1.301	1.136	0.944
Mom4	N	4.163 ↓	5.690 ↓	3.276 ↓	1.738 ↑
	L	2.429	2.353	0.400	2.651

Table 2: Z-values obtained by the Mann-Whitney test for different spectral parameters over the complete audio data as well as the vowel, sonorant and obstruent distributions.

Parameter	Complete Distribution	Vowels	Sonorants	Obstruents
ST	15,9159	35,1518	10,9377	9,4189
COG	25,3868	48,2942	10,3844	-6,5763
ER	11,7371	25,007	3,7133	-2,7387
Mom1	25,3836	48,2937	10,3844	-6,5796
Mom2	9,8721	19,4614	9,1243	-4,2063
Mom3	-17,0732	-32,1066	-9,952	5,2644
Mom4	-16,7922	-27,3897	-9,8523	7,3907

References

- Wesker, T., Meyer, B., Wagener, K., Anemüller, J., Mertins, A., Kollmeier, B. (2005): Oldenburger logatome speech corpus (OLLO) for speech recognition experiments with humans and machines. Interspeech.
- Son, R. J. J. H. ; Pols, L. C. W.(1999): An acoustic description of consonant reduction. In: Speech Communication 28, p. 125–140.
- Wenndt, S. J. ; Cupples, E. J. ; Floyd, R. M.(2002): A Study on the Classification of Whispered and Normally Phonated Speech. In: International Conference on Spoken Language Processing.
- Forrest, K., Wiesmer, G., Milenkovic, P., Dougall, R. (1988): Statistical analysis of word-initial voiceless obstruents: Preliminary data. *J. Acoust. Soc. Am.* 84(1), p. 115-123.
- Hollander, M., Wolfe, D. A. (1999), "Nonparametric Statistical Methods", John Wiley Sons, Inc..
- Harwardt, C. (2011): "Comparing the Impact of Raised Vocal Effort on Various Spectral Parameters", Interspeech.