Investigating Robustness of Spectral Moments on Normal- and High-Effort Speech

Frederike Gottsmann, Corinna Harwardt

Fraunhofer-Institut für Kommunikation,
Informationsverarbeitung und Ergonomie FKIE, Germany
frederike.gottsmann@fkie.fraunhofer.de, corinna.harwardt@fkie.fraunhofer.de

Abstract

In this paper we are looking for a robust value of the spectral moments that does not change when a speaker varies his vocal effort from normal to loud speech. To do this we first calculate the first four spectral moments for normal and loud speech. Then we compare the results for each single phoneme. After this, we do a correlation analysis to check whether normal and loud speech are linked with each other linearly. The results of the investigations show that plosives and fricatives are robust to changes of vocal effort. Vowels and sonorants demonstrate significant differences in vocal effort.

Index Terms: vocal effort, spectral moments

1. Introduction

In general, the term vocal effort describes the quantity of voice a speaker uses when he adopts it to the communication situation. Changing vocal effort means that the speaker varies his voice from normal to soft/ whispered speech or from normal to loud/ shouted speech. Whispering is associated with low vocal effort, whereas loud speech corresponds to high vocal effort. Those changes might be induced for example by the presence of background noise, by varying the distance or by emotions.
Several studies have been made by different authors which investigated the impact of vocal effort changes on glottal, articulatory and acoustic parameters [6], [5]. Unfortunately the results over the different studies are not always comparable to each other due to different preconditions, e.g. gender of speaker or setting of recordings. In this investigation we just look at those studies which focus on spectral changes. Several of those studies found a shift of frequencies from lower to higher frequency bands [4]. Often investigated spectral parameters are spectral center of gravity [4], [2], spectral tilt [6], [2] or spectral emphasis [7]. The spectral moments as proposed by Forrest et al. [1] have not been part of an investigation of vocal effort changes, as far as we know. Hence, we decided to analyze the impact of raised vocal effort on spectral moments.
In this paper we describe the results of this analysis. At first we describe the methods we used in section 2. This incorporates presentation of the audio data we used, specification of the calculation of spectral moments and definition of the statistical tests needed for evaluation of results. Next we present the results in section 3. We show average spectral moments for each different phoneme and the associated t-tests. For deeper analysis of the relationship between normal and loud speech we do a correlation analysis. After presentation of results we draw a conclusion (see section 4).

2. Method

In the next sections we describe the preconditions for our investigations. First we present the corpus we used. Then we demonstrate our calculations and the method for statistical analysis.

2.1. The OLLO corpus

The Oldenburger Logatome (OLLO) speech corpus contains data of 40 German speakers which are spoken in four dialects: standard German (speaker from university population), Bavarian (speaker from Munich), East Frisian (speaker from Oldenburg) and Eastphalian (speaker from Magdeburg). The speakers articulate different vowel-consonant-vowel or consonant-vowel-consonant phoneme combinations without any semantic information. These phonemes are the vowels /a, aː, e, ɛ, i, ɪ, o, ɔ, u, ʊ, ə/, the plosives /b, p, d, t, g, k/ and the fricatives /f, v, s, ʃ, z/. As sonorants /l, m, n/ were available in the corpus. The database includes speaker characteristics like age, gender, dialect and speaker-dependent variabilities like speaking rate, speaking effort, speaking style [8]. For our investigations we used normal- and high-effort speech samples of 19 male speakers.

2.2. Spectral moments

The speech spectrum represents the signal's energy at different frequencies. Therefore the speech signal's energy distribution changes, when a speaker raises his voice. To investigate the changes on the spectrum induced by raised vocal effort, we have a closer look at the energy distribution of the spectrum. First we compute a 512-point fast Fourier transform. To analyze the distribution of the power spectrum, we compute the first four moments of the spectral distribution, which characterize the distribution of the power spectrum very well. The first moment (m1) is associated with the mean value of the spectral distribution and the unit is hertz (Hz). Whereas the second moment (m2) is linked with the variance around first moment. The third moment (m3) represents the skewness and shows the symmetry of the distribution. Positive skewness indicates a spectral tilt where we can find a higher concentration of energy in the lower frequency ranges. Negative skewness means a spectral tilt with a higher concentration of energy in the higher frequency ranges. Kurtosis is the fourth moment (m4) of the spectral distribution. Positive values show a compact spectrum, whereas, negative values indicate a flatter spectrum [3]. The units of skewness and kurtosis are dimensionless. The spectral moments were computed according to [1].

28 – 31 August 2011, Florence, Italy

2.3. Statistical Analysis

We accomplished a two-tailed t-test. We decided to use a critical value of $\alpha = 0.01$ which leads to considerable difference between normal and loud articulation. If the z-value in Table 2 is not located within the interval 2.58 to -2.58, we refuse the null hypothesis. In our case, the null hypothesis is true when spectral moments equal for normal and loud speech. Corresponding the alternative hypothesis says that both conditions values differ. In the case that we find significant differences, we assume a linear relation between the two vocal effort conditions. Therefore a Pearson product-moment correlation coefficient is calculated to check whether normal and loud speech correlate.

3. Results

For each single phoneme we calculated average values for the four moments for normal and loud articulation. We are looking for values that do not change between normal- and high-effort. A two-tailed t-test was used to recognize significant differences between these conditions. Then, we accomplished a correlation analysis to demonstrate an interrelation between normal- and high-effort.

3.1. Plosives

At first, we analyze the six phonemes /b, p, d, t, g/ and /k/. We start with the bilabial plosives /b/ as voiced and /p/ as voiceless phonemes. At the first, third and fourth moment the values for /b/ do not show any significant difference between normal and loud articulation (see Table 1). Hence, we have to maintain the null hypothesis. For the second moment there are significant differences between these two efforts. In our calculation we can see that the values for all four moments for normal articulation are higher than for loud (details see [**?**]). Normal values for the first moment are higher than loud values. Skewness is linked with a curve skewed to the right and kurtosis indicates a leptokurtic curve. The voiceless plosive /p/ has significant differences at all moments. For the first two moments of normal vocal effort, values are also higher than those for loud speech. However the results for the third and fourth moment illustrate another behavior. Skewness is positive for both efforts but normal values are lower. Likewise for kurtosis the curve is leptokurtic. In addition to these results, we perform a correlation analysis (see Table 2). As you can see both plosives /b/ and /p/ show positive values. The plosive /b/ includes the highest correlation between normal and loud articulation for the first moment. The lowest correlation between these two efforts has /p/ for the fourth moment. Next, we describe the alveolar plosives /d/ and /t/. The first two moments for the voiced /d/ show significant differences between normal and loud speech. Hence, we have to reject the null hypothesis. For these two moments the values for high-effort are higher than for normal-effort. On the other hand for the third and fourth moment, both conditions are not significantly different and hence, the null hypothesis has to be maintained. The normal values for the third moment are higher than those for loud speech. But both values are positive. Kurtosis shows a leptokurtic curve for both values but values for normal vocal effort are also higher. Similar results as for /d/ can be observed for the voiceless /t/ for the first, second and third moment. For the fourth moment we observe a platykurtic curve for both efforts. The correlation analysis illustrates a high correlation between normal and loud speech for the voiceless /t/ for the first moment. The lowest correlation between both efforts is evident for /d/ at the fourth moment.

As final plosives we look at the voiced velar /g/ and voiceless velar /k/. For /g/ the first, third and fourth moments illustrate significant differences between normal and loud articulation. Simply the second moment does not show any significant change. Loud values for the first two moments are higher than for normal articulation. On the other hand for the last two moments, normal values are higher. Skewness is positive, too, and kurtosis illustrates a leptokurtic curve for both. The voiceless /k/ does not present any significant difference for all four moments. Just for the first moment the loud values are higher than normal speech. For the other three moments, we observe higher values for normal-effort. A positive skewness and a leptokurtic curve are found for both, too. Here, the correlation analysis shows the highest correlation between normal and loud values for the voiceless /k/ for the first moment. This plosive indicates the lowest correlation between these two values for the second moment, too.

For all plosives, we can observe that /p/ shows always significant differences for all moments (see Table 2). The plosives /b, d, t, g/ and /k/ do not show any consistent differences for all moment. In the majority of cases the third and fourth moment do not illustrate any significant difference between normal and high vocal effort. We can also see that the first moment has the highest correlation for all phonemes between normal and loud articulation. When we look at individual phonemes we can perceive that /k/ has the highest correlation between normal and loud speech at the first moment (see Fig. 1). Hence, it could be that /k/ is most robust against changes in vocal effort for all four moments. However, we can say for the other plosives /b, d/ and /t/ that they are also robust against changes in vocal effort for the third and fourth moment.

Table 1: *Results of the t-tests for plosives for each spectral moment (m1-m4)*
s = significance, ns = no significance

	m1	m2	m3	m4
/b/	ns	s	ns	ns
/p/	s	s	s	s
/d/	s	s	ns	ns
/t/	s	s	ns	ns
/g/	s	ns	s	s
/k/	ns	ns	ns	ns

Table 2: *Correlation coefficient of plosives for each spectral moment (m1-m4)*

	m1	m2	m3	m4
/b/	0.7423	0.7112	0.6790	0.6158
/p/	0.6058	0.5536	0.4917	0.4103
/d/	0.7141	0.6347	0.6068	0.4388
/t/	0.7418	0.6450	0.7016	0.4769
/g/	0.8039	0.6056	0.7531	0.6069
/k/	0.8691	0.6034	0.8375	0.6430

3.2. Fricatives

In this section we investigate the fricatives /z, s, ʃ, v/ and /f/. First, the alveolar fricatives /z/ and /s/ are analyzed. The voiced fricative /z/ does not show any significant difference for

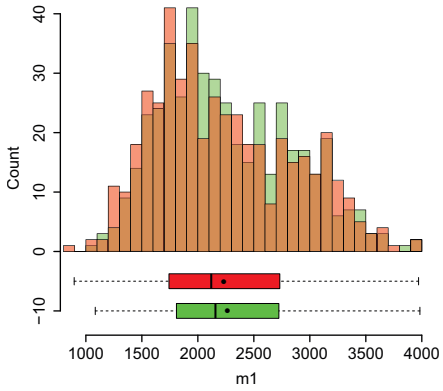

Figure 1: *Values for the first moment (m1) for loud (green) and normal (red) vocal effort for the plosive /k/*

the first, third and fourth moment (see Table 3). We can reject the null hypothesis only for the second moment. Average values for loud speech for the first and second moment are higher than average values for normal speech. For the third moment normal values are positive whereas loud values are negative. This illustrates a change in spectral tilt. Hence, the curve changes from a skew to the right to a skew to the left when a speaker changes his vocal effort from normal to loud speech. Kurtosis shows a platykurtic curve for normal as well as for loud articulation. The voiceless /s/ does not have any significant difference between normal and loud speech for the first and third moment. In contrast, the second and fourth moment offer significant differences between both efforts. The values for normal vocal effort for the first moment are slightly higher than the values for loud speech. We can observe higher values for loud articulation for the second moment. For the third moment, loud values are also higher but there is only a small difference to normal values. Both values are near null and negative skewness is linked with a curve skewed to the left. Values for the fourth moment are negative, too. This indicates a platykurtic curve for normal- and high-effort, in which normal values are higher. For these two fricatives the highest correlation between these two efforts can be found for the voiced /z/ of the second moment (see Table 4). The phoneme /s/ shows the lowest correlation between normal and loud articulation for the fourth moment.

Table 3: *Results of the t-tests for fricatices for each spectral moment (m1-m4)*
s = significance, ns = no significance

	m1	m2	m3	m4
/z/	ns	s	ns	ns
/s/	ns	s	ns	s
/ʃ/	s	ns	s	s
/v/	s	ns	s	s
/f/	ns	s	s	ns

Next we investigate the voiceless postalveolar fricative /ʃ/. The voiced variant /ʒ/ is not included in our corpus. The second moment does not illustrate a significant difference whereas for all other moments, we reject the null hypothesis. The first and fourth moment offer lower values for normal-effort than for

Table 4: *Correlation coefficient of fricative for each spectral moment (m1-m4)*

	m1	m2	m3	m4
/z/	0.8360	0.8601	0.7537	0.7205
/s/	0.7876	0.5959	0.7508	0.6043
/ʃ/	0.6826	0.8128	0.5601	0.5528
/v/	0.2287	0.6211	0.3210	0.2979
/f/	0.6314	0.8063	0.6139	0.5486

loud speech. Kurtosis displays a platykurtic curve for both values. For the second and third moment the values for normal speech are higher. We can also see a change in spectral tilt from normal to loud articulation, where the values for normal vocal effort are positive and the values for loud articulation are negative. The highest correlation between normal- and high-effort is for the second moment.
As labiodental fricatives we analyzed the voiced /v/ and voiceless /f/. The first, third and fourth moment for /v/ show significant differences between normal and loud articulation. For the second moment there are not any significant differences. We observe only for the first moment higher values for loud speech. Skewness is positive and linked with a curve skewed to the right. Kurtosis shows a leptokurtic curve. The voiceless /f/ has not any significant difference for the first and fourth moment. For the other two moments we reject the null hypothesis. The first and second moment illustrate lower values for normal effort. The third moment describes a positive skewness. Kurtosis is negative and therefore, we find a platykurtic curve for both efforts. The correlation analysis represents for each moment higher values for the voiceless /f/. The highest value is for the second moment. For the voiced fricative /v/ we observe the lowest value for the first moment.
In summary, we can see in Table 3 that the fricatives /z, s/ and /f/ do not show any significant difference for the first moment. The second, third and fourth moment have not always significant differences, too. In particular, it seems that for the first, third and fourth moment the fricative /z/ is robust against a change in vocal effort (see Fig. 2). There is also a high correlation between normal and loud articulation for these fricatives.

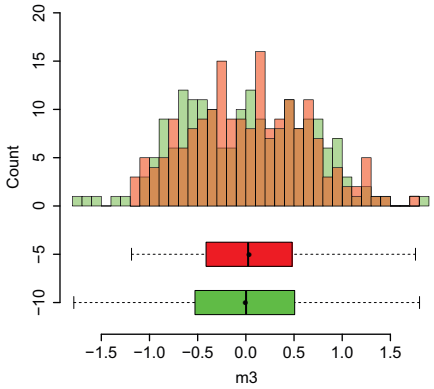

Figure 2: *Values for the third moment (m3) for loud (green) and normal (red) vocal effort for the fricative /z/*

3.3. Vowels

Vowels are the largest class of phonemes with the 11 single sounds /a, aː, e, ɛ, i, ɪ, o, ɔ, u, ʊ, ə/. At all moments all phonemes show significant differences between normal and loud articulation and hence, values for high vocal effort are always higher than for normal vocal effort. So, we can see, that they are not robust. There is a change in spectral energy when a speaker raises his vocal effort. Loud values for all phonemes for the first moment are higher than values for normal speech. As an example we can see in Fig. 3 the values for vowel /e/ for the first moment. This results we observe for the second moment, too. The third moment shows always a positive skewness and for all vowels the values for normal-effort are higher than for high-effort. Kurtosis indicates a leptokutic curve for all phonemes, where values for normal articulation are higher than for loud articulation. Correlation analysis demonstrates a positive correlation between normal and high-effort for all vowels, too. The highest correlation between these two efforts offers the central vowel /ə/ for the first, third and fourth moment. The vowel /ɛ/ possesses the highest correlation between normal and loud articulation for the second moment. In summary, we found that

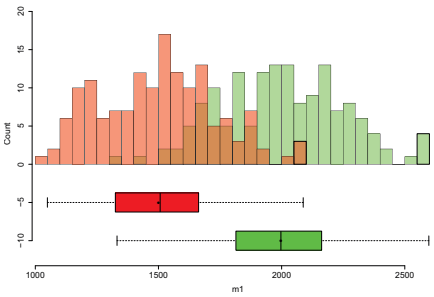

Figure 3: *Values for the first moment (m1) for loud (green) and normal (red) vocal effort for the vowel /e/*

vowels show no robust pattern when a speaker raises his vocal effort.

3.4. Sonorants

The sonorants /l, m/ and /n/ are the last phoneme class which is investigated. All moments show significant differences between normal and loud articulation. We observe for all sonorants lower values for normal speech than for loud speech for the first and second moment. In Fig. 4 we see an example for the first moment of the sonorants /l/. The values for the third and fourth moment for normal-effort are higher than the values for high-effort. For both moments, we observe positive values for normal and loud articulation. The correlation analysis for all sonorants is positive for all moments, too. These observations are comparable to those regarding the vowels. Hence we can assume that these values are not robust.

4. Conclusions

The aim of this investigation was to find out whether there is a robust value which is constant when a speaker changes his vocal effort from normal to loud speech. The results indicate that plosives and fricatives do not always show significant differences between normal and loud articulation. In particular, the plosive /k/ does not indicate any change for all moments. But also,

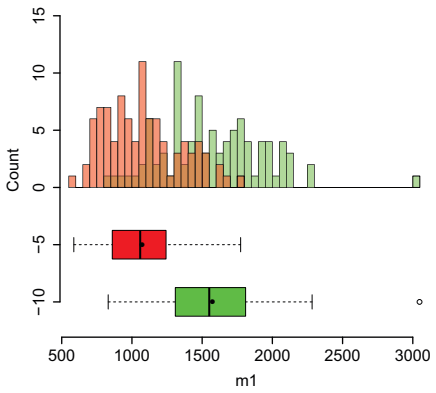

Figure 4: *Values for the first moment (m1) for loud (green) and normal (red) vocal effort for the sonorant /l/*

/b, d/ and /t/ have similar values for the third and fourth moment. The fricatives /z, s/ and /f/ are not significant for the first moment. This means for these plosives and fricatives, that they are consistent for each speaker when there is a change from normal to loud speech. Vowels and sonorants show significant differences and therefore the speech signal's energy distribution changes, when a speaker raises his vocal effort. Hence, we find no robust value for vowels and sonorants. In comparision to unvoiced phonemes we can see that voiced phonemes are less stable than unvoides ones. All phonemes and all moments show that the two efforts correlate with each other.

5. References

[1] Forrest, K., Wiesmer, P., Milenkovic, P. and Dougall, R, "Statistical analysis of word-initial voiceless obstruents: Preliminary data", J. Acoust. Soc. Am., 84(1):115–123, 1988.

[2] Junqua, J.C., "The Lombard reflex and its role on human listeners and automatic speech recognizers", J. Acoust. Soc. Am., 93(1):510–524, 1993

[3] Kardach, J., Wincowski, R., Metz, D.E., Schiavetti, N., Whitehead, R.L. and Hillenbrand, J. "Preservation of place and manner cues during simultaneous communication: a spectral moments perspective", Journal of Communication Disorders, 35 :533–542, 2002

[4] Liénard, J.-S. and Di Benedetto, M.-G., "Effect of vocal effort on spectral properties of vowels", J. Acoust. Soc. Am., 106(1):411–422, 1999.

[5] Schulman, R., "Articulatory dynamics of loud and normal speech", J. Acoust. Soc. Am., 85(1):295–312, 1989.

[6] Summers, W. van, Pisoni, D.B., Bernacki, R.H., Pedlow, R.I. and Stockes, M.A., "Effects of noise on speech production: Acoustic and perceptual analyses", J. Acoust. Soc. Am., 84(3):917–928, 1988.

[7] Traunmüller, H. and Eriksson, A., "Acoustic effects of variation in vocal effort by men, women and children", J. Acoust. Soc. Am., 107(6):3438–3451, 2000.

[8] Wesker, T., Meyer, B., Wagener, K., Anemüller, J., Mertens, A. and Kollmeier, B., "Oldenburger logatome speech corpus (OLLO) for speech recognition experiments with humans and machines", Interspeech,1–4, 2005.

[9] Gottsmann, F., "Realisierung einer Messmethode für den Stimmaufwand eines Sprechers", Magisterarbeit, Universität Bonn, 2010.