

# Comparing the Impact of GSM Transmission on Normal- and High-Effort Speech

Corinna Harwardt

FGAN - Research Establishment for Applied Science  
FKIE - Research Institute for Communication, Information Processing and Ergonomics  
Neuenahrer Str. 20  
Wachtberg, Deutschland  
harwardt@fgan.de

## ABSTRACT

Speech processing applications in mobile environment strongly depend on the quality of the transmission channel. One standard method used for mobile phones is the GSM transmission. Having an extended knowledge about the changes of acoustic measurements induced by the GSM transmission is the baseline for the development of robust features for successful speech processing tasks in mobile environment. Additionally, the knowledge about these changes may help to compensate for a mismatch in transmission channels in training and test data. Therefore we illustrate the influence of GSM transmission on acoustic measurements in this paper. The acoustic measurements we present are the first three formants and their bandwidths.

For the first and third formant we found significant differences between GSM-transmitted and studio-recorded speech. Both mean values increased, whereas the second formant did not show a significant change. The bandwidth mean values get decreased by GSM transmission for all three bandwidths. Additionally we evaluated differences concerning the impact of GSM transmission on different degrees of vocal effort. This difference is important in mobile environment, because the caller is often forced to speak up due to loud background noise.

In this investigation we observed the greatest difference between measurements from signals with different degrees of vocal effort for the first formant. The other two formants and their bandwidths differ significantly, too. The mean values of the first formants bandwidth are not significantly different for diverse degrees of vocal effort in GSM-transmitted speech.

## Categories and Subject Descriptors

I.2.7 [Natural Language Processing]: Speech recognition and synthesis

## General Terms

EXPERIMENTATION

## Keywords

Acoustic measurements, GSM transmission, vocal effort

## 1. INTRODUCTION

Speech in mobile environment is always effected by the transmission channel as well as by background noise and the changes in a speakers behavior induced by background noises (Lombard effect). The effect of the transmission channel and the Lombard effect on formants and their bandwidths has been investigated and will be presented in this paper. Previous studies about the impact of GSM, or more general telephone transmission on acoustic measurements have been mostly provided by forensic scientists, who focussed on manual measurements of vowel formants or at least on manual annotation of vowels before measuring the formants automatically [1], [6]. This study will completely rely on automatic measurements of formants and bandwidths, to generate reasonable results for speech processing applications. Additionally we used a larger set of test utterances and speakers. The influence of GSM transmission on F0 mean values has already been described in [4].

The second problem mentioned earlier does occur frequently, because mobile phone users are often forced to speak up, due to background noise (Lombard effect). The difference between acoustic measurements in diverse degrees of vocal effort is, hence, very important to speech processing in mobile environment. As further expansion to previous studies we therefore investigated the differences between acoustic measurements in GSM-transmitted speech with normal- and high-effort.

## 2. METHOD

For our investigations we used parts of the Pool 2010 corpus [5]. The database contains spontaneous speech from two different tasks concerning the degree of vocal effort. The "normal-effort" task was recorded without any additional equipment. In the "high-effort" task Lombard speech was recorded by inducing 80 dB white noise to the speaker via headphones. These two signals per speaker are available as studio recordings obtained with a high quality microphone and as GSM transmission. For the GSM transmission different mobile phones have been used to guaranty different recording conditions as they are common in forensic case work. The GSM codec is not known. The advantage of this procedure is that it leads to realistic conditions for automatic speech processing tasks in mobile environment, too. In this study we used spontaneous speech with two differ-

**Table 1:** Wilcoxon rank sum test results on studio recordings and GSM-transmitted speech for normal and high vocal effort.

	W	p-value	significance
F1 (normal)	177	$1.429e^{-13}$	***
F1 (high)	403	$5.359e^{-9}$	***
BW1 (normal)	2496	$2.2e^{16}$	***
BW1 (high)	2497	$2.2e^{16}$	***
F2 (normal)	1190	0.6817	ns
F2 (high)	995	0.07935	.
BW2 (normal)	2179	$1.545e^{10}$	***
BW2 (high)	2094	$6.066e^{09}$	***
F3 (normal)	689	0.0001116	***
F3 (high)	808	0.002337	**
BW3 (normal)	2086	$8.422e^{09}$	***
BW3 (high)	2065	$1.965e^{08}$	***

ent degrees of vocal effort from 50 speakers in studio and in GSM quality. The two speech signals per speaker with normal- and high-effort speech do not contain the same utterances. In contradiction to this, the two signals from one speaker including different channel characteristics contain exactly the same utterance.

To measure the formants we used the Snack toolkit [8]. For the calculation of the mean of each formant we used just the voiced frames. The decision which frame to characterize as voiced has been done with the help of the voicing probabilities provided by the Snack toolkit when calculating  $f_0$  and, additionally by the formant measurements which partly just provided a standard value for voiceless frames. These decisions were quite different for studio and GSM-transmitted speech. In GSM transmission we found much more formant measurements that included the standard value than in the studio speech task. This resulted in an average of 4 - 6 % samples that have been characterized as voiced in the signal of one kind of quality but not in the corresponding signal of the other quality. Those samples have not been used for the calculation of the mean. Similar problems have already been reported by [3].

After obtaining all mean values for the first three formants and their bandwidths in both qualities and both degrees of vocal effort, we did statistical significance tests to evaluate the differences between these groups of measurements. For all those tests we used the Wilcoxon rank sum test<sup>1</sup>.

### 3. RESULTS

The next subsections describe the influence of the GSM transmission on the formant mean values of the first three formants and the corresponding mean bandwidths. A special focus is set on the impact of GSM transmission on normal- and high-effort speech to show emerging problems in mobile communication when the caller speaks up. Table 1 summarizes the statistical significance test results when testing the difference between studio-recorded and GSM-transmitted speech. The value W represents the rank sum of the samples from the studio-recorded speech task.

Table 2 summarizes the statistical significance test results when testing the difference between normal- and high-effort

<sup>1</sup><http://sekhon.berkeley.edu/stats/html/wilcox.test.html> (08.04.2009)

speech. The value W represents the rank sum of the samples from the normal-effort task. We will refer to the results of both tables in the following sections. For both tables we used the following significance codes for the p-values: '\*\*\*' < 0.001, '\*\*' < 0.01, '\*' < 0.05, '.' < 0.1, 'ns' >= 0.1; ns = not significant. Notice that the rank sums for the formants are different to those of the bandwidths. Therefore the W values of formants and bandwidths are lying in different ranges.

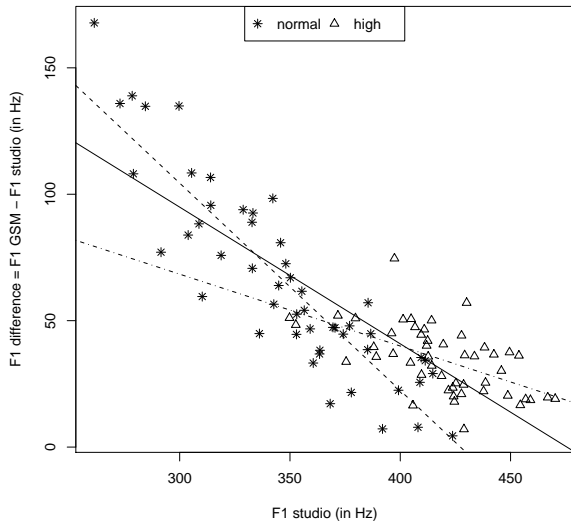
**Table 2:** Wilcoxon rank sum test results on the difference of normal- and high-effort speech for studio recordings and GSM-transmitted speech.

	W	p-value	significance
F1 (studio)	179	$1.585e^{-13}$	***
F1 (GSM)	196	$3.796e^{-13}$	***
BW1 (studio)	2058	$2.595e^{08}$	***
BW1 (GSM)	1281	0.8335	ns
F2 (studio)	892	0.01372	*
F2 (GSM)	755	0.000652	***
BW2 (studio)	2320	$1.670e^{13}$	***
BW2 (GSM)	2252	$5.05e^{12}$	***
F3 (studio)	752	0.0006043	***
F3 (GSM)	825	0.003429	**
BW3 (studio)	2067	$1.815e^{08}$	***
BW3 (GSM)	2083	$9.518e^{09}$	***

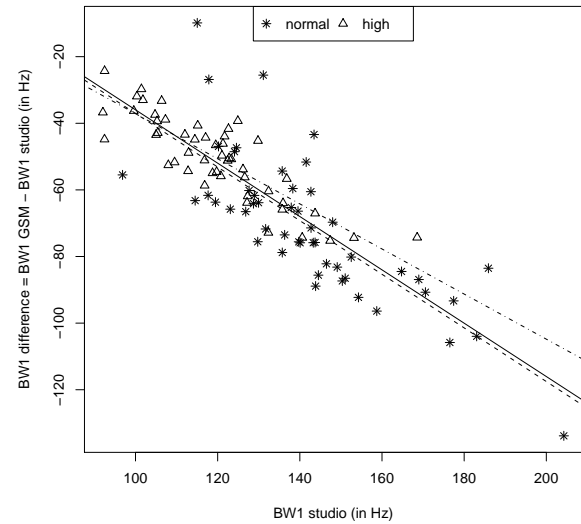
#### 3.1 The first formant and its bandwidth

This subsection presents the effects of the GSM transmission on the F1 measurements and its bandwidth (BW1) on all voiced frames of a speech signal. Figure 1 illustrates the differences between the F1 mean values of studio-recorded speech and the F1 mean value of same signal transmitted via GSM. The figure contains data from the normal- and the high-effort task and presents regression lines for the different data sets. The mean values of the bandwidth of the first formant are plotted in Figure 2. Just like Figure 1, Figure 2 presents mean values of normal- and high-effort speech and the corresponding regression lines.

The results for the normal-effort speech can be approximated by a regression line with negative slope (dashed line) for both, F1 and BW1. Examining just the results of the high-effort task for F1, we see that the means rather appear as some kind of cluster with F1 difference values mostly between 0 and 50 Hz. This clustering is just a matter of the size of the x-axis and hence, the samples can be approximated by a regression line very well (dotdashed line). The x-axis needs to be large to present both data sets. The cluster of high-effort samples is located in the higher region of the F1 mean values plotted on the x-axis, because raising vocal effort leads to a raised first formant [5], [7], [9]. This raising of the F1 mean values is highly significant for studio speech as well as for GSM-transmitted speech (see Table 2). Analogous we find a cluster of BW1 mean values for the high-effort task, with the discrepancy, that the cluster is located in the lower frequency regions between 0 and 140 Hz, with difference values between -20 and -80 Hz. This shows that the higher F1 values in Lombard speech lead to smaller bandwidths. Furthermore the F1 as well as the BW1 measurements in high-effort speech are less affected by the GSM



**Figure 1:** Differences between F1 mean values for studio-recorded and GSM-transmitted speech; dashed line = regression line for normal-effort speech, dot-dashed line = regression line for high-effort speech, solid line = regression line for all samples.



**Figure 2:** Differences between BW1 mean values for studio-recorded and GSM-transmitted speech; dashed line = regression line for normal-effort speech, dot-dashed line = regression line for high-effort speech, solid line = regression line for all samples.

transmission than normal speech. Notice that the direction of change is different for both measurements: F1 is raised up, whereas BW1 decreases. The difference between the two degrees of vocal effort for BW1 is significant for studio speech, but not for the GSM task (see Table 2).

Regarding all samples plotted in Figure 1 as one set of F1 values independent of the degree of vocal effort and all samples plotted in Figure 2 as one set of BW1 values you can draw the solid regression lines, both with a negative slope. We conclude a general tendency that higher F1 mean values lead to lower F1 difference values and lower BW1 values cause smaller deviations from the actual (studio) BW1 value. The dispersions around the regression lines point out the different influence of the GSM transmission on different speakers or, depending on the GSM codec, it might express the different bit rates used to transmit different speech signals.

The difference between studio-recorded and GSM-transmitted speech is for both degrees of vocal effort and both parameters (F1 and BW1) highly significant (see Table 1). For F1 the p-value as well as W is greater for high vocal effort than for normal-effort speech, indicating that the differences between the studio and GSM recordings are smaller for the high-effort speech. For BW1 both values are almost equal for different degrees of vocal effort. Therefore the three regression lines presented in Figure 2 are very similar.

The significant difference for F1 induced by GSM transmission is conform with the findings of [6] and [1]. The decrease of formant frequencies stated by [2] can not be supported by this study.

### 3.2 The second formant and its bandwidth

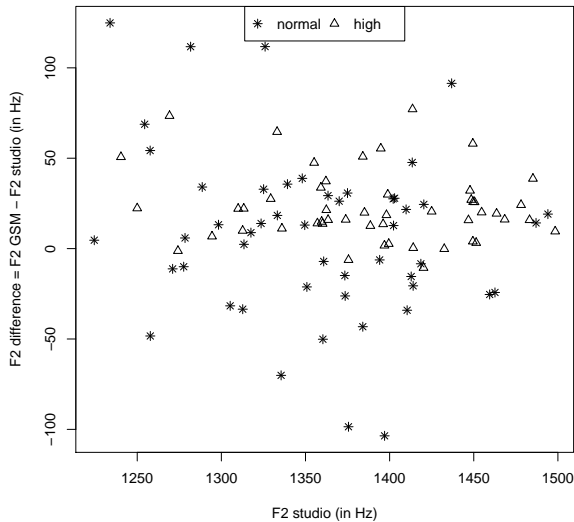
The influence of the GSM transmission on the F2 mean values of normal- and high-effort speech is illustrated in Figure 3. As you can see, there is no clear pattern for both tasks. Most of the F2 difference values are clustered around 20 Hz, independent of the input frequency plotted on the x-axis and independent of the degree of vocal effort. The frequency shift of high-effort speech that occurred for the input frequencies of the F1 mean values does not exist for F2 means.

One difference that can be observed between the two degrees of vocal effort is the occurrence of many negative difference values for normal-effort speech, which is not analogous in high-effort speech. This is supported by the Wilcoxon test investigating the degree of vocal effort, which gave a significant p-value for the studio recordings and a highly significant one for GSM transmission (see Table 2).

The difference between studio and GSM-transmitted speech shows no significance for F2 in the normal vocal effort task and just a slight significance ( $p < 0.1$ ) for high vocal effort (see Table 1). These findings are in the line with the results of [6] and [1] for normal-effort speech and again not in line with the findings of [2].

The mean values of the bandwidth (BW2) are plotted in Figure 4. The BW2 means are, as already observed for BW1, arranged linearly. For normal-, high-effort speech and for both tasks pooled together linear regression lines with negative slopes can be drawn. The regression lines are, again, very similar to each other. The difference between studio and GSM-transmitted speech is highly significant for both degrees of vocal effort (see Table 1).

The BW2 values for high vocal effort are clustered in the lower frequency range, whereas the mean values for normal-vocal effort are lying in the higher frequency range. The



**Figure 3:** Differences between  $F2$  mean values for studio-recorded and GSM-transmitted speech.

difference between normal- and high vocal effort is highly significant for GSM as well as for studio speech (see Table 2).

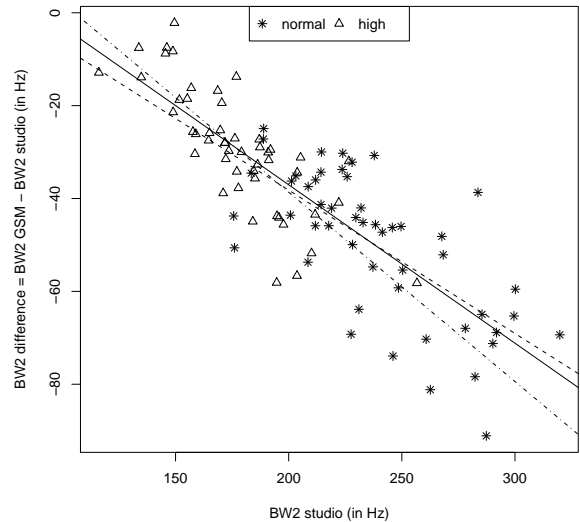
### 3.3 The third formant and its bandwidth

In Figure 5 the effect of GSM transmission on the  $F3$  mean values is expressed. The figure contains samples of normal- and high-effort speech. The samples with high vocal effort have a small shift upwards for the input frequencies on the x-axis. This shift is similar to the shift observed for  $F1$  means, but not that large. This is approved by the Wilcoxon test results which present low p-values for studio and GSM-transmitted speech (see Table 2).

The difference between  $F3$  mean values, induced by GSM transmission, ranges from approximately 0 to 200 Hz. There is no consistent trend in the distribution of all  $F3$  mean values and hence, no clear pattern is visible in Figure 5.

The difference between studio recordings and GSM transmission is statistically significant (see Table 1). These results are contrary to those of [1]. They found that  $F3$  was generally unaffected, except for speakers with high  $F3$  values. Equally the results from [2], who found a high decrease of formant frequencies for  $F3$ , could not be approved.

The mean values of the third formants bandwidth ( $BW3$ ) is shown in Figure 6. Most of the  $BW3$  mean values are clustered in the lower frequency range of the x-axis. Those lying in the higher frequency range are only normal-effort samples. The high-effort samples are lying completely in the lower frequency range. The difference between normal and high vocal effort is significant for studio and GSM-transmitted speech (see Table 2), although the regression lines for the normal-, the high-effort task and both combined are almost equal. The difference between studio and GSM-transmitted speech for  $BW3$  is significant, too (see Table 1).



**Figure 4:** Differences between  $BW2$  mean values for studio-recorded and GSM-transmitted speech; dashed line = regression line for normal-effort speech, dot-dashed line = regression line for high-effort speech, solid line = regression line for all samples.

### 3.4 Comparing the results for all three formants and their bandwidths

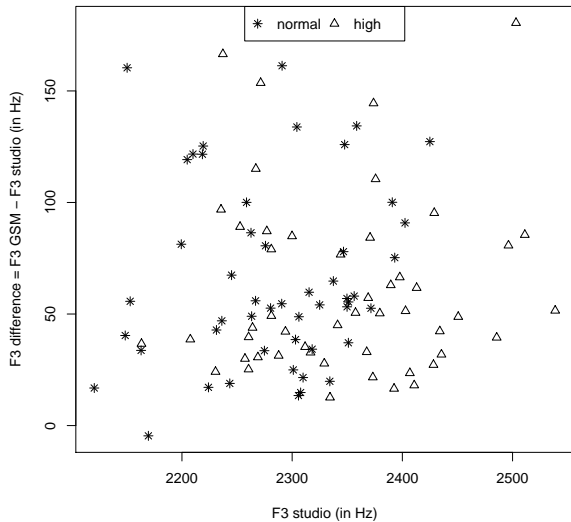
Altogether the formant  $F1$  has been affected by GSM transmission the most. The difference between normal- and high-effort speech is highest for  $F1$ , too.  $F2$  was not affected significantly by the GSM transmission, but the difference between normal- and high-effort speech was highly significant for GSM transmitted speech and slightly significant for studio speech. The third formant was again affected significantly by the GSM transmission. For of  $F1$ ,  $F3$  and  $F2$  in the high-effort task the mean values got increased. The  $F2$  values of the normal-effort task are partly decreased, the rest is increased.  $F1$  can be approximated by a regression line whereas no clear pattern is observable for the other two formants.

The significant changes induced by GSM transmission for  $F1$  and  $F3$  can be explained by the bandpass filter of telephone speech which just transmits frequencies between 300 and 3400 Hz. For both formants some of the single values used for the calculation of the mean might lie out of this range. Especially for these values, we assume more measurement errors and, hence greater differences. This thesis is supported by the fact that we find greater differences for lower  $F1$  mean values, which of course include more values below 300 Hz.

The three bandwidths lead to similar patterns, namely linear regression lines with negative slopes. The difference range of the bandwidth is similar for  $BW1$  and  $BW3$ . The range of  $BW2$  is smaller than those of the other two bandwidths.

## 4. CONCLUSIONS

This paper investigated the influence of GSM transmission on acoustic features in speech with different degrees of vo-



**Figure 5:** Differences between  $F3$  mean values for studio-recorded and GSM-transmitted speech.

cal effort. The acoustic features under consideration are the first three formants and their bandwidths.

We showed that the impact of the GSM transmission on formants is highest for  $F1$ . For  $F2$  no significant changes can be stated. In contradiction to the findings of [1] we observed significant changes for  $F3$  mean values. This contrast might be explained by the different settings of the experiments. The changes of  $F1$  due to GSM transmission can be approximated by a regression line (see Figure 1), but for  $F2$  and  $F3$  we did not find a reasonable pattern to describe the changes. Using non-linear regression techniques might help finding a pattern.

For the three bandwidths, we observed almost the same results. They all get influenced by the GSM transmission significantly. The mean values are decreased for all three bandwidths. The regression lines with negative slope imply that lower bandwidth mean values lead to lower deviations from the actual mean value in studio-recorded speech.

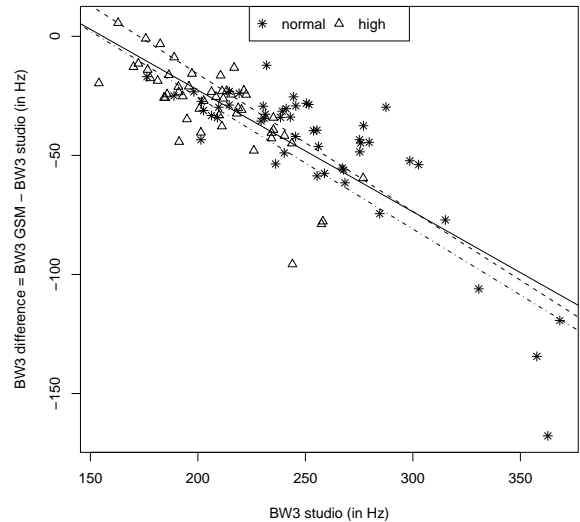
Concerning the different degrees of vocal effort we found for all formants significant changes for studio and GSM-transmitted speech.

The measurements of the bandwidths for normal- and high-effort speech lead to significant changes for all tasks, except for  $BW1$  in GSM-transmitted speech.

The most important observation for the investigating of vocal effort is that the absolute mean values of  $F1$  and all three bandwidths are in average smaller than their counterparts in normal-effort speech. Therefore we conclude that high-effort speech is less affected by GSM transmission than normal-effort speech.

## 5. REFERENCES

- [1] C. Byrne and P. Foulkes. The 'Mobile Phone Effect' on Vowel Formants. *Speech, Language and the Law*, 11(1):83–101, 2004.
- [2] B. J. Guillemin and C. I. Watson. Impact of the GSM



**Figure 6:** Differences between  $BW3$  mean values for studio-recorded and GSM-transmitted speech; dashed line = regression line for normal-effort speech, dot-dashed line = regression line for high-effort speech, solid line = regression line for all samples.

- AMR Speech Codec on Formant Information Important to Forensic Speaker Identification. In *11th Australian International Conference on Speech Science & Technology*, 2006.
- [3] B. J. Guillemin, C. I. Watson, and s. Dowler. Impact of the GSM AMR speech codec on acoustic parameters used in forensic speaker identification. In *8th International Symposium on DSP and Communications Systems*, 2005.
- [4] C. Harwardt. The Impact of GSM Transmission on Automatic  $f0$  Measurements. *to be published, IAFPA (International Association for Forensic Phonetics and Acoustics), 18th Annual Conference*, 2009.
- [5] M. Jessen, O. Köster, and S. Gfroerer. Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, 2005.
- [6] H. J. Künzel. Beware of the 'telephone effect': the influence of the telephone transmission on the measurement of formant frequencies. *International Journal of Speech Language and the Law*, 8:80–99, 2001.
- [7] R. Schulman. Dynamic and perceptual constraints of loud speech. *Journal of the Acoustical Society of America*, Supplement 1, 78, 1985.
- [8] K. Sjölander and J. Beskow. WAVESURFER- AN OPEN SOURCE SPEECH TOOL. In B. Yuan, T. Huang, and X. Tang, editors, *Proceedings of ICSLP 2000, 6th Intl Conf on Spoken Language Processing*, pages pp. 464–467, Beijing, 2000.
- [9] H. Traunmüller and A. Eriksson. Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America*, 107(6):3438–3451, 2000.