The Effect of Use of Drugs on Speaker's Fundamental Frequency and Formants

Andrey Raev[1], *Yuri Matveev*[1], *Tatiana Goloshchapova*[2]

[1] Speech Technology Center, St. Petersburg, RUSSIA
[2] Federal Drug Control Service of the Russian Federation, Moscow, RUSSIA

raev@speechpro.com, matveev@speechpro.com

Abstract

In this paper we investigate speech recordings before and after speaker's drug-abuse treatment, and show that there is no statistically significant dependency between distortions of speaker's fundamental frequency and formants on the one side, and different groups of drugs and on the degree of drug intoxication on the other. Changes of the fundamental frequency are not regular and do not have a general nature. The main reasons for these changes are changes in the emotional state of speakers, rather than a drug addiction treatment. Exploring the effect of the duration of narcotic drugs usage on the speaker's fundamental frequency showed that voices of speakers with prolonged use of drugs of the heroin group tend to decrease the fundamental frequency by about 3% per year.

Index Terms: speaker, fundamental frequency, formants, drug intoxication

1. Introduction

It is a widely accepted opinion that drug intoxication as well as other factors such as emotional (for example, stress) and functional (for example, illness) states influence speaker's vocal parameters and the way a person speaks. For example, in [1] general theoretical and applied research of the impact of emotional and functional states of the person on the acoustic characteristics of his speech were analyzed. It was shown that in most cases as the most informative acoustic correlates of the emotional and functional states considered some of the frequency, timing and power characteristics of the voice signal. It was demonstrated that the sthenic state typically tends to increase (and the asthenic to decrease) the fundamental and formant frequencies, as well as the intensity.

As noted in [2], a number of vocal parameters can be used as indicators of the effects of psychoactive and antidepressant drugs, in particular the fundamental frequency, the spectral energy distribution, and the pause length, changed with the patient's mood over different therapy periods. The following acoustic parameters were analyzed in [3]: mean values and standard deviations of signal power, the fundamental frequency, frequencies of the first and second formants, and utterance duration. After comparing emotional state against neutral state in terms of their underlying intonation parameters, it was found that fundamental frequency and the first formant deviations had the strongest discriminative power. It was shown that the direction of these deviations did not depend on the semantic content of the utterance and its duration, speaker age or gender, although the individual features of speakers influenced the absolute values of the frequencies. In [4] were investigated the fundamental frequency changes in presence of person's stress and emotional arousal changes.

There are only few studies where the problem of how drug intoxication influences the speaker voice parameters is investigated. There are no studies on how these parameters are depend on different groups of drugs and the degree of drug intoxication, as well as the duration of narcotic drugs usage. This is mainly due to the lack of speech databases for such studies.

To conduct these investigations, the authors collected a specialized database, a Speech Database Narco (SDBN), containing the voice samples of speakers who were under the influence of narcotic drugs. The SDBN was collected by STC Ltd (St. Petersburg, Russia) in a several specialized health care institutions by the request of the Federal Drug Control Service of the Russian Federation.

This paper is a continuation of [5] and presents the results of exploring the effect of narcotic drug intoxication on the speaker's fundamental and formant frequencies using the collected database. The results of these studies are important for expert practice, and should be considered when developing automatic speaker identification systems [6], [7] and [8].

2. Description of the speech database

2.1. Collecting data

All the enrolled speakers in the SDBN have a personal number, accompanied by the following information:

- Speaker's gender.
- Estimated degree of drug intoxication.
- Hypothetical type of the drug substance used.

We recorded only those speakers who were under the influence of drugs at that moment. The degree of drug intoxication was evaluated by narcologists (therapists for drug abusers at hospitals, experts at out-patient departments for drug addicts). The duration of the utterances is not less than 32 sec. Each speaker was recorded either once (while intoxicated), or twice (the first time while intoxicated, the second time - in sober state). Speaker voice recordings were made with a digital recorder at a distance of 1-2 meters from the speaker's lips, under the supervision of specially trained operators, in accordance with the instructions. The recorded speech material was carefully segmented with an audio editor program. The operator's speech, pauses and acoustic noises were deleted from the records.

2.2. Characteristics of the speech database

The present study is based on the SDBN that contains the speech of 156 Russian speakers of both genders (67 female and 89 male speakers), recorded while drug intoxicated and in sober state. The aim was to analyze the effect of narcotic drug intoxication on the speaker's F0 with regard to the gender and the duration of usage.

Aside from the speech signal we collected a number of metadata about speakers to allow statistical cross-testing for other factors than the main factor sober/intoxicated. In Tables 1-2 these metadata are summarized for the SDBN speakers.

Primarily, the SDBN speakers had different degrees of intoxication (see Table 1): mild, moderate, severe. The degrees of drug intoxication were determined by narcologists. Secondary, the speakers used different groups of drugs (see Table 2): plant-based (marijuana, hashish), semi-synthetic (heroin) and synthetic (amphetamine). In addition, in the SDBN speakers are of different genders and had different duration of narcotic drugs usage (see Table 3).

Table 1. *Number of speakers with different degrees of intoxication.*

	Degree of drug intoxication		
	mild	moderate	severe
Number of speakers	39	64	52

Table 2. *Number of speakers with different groups of drugs.*

	Type of drugs used		
	plant-based (marijuana, hashish)	semi-synthetic (heroin)	synthetic (amphetamine)
Number of speakers	39	64	52

Table 3. *Number of speakers with different length of time of using drugs.*

	Duration of narcotic drugs usage, years				
	1	2	5	10	25
Number of speakers	66	21	24	34	11

3. Analyzing voice of intoxicated speakers

Basing on the SDBN we analyzed the effect of the level of drug intoxication, different groups of drugs and the duration of narcotic drugs usage on speaker's fundamental frequencies and formants.

2.1. The speaker voice features while drug intoxicated

Analysis of the speaker's speech recordings before and after drug treatment indicates that the speaker's voice changes slightly and irregularly.

Figures 1a and 1b show the histograms of fundamental frequencies of four men before and after drug treatment. After drug treatment, the fundamental frequency of one of the speakers has increased, another has reduced, and for other speakers it remained unchanged.

Figure 1c shows the averaged histogram of fundamental frequencies for the same four men before and after drug treatment. Parameters of the histogram have changed slightly: the mean and median of the distribution of the frequencies remained the same, and the variance decreased. But this was most likely due to the changes in the emotional state.

The situation is similar for women: the mode of the fundamental frequencies distribution increased from 180 Hz to nearly 200 Hz, but the mean and median of the distribution remained almost unchanged. Statistically significant regular changes were not observed. This suggests that the main reason for the fundamental frequency changes is the emotional state changes of speakers, rather than drug addiction treatment.

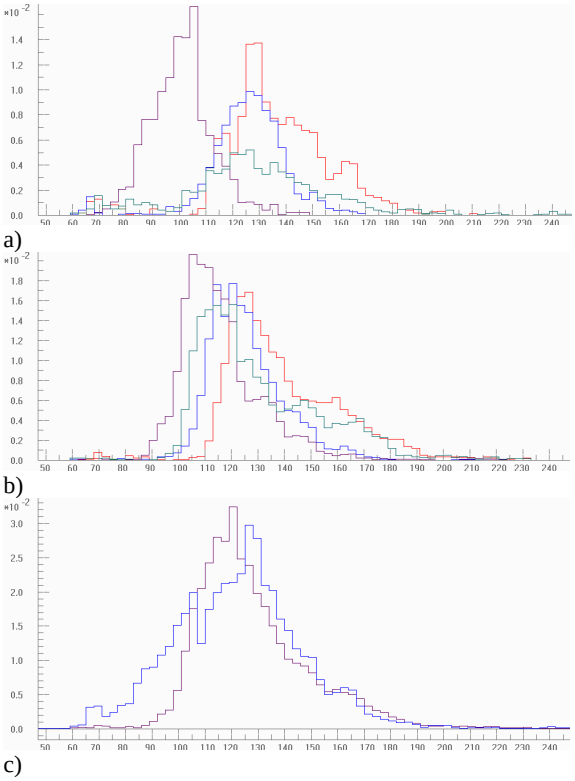

Figure 1: *Histograms of the fundamental frequencies of four men a) before drug treatment and b) after drug treatment, c) averaged histogram of the fundamental frequencies of men before (the right mode) and after (the left mode) drug treatment.*

To illustrate the effect of drugs on the formants was chosen speaker, which was in a state of severe drug intoxication before the treatment. Figure 2 shows the formant trajectories of two utterances of the same sentence before and after drug treatment.

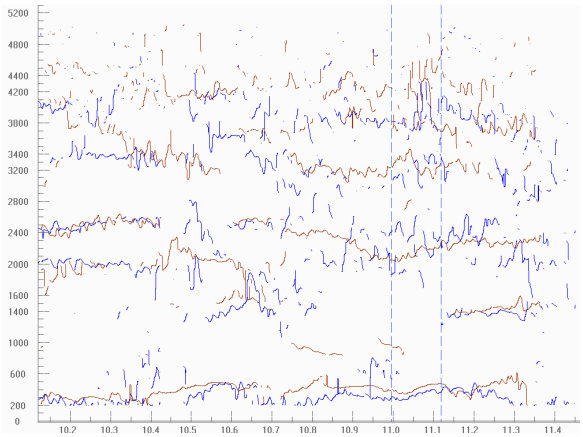

Figure 2: *Formant trajectories of two utterances of the same sentence before (blue color) and after (brown color) drug treatment.*

The phrase was pronounced in two completely different styles. Before treatment, the speaker speaks depressed and quiet, hardly articulates phrases. Signal-to-noise ratio hardly reaches 15 dB, so the formant trajectories are noisy and often interrupted or break up into disconnected pieces. After treatment, the speaker speaks freely, loudly and repeatedly laughing. Signal-to-noise ratio in the same recording

conditions increases up to 25 dB. Formant trajectories become stable and continuous. Nevertheless, a systematic shift of formants is not observed. The first formant is shifted slightly, from 250-400 Hz to 300-500 Hz. Formants in the areas of 1300 Hz, 2000 Hz, 2400 Hz, 3400 Hz retain their positions.

The histograms of formant frequencies are shown in Figure 3. The red color shows the formants of speech, which was recorded after the speaker has been treated for drug addiction. Noticeable shift of the maxima is not observed. There are quite a chaotic redistribution of power between the second (1400-1500 Hz) and third (2400-2500 Hz) formants. It is due to low signal-to-noise ratio in the recording that was made during a severe drug intoxication of the speaker.

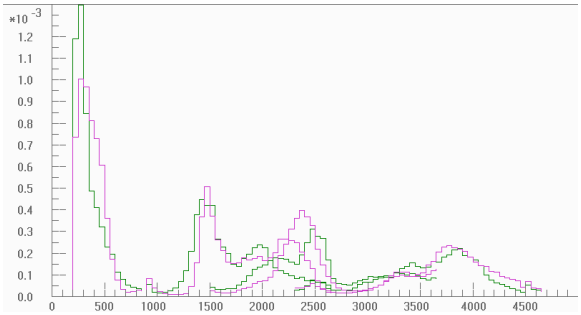

Figure 3: *Histograms of formants for speakers in a state of severe drug intoxication (green color) and in a normal state (red color).*

2.2. The dependence of the speaker's fundamental frequency on drug intoxication.

We have carried a statistical study of the effect on the fundamental frequency of the following characteristics:

- The drug intoxication degree.
- Type of the drug substance used.
- The duration of narcotic drugs usage.

According to narcologists, there are no methods for determining the degree of drug intoxication, therefore we used a Mean Opinion Score (MOS) for the degree of intoxication and for the length of time of using drugs.

To determine the speaker intoxication degree, primarily we focused on the following symptoms: speech disorders (slurred speech), inability to understand what the doctor asks him to do, omission or misunderstanding of some items of the test, own statement that he had recently used drugs.

All of these symptoms, including a statement of the recent drug use, are only indirect evidence that the speaker is drug intoxicated. Therefore, some recordings marked as belonging to a drug intoxicated speaker, may belong to a sober speaker, and vice versa. Nevertheless, the conclusions made on these samples are statistically valid.

The time of using drugs was estimated taking into account the following factors:

- Statement of the speaker that he is registered in the drug treatment clinic with a certain year.
- Statement of the speaker that he has been using drugs from a certain year.
- Statement of the speaker that he was infected with hepatitis B or C in a certain year.
- Statement of the speaker that he was infected with HIV in a certain year.

Using all these factors we can only roughly estimate the duration of drug use. Many speakers repeatedly treated for drug addiction, and they said they had had remission (a complete rejection of drug use), including long-term.

Almost all recordings of the speaker's in severe drug intoxication were very quiet, with an average amplitude of about 1-2 thousands of bins. However, a characteristic, such as energy, depends much more on the distance from the microphone to speaker's lips than on the state of the speaker. Therefore, below this characteristic is not considered.

Histograms of the fundamental frequency of the females are shown in Figure 4. These histograms do not show any statistically significant anomalies. Median of the fundamental frequency distributions is in the range 190-220 Hz, which fits in with allowable limits for normal female voices. With some stretch only one histogram can be regarded as an anomalous, which has a two-mode distribution with the modes at the points of 120 Hz and 225 Hz. However, according to the experts, a few percent of the female voices have a distribution with two modes.

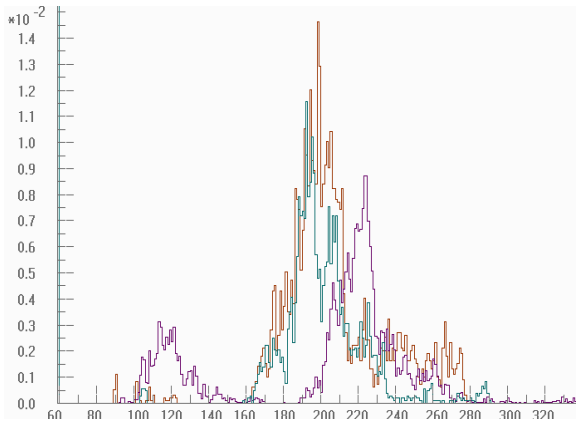

Figure 4: *Histograms of the fundamental frequency of the females in a state of severe drug intoxication.*

Figure 5 shows the examples of histograms on the fundamental frequency of 6 males in a state of severe drug intoxication. These histograms do not show any statistically significant anomalies. Median of the fundamental frequency distributions is in the range of 110-130 Hz, which fits in with allowable limits for normal male voices. Only one histogram of six can be regarded as anomalous that does not allow statistically significant conclusions.

A total out of three histograms of the female voices and 22 histograms of male voices in a state of severe drug intoxication were found only one abnormal histogram for female and 6 abnormal histograms for males. Anomaly for men is in a low average fundamental frequency (below 100 Hz), which is rare for men not using drugs (according to experts - not more than 5% of men).

As a result, it may be concluded that the state of drug intoxication does not directly affect the fundamental frequency.

At the same time, the analysis of anomalies shows that of the 6 anomalies in men, in four cases, it takes place for the speakers who took drugs five or more years. The only anomaly in female voices belong to the women who had used drugs within four years. This suggests another reason for the anomalies, namely, not the state of drug intoxication, but the duration of narcotic drugs usage.

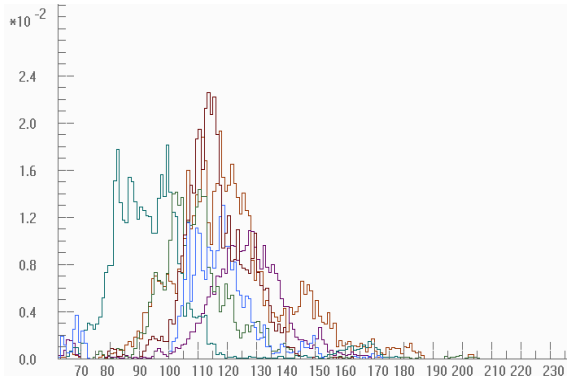

Figure 5: *Histograms of the fundamental frequency of six males in a state of severe drug intoxication.*

Figure 6 shows histograms of the fundamental frequency of 15 males who use drugs five and more years. This is a completely different situation. The main mode for more than 2/3 of speakers moves from the 120-140 Hz range to 90-100 Hz range.

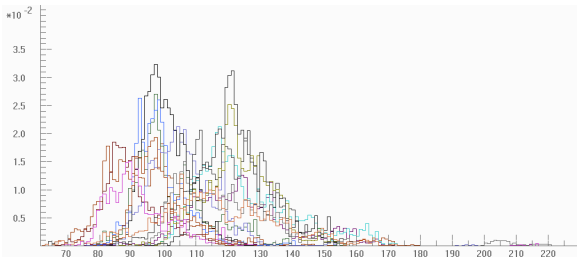

Figure 6: *Histograms of the* fundamental frequency *of 15 males who use drugs five and more years.*

The presence of the second mode at 120 Hz can be explained by the fact that the duration of narcotic drugs usage is not known exactly, and obtained using indirect data. In addition, some speakers were treated for drug addiction, and they had a long-term remission.

If we sum up all the histograms in Figure 6, then after normalization we have the histogram in Figure 7, where the mode of fundamental frequency distribution is at a frequency of about 97 Hz.

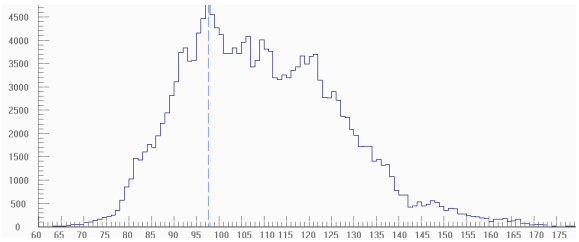

Figure 7: *Averaged histogram of the* fundamental frequency *of 15 males who use drugs five and more years*

All this suggests that prolonged use of drugs of the heroin group leads to a decrease in the fundamental frequency.

The average rate of the decrease can be very roughly estimated as 4.5 Hz (3%) per year. The linear dependency of the estimate of the decrease is chosen as the most simple. Its applicability is limited to 2-8 years of intensive using of drugs of the heroin group. To explore more complex dependencies and to evaluate the impact of a longer duration of narcotic drugs usage, the obtained data are insufficient.

A similar conclusion for other types of drugs cannot be done due to a lack of data. The average rate of the decrease of the fundamental frequency may be somewhat underestimated and may be nonlinear. Analysis of types of narcotic drugs and the duration of their use, suggests that narcomaniacs are gradually switched to heroin from "light" drugs.

4. Conclusions

Using the collected speech database with speaker speech recordings while narcotic drug intoxicated and being in sober state, with different degrees of drug intoxication, we analyzed the speaker's voices before and after drug treatment and investigated the effect of the duration of narcotic drug usage on the speaker's fundamental frequency.

The analysis showed that there is no statistically significant dependency between distortions of speaker's vocal parameters and different groups of narcotic drugs, and the degree of drug intoxication.

The fundamental frequency changes are not regular and do not have general nature. The main reason for changing the fundamental frequency is the speaker emotional state, rather than drug addiction treatment.

Changes in formant trajectories are not regular and do not have a common nature, that is, state of drug intoxication does not directly affect the formant frequencies. Systematic bias of formants was not found. The first formant is shifted slightly, from 250-400 Hz to 300-500 Hz. Formants in the areas of 1300 Hz, 2000 Hz, 2400 Hz, 3400 Hz retain their positions.

The study of the effect of the duration of narcotic drugs usage on speaker's voice shows that speakers with prolonged use of drugs of the heroin group have a tendency to decrease the fundamental frequency by about 3% per year.

5. References

[1] Adashinskaya, G.A. and Chernov, D.N., "Acoustic correlates of person-distinctive functional and emotional states", Aviakosmicheskaja i ekologicheskaja meditsina, 41(2): 3–13, 2007.

[2] Helfrich, H., Standke, R. and Scherer, K.R., "Vocal indicators of psychoactive drug effects", Speech Communication, 3: 245-252, 1984.

[3] Gelman, V.Ya., Dmitrieva, E.S., Zaitseva, K.A. and Orlov, A.M., "Influence of Human Personal Features on Acoustic Correlates of Speech Emotional Intonation Characteristics", I.P. Pavlov Journal of Higher Nervous Activity, 59(5): 538-546, 2009.

[4] Raev, A., Matveev, Yu. and Goloshchapova, T., "The Effect of Long-Term Use of Drugs on Speaker's Fundamental Frequency", Proc. SPECOM'2011, Int. Conf. "Speech and Computer", Kazan, Russia, Sept. 2011.

[5] Tolkmitt, F.J. and Scherer, K.R., "Effect of Experimentally Induced Stress on Vocal Parameters", Journal of Experimental Psychology and Human Perceptual Performance, 12(3): 302–313, 1986.

[6] Labutin, P., Koval, S. and Raev, A.," Speaker identification based on the statistical analysis of f0", Proc. IAFPA 2007, The College of St Mark & St John, Plymouth, UK, 2007.

[7] Smirnova, N., Starshinov, A., Oparin, I. and Goloshchapova, T., "Speaker Identification Using Selective Comparison of Pitch Contour Parameters", Proc. ICPhS XVI, ID 1138, Saarbrücken, 2007.

[8] Matveev, Yu.N. and Simonchik, K.K., "The speaker identification system for the NIST SRE 2010", Proc. GraphiCon-2010, St. Petersburg, Russia, 2010.