Effect of increased vocal effort on average and range of fundamental frequency in a sample of 100 German-speaking male subjects

Michael Jessen, Olaf Köster and Stefan Gfroerer

Bundeskriminalamt, Department of Speaker Identification and Tape Analysis, Wiesbaden, Germany

E-mail: Michael.Jessen@bka.bund.de, Olaf.Koester@bka.bund.de, Stefan.Gfroerer@bka.bund.de

ABSTRACT

It is known that increased levels of speaking loudness (vocal effort) lead to an increase in mean fundamental frequency. Less is known about the effect of vocal effort on f0-range and on the amount of speaker variation with respect to the relation between vocal effort and f0. Based on a corpus of 100 German-speaking male subjects it was found that, firstly, increased vocal effort does in fact lead to an increase in mean f0 across all subjects, secondly, that f0-standard deviation increases for about 90 percent of all subjects, and thirdly, that there are substantial speaker differences with respect to the size of the f0 differences in normal vs. loud speech. Some speaker differences occur also when amplitude measurements are included in the evaluation, i.e. for the same increase in signal amplitude some speakers show more substantial differences in f0-mean and f0-standard deviation than others.

1. INTRODUCTION

Speaking loudly (i.e. with increased vocal effort) is usually accompanied by an increase of mean fundamental frequency (f0) compared to a neutral vocal effort level (see [1] for the early literature and some possible physical explanations). This effect has important implications for forensic phonetics. f0 is an important carrier of speaker-specific information (with a strong anatomical/ physiological foundation) but in forensic voice comparisons the speech of the unknown speaker is often produced at a different level of perceived vocal effort than the speech sample recorded of the suspect speaker (mostly more vocal effort in the former than the latter, due to factors such as emotional stress). Ignoring the influence of vocal effort on f0 could contribute to false rejection statements (different mean f0 at different vocal effort levels incorrectly attributed to different speakers) or, worse, false identification statements (similar mean f0 at different vocal effort levels incorrectly attributed to the same speaker). The f0-raising effect of increased vocal effort has been shown, among others, by [2,3,4,5,6]. Despite the consistency of this effect across studies, the size of the effect can differ. Some of the differences are probably due to factors such as experimental design or measurement methodology, while some also seem to be purely speaker-dependent. For example, it was shown by [2] that within a Lombard design, exposure to 80 dB white noise lead to an amplitude increase of about 5 dB in the speech production of both subjects used in that study, while the increase in mean f0 from the baseline condition to the Lombard condition was only 1 Hz for the first subject but 16 Hz for the second subject. Speaker differences have also been shown by [3]. Whereas the influence of vocal effort on *mean* f0 is quite well investigated, only a few studies have included measures of f0-*range*. According to [4] there is an increase in f0-standard deviation with increasing levels of vocal effort (similarly [5]). According to [6], on the other hand, f0-range in loud vs. normal speech is essentially the same. In the present study mean f0 and f0-standard deviation were measured and evaluated with special emphasis on speaker differences among the 100 subjects investigated.

2. METHOD

The results presented in this paper are based on about six minutes of (semi)spontaneous speech produced by 100 male German speakers each in a neutral-speech and a loud-speech condition. This sample is part of a larger data corpus recorded at the BKA in 2001, which also includes read speech and speech produced in telephone communication. (Semi)spontaneous speech was elicited by asking subjects to describe a picture – while avoiding certain pre-given terms – to a conversation partner, who had to name the picture on the basis of the description of the subject. This task was carried out both in a baseline condition and a Lombard (loud-speech) condition. In the Lombard condition subjects were exposed to 80 dB_{SPL} white noise over headphones in order to induce an increase in vocal effort. Closed headphones were used to prevent a situation where Lombard-condition noise leaking from the headphones would end up on the recording. A high-quality condenser microphone was attached to a helmet to ensure constant distance to the speaker, which was important to enable valid amplitude measurements. The headphones where attached to the helmet separately from the microphone and could be removed from the ears for the baseline condition without affecting microphone placement. The order of the two

ISBN 1-876346-48-5

conditions of the experiment was varied randomly from subject to subject. DAT (digital audio tape) recordings were made and analysed with ESPS/waves+. Over each of the two conditions (Lombard and baseline) mean f0 and f0-standard deviation were determined based on the labelled speech signal. Labeling was performed manually on the basis of the audio output and the graphical f0 representation, with the goal to exclude certain portions of the f0 signal from statistical analysis. Excluded from analysis were f0 values stemming from non-verbal vocalisations such as laughter and cough, from invalid measurements (e.g. f0 during voiceless portions of the signal), from abrupt f0 changes due to creaky voice, voiceless obstruents, etc., and, of course, from utterances produced by the conversation partner. The f0 statistics (means and standard deviation) obtained in this manner are the basis of the results to be shown in Figures 1-8. To obtain the results presented in Figs. 9-10, signal amplitude was measured by RMS in decibels and mean amplitude across the entire condition was determined. For this purpose, close-to-zero amplitude portions (such as in pauses etc.) were excluded from analysis. Amplitude measurements were only made in voiced speech and only within those sections of the signal that were already included in the f0 analysis mentioned above.

3. RESULTS

As will be shown in this section, vocal effort has an important influence on the distribution of average and range of f0. Figure 1 shows a histogram of mean f0 among the 100 subjects investigated. The results shown in Fig. 1 originate from spontaneous speech spoken at a neutral loudness level (baseline condition). For example, for 32 out of the 100 speakers (i.e. 32 percent of all speakers) f0-mean lies between 111 Hz and 120 Hz (cf. [7] for similar results based on neutral loudness levels). Due to the Gaussian-like shape it can be assumed that the distribution in Fig. 1 is a good representation of the f0-mean values of neutral-style speech produced in the German-speaking male population. As such, the results in Fig. 1 are useful for forensic-phonetic interpretations of f0 evidence in casework, by providing an indication of the strength of the speaker identification evidence. For example, if both suspect and unknown produce mean-f0 between 151 and 160 Hz in neural-style speech, the speaker identification evidence is stronger than if the f0 values of both speakers occur between 111 and 120 Hz.

Fig. 2 shows that there is a substantial shift towards higher f0-mean values when turning from neutral to loud speech (Lombard condition). For example, the range of values up to 120 Hz, which was found in the majority of speakers in neutral speech, is only covered by 2 percent of all speakers in loud speech.

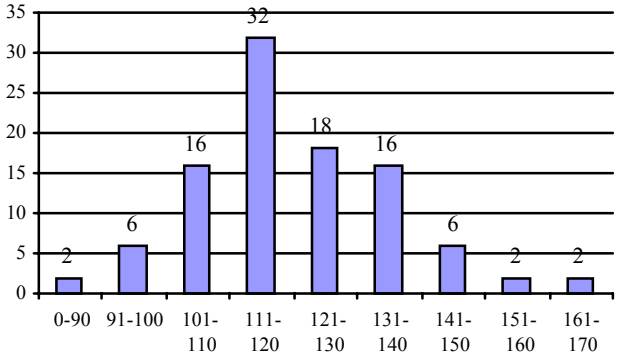

Figure 1: Histogram of mean f0 in spontaneous speech with neutral vocal effort among 100 male subjects. Intervals of f0 values are plotted on the x-axis, number of speakers per interval on the y-axis and on top of the columns.

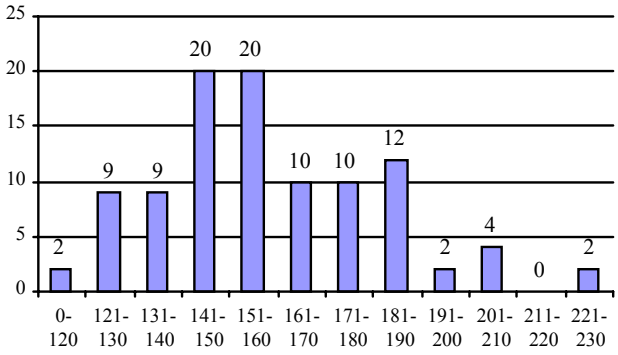

Figure 2: Mean f0 in spontaneous speech with high vocal effort (Lombard condition) among 100 male subjects (cf. Fig. 1 for further conventions).

Fig. 3 shows that the size of the upshift in mean f0 from neutral to loud speech is not uniform across speakers. For some speakers (5 percent) the increase in mean f0 from normal to loud speech is as small as up to 10 Hz, for some (1 percent) it is as large as between 81 and 90 Hz. The interval on the histogram in Fig. 3 occupied by most speakers lies between a 21 and a 30 Hz difference between normal and loud speech.

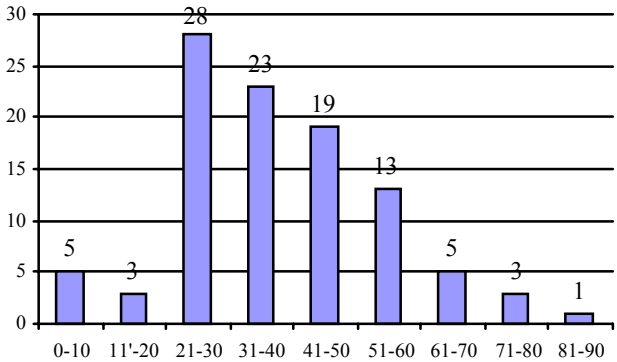

Figure 3: Speaker-dependent size of difference in mean f0 (in Hz) between neutral and loud spontaneous speech.

Fig. 4 shows the difference expressed in terms of semitones instead of Hz – offering a picture of the difference between neutral and loud speech that is more neutral with respect to the particular mean f0 values different speakers start out with in the neutral speech condition.

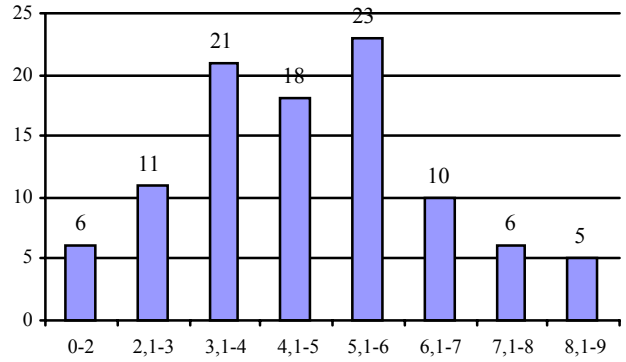

Figure 4: Speaker-dependent size of difference in mean f0 (in semitones) between neutral and loud spontaneous speech.

Turning from mean f0 to the results for f0-range, Fig. 5 shows that the f0-standard deviation measured in neutral speech lies between 16 and 25 Hz for most speakers. This results is consistent with [7], according to whom an f0-standard deviation of 20 Hz constitutes a neutral value.

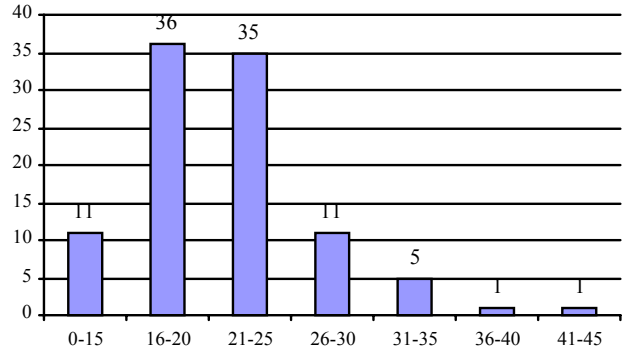

Figure 5: f0-standard deviation in spontaneous speech with neutral vocal effort among 100 male subjects.

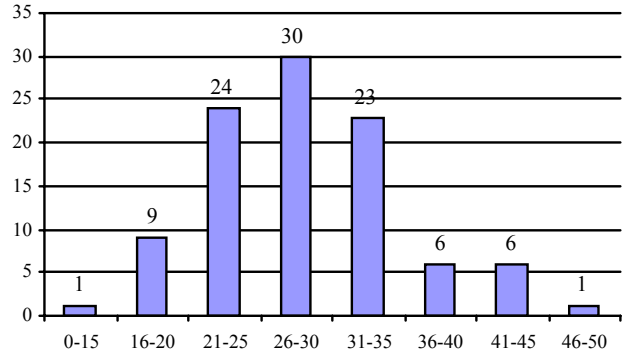

Figure 6: f0-standard deviation in spontaneous speech with high vocal effort (Lombard condition) among 100 male subjects.

Fig. 6 shows that loud speech brings about an upward shift in the distribution of f0-standard deviation values relative to neutral speech. In Fig. 7 it can be seen that not every speaker produced such an upward shift; for 9 percent of the speakers f0-standard deviation is lower or equally high in loud compared to neutral speech. In Fig. 8 this difference between neutral and loud speech is expressed in terms of semitones.

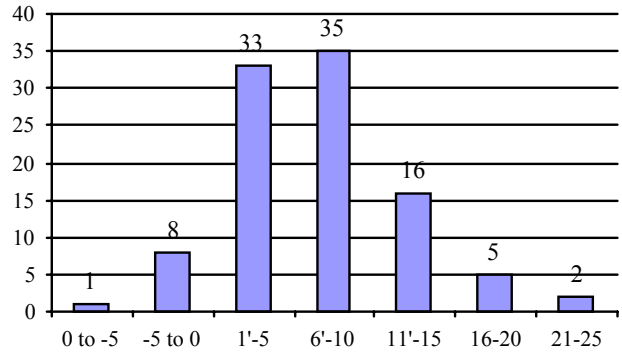

Figure 7: Speaker-dependent size of difference in f0-standard deviation (in Hz) between neutral and loud spontaneous speech.

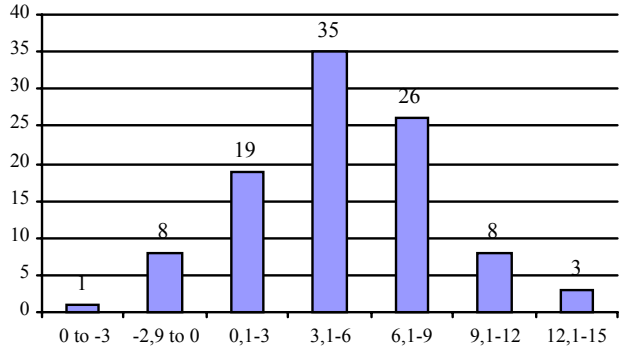

Figure 8: Speaker-dependent size of difference in f0-standard deviation (in semitones) between neutral and loud spontaneous speech.

Some proportion of the speaker-specific patterns with respect to the size of the difference in f0-mean and f0-standard deviation between neutral and loud speech seen in Figs. 3,4 and 7,8 might be the result of the fact that different speakers react to the Lombard condition with different levels of loudness increase. To control for this possibility, the measured size of the amplitude increase from neutral to loud spontaneous speech was related to the size of the change (mostly an increase) in f0-mean (Fig. 9) and f0-standard deviation (Fig. 10). Figs. 9 and 10 show (for values greater than 0) that there is a positive dependence between the measured loudness increase from neutral to loud speech and the associated increase in mean and standard deviation of f0. They also show that there are substantial speaker differences with respect to the relation between the upshift in loudness and the upshift in f0 values. For example, a constant 12 dB increase in loudness can lead to a 70 Hz increase in mean f0 for one speaker but only a 30 Hz increase for another.

ISBN 1-876346-48-5

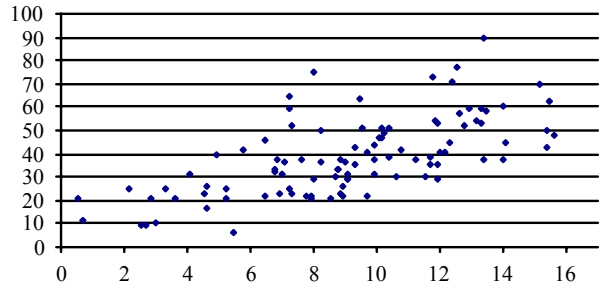

Figure 9: Scatterplot of difference in mean amplitude (in dB on x-axis) against difference in mean f0 (in Hz on y-axis) between loud and neutral speech. Each point represents the results for one speaker.

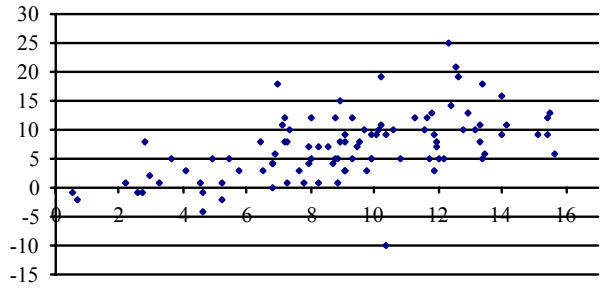

Figure 10: Scatterplot of difference in mean amplitude (in dB on x-axis) against difference in f0-standard deviation (in Hz on y-axis) between loud and neutral speech.

The one value in Fig. 10 that deviates from the general trend (-10 Hz) results from a speaker who had an above-average mean f0 in the baseline condition (137 Hz; cf. Fig. 1). He had a complex intonation pattern in the baseline condition, while in the Lombard condition his intonation was more monotonous – probably due to a ceiling effect (reaching the upper limit of the modal-voice f0-range).

4. CONCLUSIONS

The results of the present study have shown that an increase in vocal effort stimulated by a Lombard design leads to an increase in mean f0 uniformly across speakers and an increase in f0-standard deviation for about 90 percent of all speakers. The size of this increase in f0 values differs quite substantially between speakers. This was shown both on the basis of histograms in which the increase in vocal effort is treated as a categorical variable (Lombard vs. baseline) and on the basis of plots where the increase in vocal effort is expressed as a continuous variable – based on measurements of mean signal amplitude.

For the purpose of forensic speaker identification it would be desirable to have a normalization procedure whereby speech samples that differ in vocal effort can be normalized with respect to the f0 changes caused by these differences. Apart from the difficulty of quantifying vocal effort levels acoustically or auditorily in forensic material, the success of developing such a normalization procedure will be limited by the fact that the size and shape of the association between increased loudness and changes in f0 can differ between speakers. For example, if the unknown speaker only speaks with high and the suspect only with moderate vocal effort, f0-normalization is constrained by the fact that one cannot know in advance how the loudness-f0 association of the particular speakers(s) involved looks like. If, on the other hand, both unknown and suspect speak with different degrees of vocal effort, the speaker-specific loudness-f0 association can be estimated. This also has the advantage that the particular shape of this association can be used as a speaker-specific feature in its own right.

ACKNLOWLEDGMENTS

We thank our colleagues at the BKA who volunteered as subjects for their cooperation. We also gratefully acknowledge the contribution of Martina Huber, Kirsten Kunze, and Jutta Quint, who were financially supported by the BKA and by a grant from the International Association for Forensic Phonetics, for their assistance in recording the corpus and performing labeling.

REFERENCES

[1] I. Titze, “On the relation between subglottal pressure and fundamental frequency in phonation”, *JASA,* vol. 85, pp. 901–906, 1989.

[2] W. Van Summers, D.B. Pisoni, R.H. Bernacki, R.I. Pedlow and M.A. Stokes, “Effects of noise on speech production: Acoustic and perceptual analyses”, *JASA,* vol. 84, pp. 917–928, 1988.

[3] P. Gramming, J. Sundberg, S. Ternström, R. Leanderson and W.H. Perkins, “Relationship between changes in voice pitch and loudness”, *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH Stockholm*, vol. 15, pp. 39–55, 1987.

[4] C. Dromey and L.O. Ramig, “Intentional changes in sound pressure level and rate: Their impact on measures of respiration, phonation, and articulation”, *Journal of Speech, Language, and Hearing Research,* vol. 41, pp. 1003–1018, 1998.

[5] S. Köster, “Acoustic-phonetic aspects of Lombard Speech for different text styles”, *The Phonetician*, vol. 85, pp. 9–16, 2002.

[6] D.R. Ladd and J. Terken, “Modelling intra- and inter-speaker pitch range variation”, *Proceedings of the International Congress of Phonetic Sciences* 13, vol. 2, pp. 386–338, 1995.

[7] H.J. Künzel, *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*, Heidelberg: Kriminalistik Verlag, 1987.