Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages

Anita Wagner[†] **and Angelika Braun**[‡]
† MPI for Psycholinguistics, Nijmegen
‡ University of Marburg, Marburg
E-mail: anita.wagner@mpi.nl, Braun3@mailer.uni-marburg.de

ABSTRACT

This paper addresses the relation between voice quality features and language, or, more precisely, stereotypes concerning voice existing in different speech communities. Results are presented of a study analyzing the acoustic parameters F0, F0 modulation, HNR, jitter and shimmer for male voices of native speakers of German, Italian and Polish. Speakers of these languages have the reputation of sounding either more "rough" (Italian) or more "clear" (Polish) than the other groups. The parameters examined in the presented study correlate with the psycho-acoustic impression of "roughness" in voice. Significant differences between the language groups have been found showing that different parameters predominate in voices of speakers of different languages. In the absence of more "superficial" explanations for these findings like voice pathologies, smoking/drinking habits, etc. we conclude that the differences found are caused by cultural stereotypes in the languages involved.

1 INTRODUCTION

A speaker' s voice depends primarily on his/her anatomical and physiological characteristics. Changes within the vocal apparatus, due to age related processes or diseases affect the voice quality. On the other hand also long-term muscular adjustments to a speaker's specific way of phonation [cf. 1] cause changes in voice quality which can serve as a social marker [2] and convey characteristics of a community.

There are surprisingly few studies focussing on language dependency of voice quality features. Previous research found language specific differences with respect to F0 between speakers of several languages [3, 4, 5]. Wardrip-Fruin [6] reports a language dependency rather for F0 modulation than for F0. The present study examined in addition to these parameters of F0 also parameters on a more delicate level. Harmonic-to-Noise ratio and the micro fluctuations of amplitude (shimmer) and of frequency (jitter) are parameters correlated with the impression of "roughness" in voice [cf. 11]. They are used in phoniatry as parameters of acoustic diagnostics but they are part of every naturally generated voice signal.

The voice signals examined in the presented study are voices of speakers of German, Italian and Polish, languages which are stereotypically very different from one another. Italian speakers have the reputation of having a rather "rough" voice while Polish speakers are known as having a more "clear" or "bright" voice quality.

The research question of the presented study is whether differences in the acoustic manifestations of "roughness" in voice can be found between language groups in absence of other external factors affecting voice quality, like smoking habits or the influence of alcohol.

2 METHOD

2.1 Speakers

A total of 145 adult male speakers were studied. The study was confined to male subjects in order to avoid vocal changes in females which are dependent on the hormonal monthly cycles [7]. The German-speaking group consisted of 51 speakers with an average age of 26.8 (range: 20-38). The Polish-speaking group consisted of 48 speakers, average age is 26,3 (range: 20-37) and the Italian-speaking group consisted of 46 speakers, with an average age of 26.1 (range: 18-43). To assure group homogeneity special care was taken. Only younger males were considered in order to avoid changes in vocal behaviour owing to vocal overuse syndromes or long-term (ab)use of alcohol. Speakers were recorded at approximately the same time of day in order to avoid effects on F0 induced by time of day [cf.8]. University students were chosen so as to avoid potential social differences in vocal behaviour. Subjects who had consumed alcohol within the last 24 hours prior to the recording were excluded from the experiment. Smokers were excluded from the experiment (subjects who had not smoked for a year were admitted, though) [12], as well. Finally, speakers with speech/language defects or hearing deficits were excluded from the study.

2.2 Recordings

Recordings were made at the Universities of Katowice

ISBN 1-876346-48-5

(Poland), Trier (Germany), Trieste and Verona (Italy) in a controlled environment. The individual locations were not completely sound-proof, but exhibited an ambiance which was characterised by low reverberation levels. The mouth-to-microphone distance was kept constant at approximately 5". A digital audio recorder Sony TCD-D7 was used for all recordings in conjunction with a condenser microphone Sony ECM-959.
Subjects read a neutral phonetic text ("The North Wind and the Sun") first and then were asked to phonate the vowel /a/ for at least five seconds at a comfortable effort level.

2.3 Analyses

The data were recorded into two different computer systems for analysis. The F0 analyses were carried out using a stand-alone computer MEDAV Spektro 3000. The sampling rate was 12.5 KHz with a 16 bit quantization. The *Soundscope*® for Apple MacIntosh software package was used for the remainder of the analyses. Fundamental frequency values were determined using two independent algorithms: a peak-picking (SIFT) and a cepstrum-based algorithm. A maximum difference between algorithms of 3Hz was considered acceptable; the more plausible one of the two (based on visual inspection of the corresponding spectrograms) was used in the statistical evaluation. Jitter, shimmer, and HNR were analysed based on the full bandwidth provided by the DAT recordings (44 kHz) using the Yumoto et al. algorithm for HNR [9], and the algorithms for RAP (Relative Average Perturbation) and APQ (Amplitude Perturbation Quotient) [10]. Only the central part of the sustained vowel was used for these measurements in order to avoid onset and offset effects (i.e. the initial and final 500 ms were discarded).

3 RESULTS AND DISCUSSION

3.1 Average Fundamental Frequency

F0 was measured for two conditions: reading of the neutral text and during the center section of phonating a sustained vowel /a/. The F0 measurements in the sustained vowel were carried out in order to control for a possible deviation of subjects from their habitual F0. If large differences had been found, this might have presented a problem for further measurements which were derived from the sustained vowel condition. Mean values for this condition exceed those done in the text reading condition by 2 Hz on average, but owing to individual variation this difference is far from significant. It thus seems safe to assume that subjects did not utilize unnatural phonatory mechanisms to produce the sustained vowels and that the latter can be analyzed without reserve. Table 1 displays the means for F0 measured in the reading text condition. As far as average voice fundamental frequency is concerned, the German group exhibited the lowest values whereas the highest values were measured for the Polish subjects.

	German	Italian	Polish
mean(Hz)	117.4	120.90	122.40
s (Hz)	10.60	11.42	11.74
Range	96.25 - 143.75	102.25 - 145	100.5 - 153

Table 1: Results of F0 measurements

Figure 1 shows the distribution of F0 values according to language groups. They exhibit quite a constant shift of the Polish group to higher values as compared to the German group. Analyses of variance and post host testing according to Tukey show that the differences between these two groups are significant at the 10% - level ($F(2,142) = 2.45$, $p = 0.08$).
These findings confirm the results found in previous

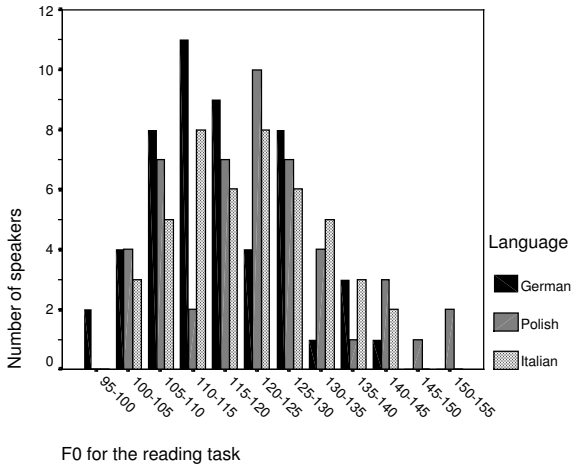

Figure 1: Distribution of F0 values in Hz for language groups

research [3, 4, 5] and thus provides support for the hypothesis of intercultural or/and interlanguage differences in fundamental frequency. The magnitude of these differences is in good accord with the results of [3] and [4] for speakers of different language groups; it is, however, smaller than the differences found by [5] for German as opposed to Turkish speakers.

3.2 F0 standard deviation

Wardrip-Fruin [6] has been the only researcher up to now to observe between-language differences in F0 standard deviation. This finding could not be replicated in the present study. The means and the standard deviations are listed in Table 2. No significant differences were found across languages. Instead, the values show a large degree of homogeneity.

	German	Italian	Polish
mean (Hz)	11.43	11.42	11.74
s (Hz)	2.77	2.07	2.22
Range	6.50 – 17.75	7.75 – 16.80	8.00 – 16.00

Table 2: Results of F0 standard deviation measurements

ISBN 1-876346-48-5

3.3 Harmonics-to-noise ratio (HNR)

The results from the HNR analysis are shown in Table 3. The HNR values found in this study are relatively high as compared to other studies which utilized the same algorithm [e.g. 9,12]. A possible explanation for this finding is that speakers were very carefully selected, and a number of factors causing low HNRs were deliberately ruled out.

	German	Italian	Polish
mean (dB)	15.80	14.39	17.14
s (sB)	3.62	3.85	3.79
Range	8.11-23.32	4.41-20.30	7.19-25.56

Table 3: Results of HNR measurements

Figure 2 shows the box plot for the language groups. Within-group ranges prove to be very similar. The Polish group is exceptional in the sense that their median HNR is higher than the value for 75% of the German speakers, and 25% of the Polish speakers reach HNR values above the individual maximum for Italians. The differences between the Polish and the Italian group reached significance on the 1% -level ($F(2,142)= 6.31$, $p = 0.002$) in an additional analysis of variance and post hoc testing according to Tukey.

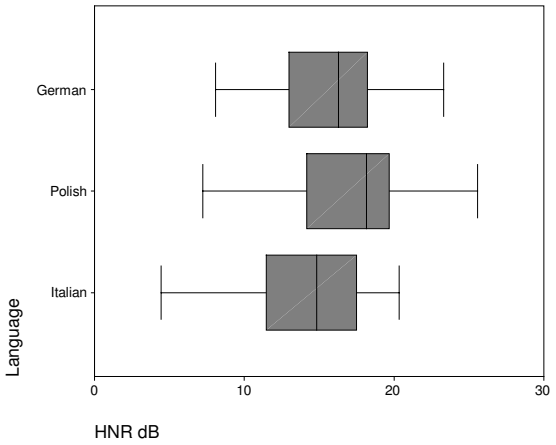

Figure 2: Box plot of HNR according to language groups

Hillenbrand [15] reports that different degrees of HNR correlate positively with the perception of "brightness". The Polish speakers studied here exhibit very high values in this respect. On the other hand, low HNR values have been shown to correlate well with the percept of “roughness” [10, 11]. The findings for our Italian speakers implicate thus, that voices of those speakers might be perceived as more “rough” with respect to ones of the Polish speakers.

3.4 Shimmer

Table 4 shows the mean shimmer-values for the language groups. Again the means tend to be lower than those found in previous studies using the same algorithm [e.g. 12]. The careful selection of subjects provides again a good explanation.

	German	Italian	Polish
mean (%)	2.49	3.75	2.73
s (%)	0.85	1.29	1.41
Range	0.99-5.24	2.15-8.25	0.65-6.29

Table 4: Results of shimmer measurements

Figure 3 shows the grouped distribution of measurement results for the different languages. The relatively high minimum as well as the range of the Italian speakers is in good accord with the values which [12] measured for smoking subjects. High shimmer values correlate with the impression of “roughness” in voice [11]. Thus, it can be assumed that shimmer would contribute to the impression of a more rough voice for an Italian speaker

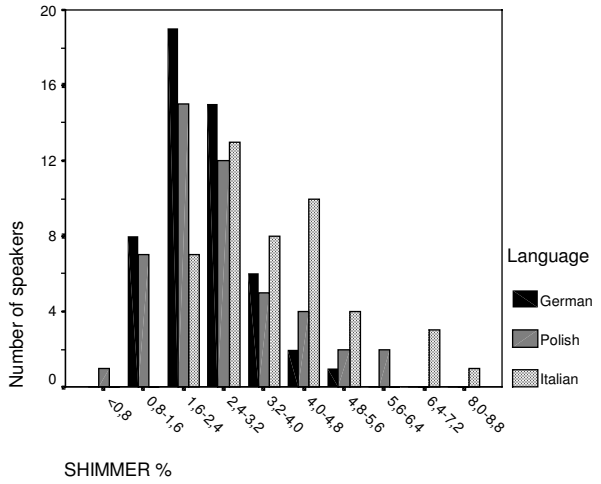

Figure 3: Distribution of shimmer-values for language groups

rather than for one of the other two groups. According to the post-hoc test (Games-Howell) the Italian group differs from the other two at the .01% level ($F(2,142) = 14.65$, $p =0.00$).

3.5 Jitter

The results for vocal jitter are summarized in Table 5. In short, no statistically significant differences were found between language groups.

	German	Italian	Polish
mean (%)	0.43	0.51	0.46
s (%)	0.34	0.42	0.35
Range	0.09-1.63	0.15-1.66	0.09-1.97

Table 5: Results of jitter measurements

Individual values for most speakers are well below the normative value of 1%, which is cited in the literature as a maximum for healthy voices [13]. The majority of speakers in all language groups can be found within that range; however, in every group there are also values well exceeding 1%. This seems to justify a rather liberal view on normative values (i.e. < 2%) as expressed by [14].

ISBN 1-876346-48-5

The role of jitter in the perception of roughness in voice is controversial. Previous research on synthetic stimuli stress the role of jitter as more relevant relative to the one of shimmer. On the other hand, studies with natural stimuli interpret shimmer and HNR to be the best predictors of roughness [11].

4 CONCLUSION

The results of the presented study confirm the hypothesis of intercultural or/and interlanguage differences in voice quality. The fact that the measurements were carried out on the basis of a comparatively large number of recordings adds weight to previous findings regarding F0 [3, 4, 5].
In addition differences on a more delicate level of voice quality have been found. The parameters HNR and shimmer varied between speakers of German, Italian and Polish making these speaker groups differing from one another by different parameters. The Polish group exhibited the highest values with respect to HNR and varied therefore from the Italian group. On the other hand the Polish speakers showed also the biggest F0 values and differed in this parameter significantly from the German group. Hillenbrand [15] reports that different degrees of HNR correlate well with the perception of "brightness". The Polish speakers studied here exhibit very high values in this respect. F0 correlates negatively with “roughness” [11], in this case high F0 value might in addition contribute to the impression of a “bright” voice of the Polish speakers. As far as parameters indicating vocal roughness are concerned, the Italian group is very different from the other two.

With respect to the perceptive domain it may be assumed – actual tests pending – that the voices of Polish speakers will be perceived as "bright", whereas those of Italian speakers will be rated as "rough".

Another interesting arising question is the one concerning the relation between the cycle-to-cycle micro fluctuations and HNR. Sources of noise components captured by HNR are said to be manifold and jitter and shimmer might be some of the sources of noise. As the Polish group exhibiting the highest HNR values at the same time shows higher shimmer and jitter values than the German group which scores rather low in HNR, these results indicate that there is no simple linear relation between minimal cycle-to-cycle fluctuations on the one hand and HNR on the other.

REFERENCES

[1] Laver, John (1980): The phonetic description of voice quality. Cambridge.

[2] Esling, John (1978): The identification of features of voice quality in social groups. In: Journal of the International Phonetic Association 8, 18-23.

[3] Hanley, T.D. / Snidecor, J.C. / Ringel, R.L. (1966): Some Acoustic Differences among Languages. In: Phonetica 14, pp. 97-107.

[4] Majewski, W. / Hollien, H. / Zalewski, J. (1972): Speaking fundamental frequency of Polish adult males. In: Phonetica 25, pp. 119-125.

[5] Braun, Angelika (1994): Sprechstimmlage und Muttersprache. In: Zeitschrift für Dialektologie und Linguistik 51, pp. 170-178.

[6] Wardrip-Fruin, Carolyn (1989b): Vocal fundamental frequency: variation by language, language group and sex. Paper presented at the Meeting of the Acoustical Society of America, Nov. 28, 1989, St. Louis [unpublished ms.].

[7] Hirson, Allen / Roe, Sam (1993): The stability of voice and periodic fluctuations in voice quality through the menstrual cycle in young women. In: Journal of the British Voice Association 2, pp. 78-88.

[8] Garrett, Kathryn L. / Healey, E. Charles (1987): An acoustic analysis of fluctuations in the voices of normal adult speakers across three times of day. In: Journal of the Acoustical Society of America 82, pp. 58-62.

[9] Yumoto, Eiji / Gould, Wilbur J. / Baer, Thomas (1982): Harmonic-to-noise ratio as an index of the degree of hoarseness. In: Journal of the Acoustical Society of America 71, pp. 1544-1549.

[10] Koike, Yasuo / Takahashi, H. / Calcaterra, T. C. (1977): Acoustic measures for detecting laryngeal pathology. In: Acta Otolaryngologica 84, pp. 105-117.

[11] Martin, David / Fitch, James / Wolfe, Virginia (1995): Pathologic voice type and the acoustic prediction of severity. In: Journal of Speech and Hearing Research 38, pp. 765-771.

[12] Braun, Angelika (1994): The effect of cigarette smoking on vocal parameters. In: Proceedings of the ESCA Conference on Speaker Identification Verification Recognition, Martigny, pp. 161 - 164.

[13] Heiberger, Vicki L. / Horii, Yoshiyuki (1982): Jitter and shimmer in sustained phonation. In: Lass, Norman J. (ed.) Speech and language. Advances in basic research and practice. Vol 7. New York etc., pp. 299-332.

[14] Klingholz, Fritz (1991): Jitter. In: Sprache – Stimme – Gehör 15, pp. 79-85.

[15] Hillenbrand, James (1988): Perception of aperiodicities in synthetically generated voices. In: Journal of the Acoustical Society of America 83, pp. 2361-2371.

ISBN 1-876346-48-5