Perceptive and acoustic measurement of average speaking pitch of female and male speakers in German radio news

Sven Grawunder, Ines Bose, Birgit Hertha, Franziska Trauselt, & Lutz Christian Anders*

Department of Speech Science and Phonetics, Martin Luther University, Halle
*Max Planck Institute for Evolutionary Anthropology, Leipzig
grawunder@eva.mpg.de,{anders;bose}@sprechwiss.uni-halle.de

Abstract

The average pitch of 68 news broadcasters (34 female / 34 male speakers) was evaluated by 6 expert listeners. Additionally, the average fundamental frequency for all samples was analyzed by means of a series of standard pitch detection algorithms. The results show a strong correlation of acoustic mean and auditory median values for male voices, whereas the auditory mean values female voices are slightly higher than their acoustic counterparts. Vice versa interlistener reliability is higher for the evaluation of female speakers, whereas listeners report more difficulties in rating for male speaking voices.

Index Terms: average speaking pitch, auditory measurement, speaking fundamental frequency, gender.

1. Introduction

The initial impetus for this pilot study was the impression that public vs. private (commercial) radio stations differ in their choice or preselection of speaker voices. In general speaking voices (male or female) in radio news broadcasts are reported as unusually deep (cf.[8], [21]). The reason generally given is role stereotypes, which include the conventional ascription of authority, competence, and reliability to deep voices [14]. Voice usage in public radio reflects the main emphasis which is placed on officiality and objectivity. For commercial radio, on the other hand, the entertainment aspect is emphasized by a more emotional use of voices. Therefore, we would assume that news broadcasters' voices on 'public' stations would be perceived as deeper than those of commercial stations.

To test this hypothesis we assumed that such evaluations of voice ranges (deep, mid-high, high) would correspond with the estimates of average speaking pitch (henceforth average pitch) [German 'mittlere Sprechstimmlage']. Thus average pitch is considered as an indicator of subjective (tonal) perception. Traditionally, (German) speech science has preferred to measure average speaking pitch by perceptual means. To be sure, the average vocal F0 values (as an acoustic parameter), also known as speaking fundamental frequency [1], is also a fairly widely used parameter. In previous phonetic work [21, p93] average pitch and speaker fundamental frequency have often been assumed to be identical or at least isomorphic. On the other hand some authors [16, 11] have argued that pitch perception is also influenced by spectral voice characteristics in addition to F0. Hence a special study was carried out, critically comparing the two methods.

We define the average speaking pitch as the perceptually (auditorily) obtained average pitch value of a speaking voice within a given utterance. Average speaking pitch is a paralinguistic feature with a strong impact on the listener's overall perception of the utterance. Here hearer's identification of the speaker's sex (gender) and his ascription of the speaker's habitual attributes and actual emotional state come into play. (Clearly the speaker also also typically adjusts speaking pitch to his/her intended effect: integrity, dominance; animating the interlocutor, evoking compassion etc.).

On the one hand, the range of average speaking pitch within a given utterance is also determined by anatomical features of the speaker (size of the larynx, length of the vocal folds). On the other hand the average pitch is also determined by the actual unintentionally affected (emotional) state or disposition of the speaker. Such an emotional disposition can result in corresponding tension states of the respiratory and phonatory muscles. In addition to this an average speaking pitch can be selected deliberately, depending on the speaker's purpose. As such it can be influenced by linguistic (textual), emotional and situation-specific factors.

For speakers of German no comprehensive listener-based source exists giving critical pitch thresholds for rating of a voice as high – mid-high – deep. An approximate idea can be gleaned from the literature [20, p99]. For male speakers a range from G2 (97.9Hz) to C3 (130.8Hz) is considered to be 'normal'. For female speakers and children a range from A3 (220Hz) or G3 (195.9Hz) to C4 (261.6Hz) is considered to be normal. From our observations these values seem to be relatively high estimates: (German) female voices may be ascribed as deep if their average speaking pitch is below G#3 (207.6Hz), (German) male voices if their average pitch is below G#2 (103.8Hz). We also observed that female pitch movements tend to be more variable than male pitch movements.

2. Corpus & Method

2.1. Corpus

In comparison with other more emotive formats of radio broadcasts, news broadcasts are meant to be purely informative and institutionalized texts which are presented in a factually and neutral speech manner. This makes news a good subject for perceptual measurement of average pitch.

Table 1: *corpus description.*

radio broadcasts Σ 68	public stations Σ 36	commercial stations Σ 32
34 female voices	18	16
34 male voices	18	16

68 audio recordings of news broadcast from different German radio stations were investigated altogether. Table 1 shows which recordings the corpus is composed of. The samples have an average duration of 81sec. The recordings were made by means of a minidisc recorder (Sony MR7). The areal selection of (Federal German) radio stations was random, but with the requirement of equal proportions for gender and station government [4, 19].

2.2. Perceptual Method

The perceptual (auditory) method of average speaking pitch measurement consisted of a subjective evaluation by a trained and experienced listener. From reports by such evaluators, we know that the average speaking pitch is a more 'virtual' parameter. Most listeners actively follow the pitch movement during the utterance (silently or half-loud); only then do they decide on the virtual center of the perceived tones. This pitch corresponds approximately to the tone which is represented most frequently within the utterance, but it also tends to be the tone of the crucial stressed syllables. Experts are able to ignore loudness and textual features as well the nontypical sequences (e.g. creaky voice) often found at the beginning and end of an utterance (sentence)(cf. [16]).

For this study 6 expert listeners (5 female /1 male, between 27 and 45 years of age) were asked individually to judge the samples in two separate runs (female/male speakers). Each of the experts (speech scientist/speech pathologist) is musically trained and has experience in obtaining the average pitch from a subject's running speech. The experts were allowed to judge after repeated listening and with the aid of a piano or a tuning fork. The average speaking pitch estimations were provided as musical notes which have been afterwards transformed into numerical values (cent, Hertz). In addition to the perceived musical note the experts were also ask for a self report about the difficulty or easiness of measurement per speaker (5 degree scale).

2.3. Acoustic Method

The speaking fundamental frequency (SFF) was obtained for all samples by means of the software PRAAT [2]. F0 was obtained in a first step over the total duration of the sample. In a second step the sample was divided into three equally long parts, which each served again for F0 measurement. In addition to the mean and median value of F0 for each sample (each news cast) also the standard deviation, maximum, and minimum were computed for each sub-sequence.

For pitch detection we choose the PRAAT standard setting of an autocorrelation algorithm with a time step window of 0.01 sec (100 pitch values per second). In order to see how SFF depends on the applied pitch detection settings the according parameters were variegated systematically in 10 different runs (accuracy: yes/no; silence threshold: 0.03; octave jump cost: 0.35, 0.37; voicing threshold: 0.45, 0.47, 0.49, 0.51; voiced/unvoiced cost: 0.14; pitch ceiling: 400). Since the maximum variation range of a given pitch measurement would not exceed 0.7 semitones the results appeared to be fairly stable and robust. Therefore we report only one run of acoustic data exertion. For measurement of speaking ranges (maximum value - minimum value) a 0.7 voicing threshold has been applied.

3. Results

3.1. Auditory and Acoustic Determination of Average Speaking Pitch

In order to compare both measurements we have chosen the cent-scale. One semitone, the twelfth part of an octave, equals 100cent, which corresponds to a frequency ratio of $1 : \sqrt[12]{2}$. Since A4 (440Hz) is given with 5700cent the basis of the cent-scale is made up by C0 (16.35159783Hz). In this way we can also use the cent-scale for nominal representations of pitch tones (cf. figure2).

As well as for the acoustic measures also for pitch values of auditory based estimations the mean ($\bar{x}$), median ($\tilde{x}$) and standard deviation (SD) value over all 6 experts have been determined.

As expected the pitch values voices of news broadcasters appear as rather low. Male speakers show mean values of acoustically obtained F0 between 89.2Hz and 132.2Hz and auditorily obtained pitch between 90.2Hz and 112.0Hz. And female speakers show F0 values between 138.5Hz and 198.Hz and pitch values 173.3Hz and 211.9Hz.

Table 2: *average values of acoustic (ac) and auditory (au) mean ($\bar{x}$) and median ($\tilde{x}$) pitch measures (in cent, Hertz and musical note), arrows indicate the derivation from the ideal center frequency of the note*

gender	ac $\bar{x}$	ac $\tilde{x}$	au $\bar{x}$	au $\tilde{x}$
male speakers	3255cent 107.7Hz A2↑	3185cent 103.5Hz G#2	3198cent 104.2Hz G#2↑	3197cent 103.9Hz G#2
female speakers	4075cent 172.6Hz F3↓	4028cent 168.1Hz E3↑	4292cent 195.6Hz G3	4286cent 194.8Hz G3↓

Nevertheless, regarding the station government there is only a significant difference in F0 mean (and median) values between public and private stations for male (for $\tilde{x}$ and $\bar{x}_3$ p=0.004)but not for female broadcasters which can be taken into account.

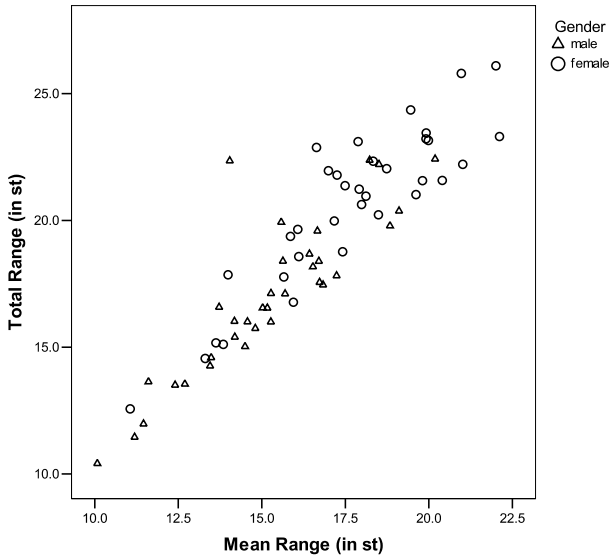

Figure 1: measured F0-ranges (in semitones); x-axis=mean range over 3 sequences of a sample; y-axis=total range over one sample

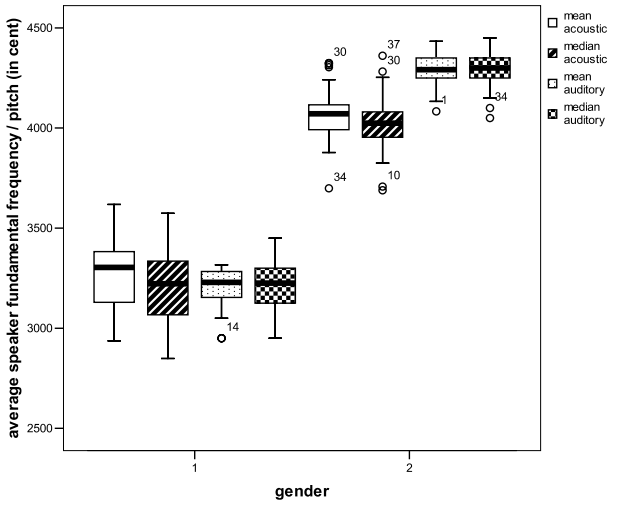

Figure 2: global results of pitch and SFF (in cent) for the whole corpus (gender: 1=male, 2=female)

The suggested differences in F0-variability which would be shown in SD of pitch values (in semitones) measured over the total duration shows significant differences between genders ($p < 0.05$) but not for stations. According to the results of a Levene's test and a given 95% confidence interval for all pitch values (mean and median) the variances between the groups of male and female speakers can be assumed as equal ($p > 0.05$).

However the range and here especially the mean range of minimum and maximum values over the three sequences illustrate these results. All broadcasters show a wide scope of fundamental frequency ranges (cf. figure1). Whereas for male broadcasters this scope is slightly smaller (10-19st) the range for female broadcasters (11-20st) stays also in the upper band. These ranges of female broadcasters are perspicuously larger than those measured in previous research (N=5) [21]. Whereas both groups show normally distributed scores female pitch and SFF values exceed a positive kurtosis vis-a-vis a negative kurtosis of male speakers. Also here no significant effect regarding the influence of public or private station can be described.

3.2. Critical Evaluation of Methods

3.2.1. Differences Within the Group of Experts

In order to assess the homogeneity of pitch evaluations per gender group of speakers the concept of interrater (or interlistener) reliability [15] was adopted. After the determination of ranks for the individual values separately for experts and gender, the middle ranks per objects (values per speaker) have been determined. Then the mean absolute difference between individual ranks and middle rank per object were compiled. In this way the null hypothesis is formulated that if the ratings of experts is homogeneous also the mean absolute differences in the two (gender) groups should be homogenous. This was then tested by means of the Mann-Whitney U-Test.

The results ($U = 405.4, N_{gender1} = N_{gender2} = 34, exactP = 0.034$) show a slightly more heterogenous expert rating of male speakers. However since we had only a small group of experts and proportion of female experts "overheaded" which leads again to the question of cross-gender difficulties in evaluation of pitch.

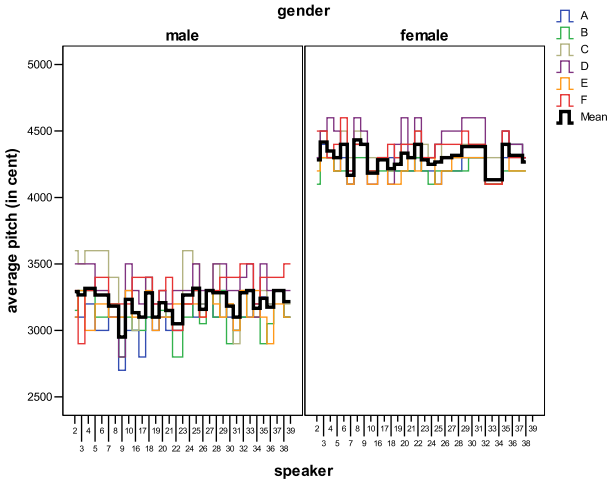

Figure 3: distribution in pitch evaluation of male and female speakers shown for 6 experts and their mean value (black bold line)

3.2.2. Expert Group vs. Acoustic Measures

A characterization regarding the relation between acoustically and perceptually measured values was achieved by comparing means of mean and median pitch values and (simple and averaged) mean and median SFF values. For male voices total mean of SFF correlates strong with the mean value of pitch evaluations and for female voices (cf. table3). If we consider differences of means we find best convergence of median of SFF together with mean and median of pitch in male voices. Otherwise all other means differ significantly.

Table 3: *correlation coefficient r values (Pearson's product-moment correlation) of acoustic (ac) and auditory (au) mean ($\bar{x}$) and median ($\tilde{x}$) pitch measures, $p > 0.001$*

measures	male voices	female voices
ac $\bar{x}$ & au $\bar{x}$	0.859	0.760
ac $\bar{x}$ & au $\tilde{x}$	0.858	0.693
ac $\tilde{x}$ & au $\bar{x}$	0.856	0.715
ac $\tilde{x}$ & au $\tilde{x}$	0.856	0.653
ac $\bar{x}_3$ & au $\bar{x}$	0.859	0.763
ac $\bar{x}_3$ & au $\tilde{x}$	0.858	0.695

There has been found no correlation between self report of difficulty (arduousness) in rating and the difference between auditory and acoustic values and no specific tendency or gender preference. Regarding these findings it is surprising that auditory and acoustic measurements for male voices are nearly identical (difference of means about 1st = 100cent) but ratings for – according to the self reports – easier ratable female voices deviate stronger from acoustic measurements (cf. figure 2). Presumably one reason may be given by the constant higher F0-variability (cf. figure1) of female speakers which may cause a different auditory impression (cf. [18]).

4. Discussion & Conclusion

Though the SFF of news broadcasters can be assigned to low voice ranges there is no general trend regarding private or public radio stations. Whether those speakers which use a deeper voice have indeed a physiologically deeper voice than others or just prefer or use the deep range of their voice stays unanswered. We assume that the latter is the case, since we see here another instant of speaking role adequate behaviour, i.e. use of register.

Over longer speaking passages ($>$ 30sec) subjective average pitch tone impression and acoustic measurement of average vocal fundamental frequency show very similar results. The occurring deviations do not lead to substantial different interpretations of speaking behaviour, but there seems to exist a remarkable deviation from the acoustic measurement for pitch evaluation of female speakers.

The self report about the arduousness of evaluations per listener and speaker vis-a-vis heterogeneity of evaluation results would lead to the assumption that evaluation of voices in cross-gender perspective causes more difficulties than in equal-gender perspective. However the evidence points to the opposite.

In order to explain these results possible influence variables on pitch evaluation (see section 1) need to be discussed. Since this study controlled for text purpose and emotive state, F0- variability, F0-range and voice quality (timbre) come into question. F0-variability (SD) showed no significant difference. In this way the obtained higher speaking ranges for female speakers and the specific spectral characteristics of female voices remain as explanatory ground.

This paradox situation needs to be further investigated. Therefore forthcoming studies are based on a larger group of experts with an equal gender proportion, whereas the studies seek also after dependencies of a range of characteristics in the stimuli (stimulus length, F0-variability, text type etc.).

5. Acknowledgements

We would like to thank the six anonymous expert raters. And we are also very grateful to Roger Mundry (MPI EVA, Leipzig) for his kind assistance with the statistics of interrater/interlistener reliability. Any mistakes and misinterpretations remaining in the text are to the authors' full responsibility.

6. References

[1] Baken, R. J., & Orlikoff, R. F. (2000): *Clinical measurement of speech and voice* (2. ed.). San Diego: Singular Thomson Learning.

[2] Boersma, P., Weenink, D. (2005): PRAAT 4.4 `http://www.praat.org`

[3] Braun, A., (1992): Zur Bedeutung des Merkmals "mittlere Sprechstimmlage" in der forensischen Sprechererkennung. In *Phonetik und Dialektologie. Joachim Göschel zum 60. Geburtstag.* Marburg: Verlag der Universitätsbibliothek, 1-26.

[4] Braun, A., (1994): Sprechstimmlage und Muttersprache. *Zeitschrift für Dialektologie und Linguistik*, 61, 170-178.

[5] Eriksson, A., & Traunmüller, H. (2002): Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Percept Psychophys*, 64(1), 131-139.

[6] Hertha, B. (2005): Auditive und akustische Untersuchungen zu den mittleren Sprechstimmlage von Nachrichtensprechern im ffentlich-rechtlichen und privaten Rundfunk. Unpublished Diploma thesis, Martin-Luther-Universitt Halle-Wittenberg, Halle.

[7] Hess, W., Sendlmeier, W. F., & Gesellschaft für Angewandte Linguistik. (1992): Beiträge zur angewandten und experimentellen Phonetik: [basiert auf Vorträgen und Vorführungen, die im Rahmen der Sektion Phonetik auf der Jahrestagung der Gesellschaft für Angewandte Linguistik (GAL) 1990 in Bonn präsentiert wurden] (*Zeitschrift für Dialektologie und Linguistik Beihefte*; 72). Stuttgart: Steiner.

[8] Hilbert, P. M. (1986): Die Aussprache des Nachrichtensprechers im Rundfunk als kulturelles Vorbild. In: F. R. Varwig (Ed.), Sprechkultur im Medienzeitalter (*Sprache und Sprechen*; 16). Frankfurt am Main: Scriptor.

[9] Hirano, M., Vennard, W., & Ohala, J. (1970): Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles. *Folia Phoniatr (Basel)*, 22(1), 1-20.

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

[11] Neppert, J. M. H. (1999): *Elemente einer akustischen Phonetik* (4., vollst. neu bearb. Aufl ed.). Hamburg: Buske.

[13] Ohala, J. J. (1983): Cross-language use of pitch: an ethological view. *Phonetica*, 40(1), 1-18.

[14] Sendlmeier, W. (2005): Sprechwirkung - Sprechstile in Funk und Fernsehen (*Mündliche Kommunikation*; 3). Berlin: Logos.

[15] Stemler, S. E. (2004): A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Practical Assessment, Research & Evaluation, 9(4). Retrieved from http://PAREonline.net/getvn.asp?v=9&n=4 (02.04.2006)

[16] Terhardt, E. (1998): *Akustische Kommunikation.* Berlin [u.a.]: Springer.

[17] Traunmüller, H., & Eriksson, A. (1995, manuscript): The frequency range of the voice fundamental in the speech of male and female adults. Retrieved from http://www.ling.su.se/staff/hartmut/aktupub.htm (12.12.2005)

[18] Traunmüller, H., & Eriksson, A. (1995): The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *J Acoust Soc Am*, 97(3), 1905-1915.

[19] Trauselt, F. (2002): Untersuchungen zur Mittleren Sprechstimmlage bei Nachrichtensprecherinnen im öffentlich-rechtlichen und privaten Hörfunk. Unpublished Diploma thesis, Martin-Luther-Universität Halle-Wittenberg, Halle.

[20] Wendler, J., & Seidner, W. (1997): *Die Sängerstimme: phoniatrische Grundlagen der Gesangsausbildung*. Berlin: Henschel.

[21] Wittlinger, I., & Sendlmeier, W.F. (2005): Stimme und Sprechweise erfolgreicher Frauen - eine akustische und auditive Analyse. In: W. Sendlmeier (Ed.), Sprechwirkung - Sprechstile in Funk und Fernsehen (Mündliche Kommunikation, Bd.4). Berlin: Logos.