Voice Pleasantness of Female Voices and the Assessment of Physical Characteristics

Vivien Zuta[1, 2]

[1] FIAS, Frankfurt Institute for Advanced Studies, Frankfurt, Germany
[2] Institute of Phonetics, Goethe University, Frankfurt, Germany
zuta@fias.uni-frankfurt.de

Abstract. It has been demonstrated that there seem to be non-linguistic patterns from which listeners refer to the appearance of a speaker. This study completes the collected voice parameters, which count so far as indicators of physical attributes and as factors for voice attractiveness. Since scientists tend to prefer male voices for such analyses this one is based on female voices and both-gendered judges. 20 female voices were played to 102 listeners, 52 male and 50 female. Because of different rating strategies of female and male listeners, the group specific rating strategies were compared; a bimodal rating dispersion was assumed and mostly proved.

Keywords: voice attractiveness, voice production, voice perception, vocal tract

1 Introduction

Most people have had the experience of surprise, by the appearance of a person they had hitherto known only from his or her voice. However, the surprised reaction indicates that there must be exact cues in speech that lead the listener to create a certain picture of the speaker. In previous studies [1] the author examined phonetic criteria of male voices and found out, that female listeners usually assign attractive body characteristics to speakers with attractive sounding voices. The rating of whether a voice was judged as attractive or not was carried out independently of F0, which was a big surprise, since 100% of the listeners declared to prefer deep sounding male voices. The cliché that attractive male voices need to have low F0 values was not proven. The reactions were splendorous and led to big media interest. (In a broader, more popular sense, see also [2]). In order to find out more interesting details about listeners judgements in terms of voices a more extensive study was carried out, which is still in progress and includes female voices as well. Some of the results are presented in the following article.

A closer look at the publications on this topic show how discordant the results are and how complex the research question is designed.

Several studies, mostly undertaken outside of the field of phonetics, deal with the topic of how feasible it is to infer the speakers' appearance from their voice. Some of these studies concentrate on the correlation of assumed physical attributes and their equivalents in vocal attractiveness. Collins & Missing [3] found that vocal attractiveness and the attractiveness of female faces are related and that female voices with higher fundamental frequency were judged as more attractive and as likely to be younger. Feinberg et al. [4] report a positive correlation between facial-metric femininity and vocal attraction, the latter being connected with higher pitched voices. Actual physical attributes and rated voice attractiveness were examined for example by Hughes et al. [5] who found a surprising correlation between Waist-to-Hip-Ratio (WHR) for women and Shoulder-to-Hip-Ratio (SHR) for men and the opposite-sex voice ratings. Voices of females with low WHRs and males with big SHRs were found to be consistently rated as more attractive. In 2002 Hughes et al. [6] found a correlation between the measurable bilateral symmetry of a speaker and his or her rated voice attractiveness. Additionally Evans et al. [7] discovered negative correlations between F0 for male voices and male body measures such as SHR. Furthermore they found that men with large body size and shape tend to have smaller formant dispersion than shorter men do. However they refer to the findings of Fitch and Giedd [8] whereby the spreading of overweight in western society weakens any statistical findings.

Other examinations that are more closely related to the field of phonetic science, deal with the question of whether physical body characteristics are related to measurable physical parameters in voice at all. So did Künzel [9] examine the correlation between the average fundamental frequency and speaker height and weight. He assumed contrarily to other examinations [10], that there is no acoustical parameter that could be counted as an indicator for height and weight estimations. Furthermore he postulates that speech signals do not contain any information about speaker's height or weight at all. Van Dommelen & Moxness [11] concluded that listeners correctly used speech rate for speaker weight estimations whereas low F0 was wrongly taken for large body estimations.

Lass & Davis [10] found that the accuracy of estimating the height of male and female subjects (in agreement with van Dommelen & Moxness [11]) and male weight on the basis of speech signals was higher than would be attributed to chance. Lass & Davis, however, were not able to determine the exact cues in the speech signals from which these conclusions were derived. Another investigation was performed by Bruckert et al. [12] who found that females were consistent in their judgment of pleasantness and in their height, weight and age estimations. Bruckert et al. concluded that pleasantness judgments were based mainly on intonation. Women tend to prefer voices with raising pitch. Another result was that female listeners were able to correctly estimate age and weight but not height. For the age estimation women would use lower-frequency formants with little formant dispersion as a sign for an older age. Bruckert et al. however were not able to determine which acoustic parameters were used to estimate the subjects weight correctly.

It is very interesting to see that some experiments do find correlations between body size and voice pitch and some do not. Mostly these differences are attributed to the varying experimental set ups and measurement methods. The hypothesis that a positive correlation between body size and F0 exists is due to the assumption that the lengths of the vocal folds increases with body size [13]. This however is not the case within the group of mammal species [14]. Another kind of experiment handles the question of whether the suspected body characteristics are consistent and/or close to the actual body characteristics of the speaker. Krauss et al. [15] examined the possibility to refer from ones voice to his/her photo (one out of two). The accuracy was better than chance and naïve listeners were able to estimate a speaker's age, height, and weight from voice samples *"nearly as well as they can from a photograph"*.

This topic is highly interesting since the correlation between body characteristics and measured values is usually quite low. It seems evident that there must be some cues in the speech signal to be identified, because, as it will be shown, correlations between actual and estimated characteristics were found.

2 Experiment

2.1 Material

The recordings for the experiment were conducted using a set of 40 native German speakers (20 female, 20 male) who were requested to narrate the fairy tale "Little Red Riding Hood". The fairy tale was spoken in standard-German at speaker specific speed, articulation and intonation. No guidelines were given. (Little Red Riding Hood was chosen because of its widespread familiarity.) The task for the listeners was to judge in reference to the sound they were listening to but not in reference to the content of what the speakers were saying. It was ensured that everybody who took part in the investigation was familiar with the fairy tale. Speakers were told to narrate the story in a natural way and they were recorded in a sound treated environment with professional equipment (see *2.6.1. Recordings)*.

2.2 Subjects

None of the subjects had a strong dialect, a speech disorder or was affected by an illness that could have had an effect on the voice quality. All of them were healthy and were recruited by the author and her co-worker. The participants were asked about their height, weight, age, hair and eye colour, education, smoking habits and if they ever had their nose broken. (Female subjects were also asked about hormonal contraceptives.)

Since only female voices are relevant for this article, the male speakers descriptive statistics is left aside.

Table 1. Descriptive statistics for subjects.

female subjects (N=20)	**mean**	**range**	
		min	**max**
age (years)	25 (5,29)	21	44
height (cm)	168 (6,30)	152	178
weight (kg)	62,4 (11,7)	50	100

The Voice Listeners were 107 adults. They were recruited through a couple of advertisements, which were placed all around the university, different institutes and via the Multimedia-System of the Goethe University Frankfurt. We also contacted different schools and companies directly with covering letters. They were allured with a little lottery, cookies, cakes and coffee.

5 listeners did not continue until the end of the evaluation and were not included in the analysis. 8 listeners were younger than 18years old. The statistic analysis was carried out twice, including and excluding those 8 listeners. Since the results did not differ significantly the age did not matter for this analysis. Leaving 52 male and 50 female listeners.

Table 2. Descriptive statistics for listeners.

	male listeners (N=52)			female listeners (N=50)		
	mean	range		mean	range	
		min	max		min	max
age (years)	24,76 (11,75)	16	63	24,52 (8,22)	17	56
height (cm)	181,89 (7,19)	165	195	168,14 (6,45)	170	185
weight (kg)	78,19 (11,33)	58	110	62,67 (10,49)	45	90

2.4 Stimuli Selection

Originally, it was planned to present the whole 40 stimuli to all listeners, but soon it became clear that this would take way too much time for each listener. So it was decided to present only 10 male and 10 female speakers to each listener and it already took them about 60 min. to go through the twenty voices. In order to get an equally distributed evaluation a script was computed which randomised the stimuli so that an equally distributed evaluation was assured. This randomisation guaranteed also that a fatiguing effect could be disregarded.

2.5 Procedure

The participants were all tested in the same manner. Each participant was seated in front of one computer and was told to follow the instructions that were written on the first page of the questionnaire. The participants were lead through the experiment on-screen and had to answer 20 questionnaires about 20 different voices. (For this article we refer to the female voices N=20 and all participants N=102 only.) The experimenter was seated in the same room to take care that participants would not communicate with each other. Since the experiment was undertaken in the rooms of the Institute of Phonetics and in different offices in the Frankfurt area, the experiment was run on Macintosh and Window PCs, depending on the availability of each company. It was made sure that the experimental premises were as analogue as possible. During the time of the experiment, there were no other activities, such as usual office business or university business, which could have sidetracked the participants. No time limit was given.

2.6 Preparing the Data

Recordings. Recording the subjects in a sound treated environment using the Microtrack 24/96 Professional 2-Channel Mobile Digital Recorder (M-Audio) created the stimuli. The voices were recorded 44,1 kHz, 16 bit mono.

Firstly the data was adjusted. In order to avoid extreme effects, such as too loud or too silent speech, all audio files were rescaled by setting the peak scale to 0,99 and the intensity to 70 dB. Afterwards the audio signals were cleared of laughs, coughs and abruptions since these particles are absolutely irrelevant for what was sought after. Given that we tend to investigate speech and voices these behavioural particles were not given further attention.

Using PRAAT software[1] all signals were segmented. Each signal obtained segmentation into two tiers. The first tier for F0, in which five segments[2] of each 5 seconds was measured in order to calculate average F0. Since a fairy tale was told it was not possible to measure F0 simply through the whole signal, because of some unwanted variations due to direct speech within the fairy tale. In order to measure F0 equally for all audio signals a script was used, to calculate F0 for all segments in a signal. The average of these results was taken respectively as F0 for the subjects. As reference served the PRAAT Voice Report function.

To be able to compare formant frequencies a second tier was created in which [a:], [e:], [i:], [o:] and [u:] were segmented. Each vowel was segmented five times, if this was not possible, it was marked in the statistics. A script ran through the tiers and calculated the mean formant values in the segmented parts of the signal. The average values were calculated and analysed.

3 Results

For the voices ANOVAs and/or Nominal Logistics were carried out. According to pre-experiments, male and female listeners are judging so differently that we are dealing with a bimodal distribution.

---

[1] www.praat.org

[2] Two segments at the beginning, one segment in the middle and two segments at the end of each signal.

3.1 Pleasantness of Female Voices

For both listener-groups a nominal logistic[3] using Chi-Square analysis and REML (Restricted Maximum Likelihood) method with *pleasantness of voice* as dependent variable and *actual age, height, weight, education, smoking habits, broken nose, F0, estimated age, estimated weight and estimated height* as independent variables and randomised subjects was carried out.

For female listeners interactions were found between *pleasantness* and: *actual age* χ^2=16.24, p=0.0001; *actual height* χ^2=12.41, p=0.0004; *education* χ^2=22.49, p=0.0004; b*roken nose* χ^2=8.09, p=0.0045; *smoking habits* χ^2=8.12, p=0.0044 and *estimated height* χ^2=6.12, p=0.0469. No interactions were found for w*eight* χ^2=2.59 p=0.1071, *F0* χ^2=1.91, p=0.1665; *estimated weight* χ^2=4.70, p=0.4530 and *estimated age* χ^2=0.49, p=0.9219.

For male listeners interactions were found between *pleasantness* and: *actual age* χ^2=11.14, p=0.008; *smoking habits* χ^2=6.18, p=0.0129 and *education* χ^2=17.58, p=0.0035. All estimated factors were significant as well: *estimated age* χ^2=16.28, p=0.0003, *estimated height* χ^2=6.22, p=0.0446, *estimated weight* χ^2=17.43, p=0.0038. No interactions were found for *actual height* χ^2=1.89, p=0.1696; *actual weight* χ^2=1.03, p=0.3095, *broken nose* χ^2=1.54, p=0.2144 and *F0* χ^2=1.01, p=0.3140.

Table 3. Dependencies for characteristics of female voices and listeners of both gender.

	female listeners (N=50)	**male listeners (N=52)**
actual age	0,0001	0.008
actual height	0,0004	n.s.
actual weight	n.s.	n.s.
education	0,0004	0,0035
smoking habits	0,0044	0,0129
broken nose	0,0045	n.s.
F0	n.s.	n.s.
estimated age	n.s.	0,0003
estimated height	0,0469	0,0446
estimated weight	n.s.	0,0038

3.2 Listeners' Judgments and Speakers' Actual Characteristics

For both listener-groups four separate nominal logistics using Chi-Square analysis and REML method with *estimated age, estimated height*, *estimated weight* and *estimated F0 (high, low, average)* as dependent variable and *actual age, height, weight, education, smoking habits, broken nose* and *F0* as independent variables were carried out.

For female listeners interactions were found between *estimated age* and: *actual age* χ^2=14.62, p=0.0022; *weight* χ^2=11.05, p=0.0115 and *smoking habits* χ^2=14.90,

[3] All Chi-Square-Tests in this article are Effect Wald Tests

p=0.0019. All other factors and interactions were not significant. Between *estimated height*, correlations were found for: *actual height* χ^2=10.0, p=0.0069; *weight* χ^2=10.05, p=0.0066; *smoking habits* χ^2=10.42, p=0.0055; *education* χ^2=23.2, p=0.0100; *broken nose* χ^2=7.27, p=0.0264; *F0* χ^2=9.29, p=0.0096. A*ctual age* did not show significant correlations with height estimations (χ^2=6.88, p=0.3324). Weight estimation was predicted only by *F0* χ^2=17.4353, p=0.0037 and *Age* χ^2=11.41, p=0.0438. For *estimated F0* as dependent variable, interactions were found for *Age* χ^2=8.08, p=0.0176, *education* χ^2=24.83, p=0.0057 and *F0* χ^2=18.40, p=0.0001.

For male listeners interactions were found only between *estimated age* and: *smoking habits* χ^2=20.52, $p \leq 0.0001$ and *F0* χ^2=27.46, $p \leq 0.0001$. No other significant interactions for this dependent factor were found. Interactions with *estimated height* were found for following independent factors: *height* χ^2=9.26, p=0.0097*, weight* χ^2=15.51, p=0.0004*, education* χ^2=35, p=0.0001*, broken nose* χ^2=12,36, p=0.0021, *F0* χ^2=11.15, p=0.0038. All other factors and interactions were not significant. Weight estimation was made using *actual age* χ^2=15.12, p=0.0084, and *height* χ^2=21.69, p=0.0168 only. For *estimated F0* as dependent variable, *Age* χ^2=20.30, $p \leq 0.0001$, *education* χ^2=28.75, $p \leq 0.0001$ and *smoking habits* χ^2=12.61, p=0.0004, showed significant relations. All other factors were insignificant.

Table 4. Interactions between actual and estimated characteristics for female voices and female listeners; independent variables vertically aligned.

	estimated age	**estimated height**	**estimated weight**	**estimated F0**
actual age	0,0022	n.s.	0,0438	0,0176
actual height	0,0115	0,0069	n.s.	n.s.
actual weight	n.s.	0,0066	n.s.	n.s.
education	n.s.	0,01	n.s.	0,0057
smoking habits	0,0019	0,0055	n.s.	n.s.
broken nose	n.s.	0,0264	n.s.	n.s.
F0	n.s.	0,0096	0,0037	0,0001

Table 5. Interactions between actual and estimated characteristics for female voices and male listeners; independent variables vertically aligned.

	estimated age	**estimated height**	**estimated weight**	**estimated F0**
actual age	n.s.	n.s.	0,0084	≤0,0001
actual height	n.s.	0,0097	0,0168	n.s.
actual weight	n.s.	0,0004	n.s.	n.s.
education	n.s.	0,0001	n.s.	≤0,0001
smoking habits	≤0,0001	n.s.	n.s.	0,0004
broken nose	n.s.	0,0021	n.s.	n.s.
F0	≤0,0001	0,0038	n.s.	n.s.

4 Discussion

Inconsistent to previous studies of preferences for women's voices the results show, that the evaluation of pleasantness of female voices occurs independently of voice pitch for both, male and female listeners. The results furthermore indicate that for male voice listeners F0 plays even a much smaller role than expected. Male listeners are not able to correctly assign a F0 description (low, average, high) to a woman's voice. F0-judgements of male listeners do not show any significant correlation with actual F0 χ^2=3.3688, p=0.1856 F0. Male listeners use F0 for height estimations and age estimations only and affiliate a lower F0 with higher age. Several studies ([3], [16]) have shown, that high-pitched female voices are indicators for female fertility and with that to be preferred. But as it is shown in the table below, men prefer average judgments instead of making a decision when the frequencies fall below 205Hz. Since men tend to have difficulties in correctly estimating F0 in general, they seem only to make differences between *high* and *not high* pitch. The pleasantness judgments show that men prefer younger estimated voices to older sounding ones, but that does not go along with F0. Actual F0 is no indicator for pleasantness (χ^2=1.01, p=0.3140).

Table 6. Pleasantness, age and F0 estimations compared with actual values for male listeners.

average estimated age	average actual age	estimated F0	actual F0	pleasantness
20	24	42% high 54% average 4% deep	205 Hz	68%
20-30	25	29% high 66% average 5% deep	200 Hz	68%
30-40	28	27% high 64% average 7% deep	187 Hz	34%

It is imaginable, that men subconsciously prefer voices above 200 Hz, (which is rather a normal than a significant high pitched female fundamental frequency) and affiliate youth with these voices. But fact is, that men do not necessarily prefer high-pitched voices, more importantly they put their focus on estimated age and favor young sounding voices. Another interpretation is that men think that young sounding voices have to have a certain high frequency. Possibly they are influenced by a cliché that leads to the judgments listed above. (Voice sounds young, must be high.) Male listeners cannot reliably differentiate between high/average and deep female voices. The physical parameter that is used by the male listeners of this experiment in order to judge female voices' age needs to be found out in further analysis of this data.

Female listeners however are able to correctly assign F0 estimates to the actual F0 value with χ^2=18.4987, p=0.0001, but in agreement to male listeners do not use F0 for pleasantness judgments.

Table 7. Pleasantness, age and F0 estimations compared with actual values for female listeners.

average estimated age	average actual age	estimated F0	actual F0	pleasantness
20	24	30% high	207 Hz	49%
		64% average		
		6% deep		
20-30	25	28% high	200 Hz	60%
		65% average		
		7% deep		
30-40	28	16% high	188 Hz	57%
		73% average		
		11% deep		

Female listeners used F0 for weight estimations but no correlation between estimated and actual weight was detectable. A deeper F0 leads female listeners to higher weight estimations. Creating two groups, the lighter judged group (50-60kg) had an average weight of 62kg and a mean F0 of 201 Hz. The heavier judged group of female speakers (60-75kg) had an *actual* average weight of 62kg and a mean F0 of 197 Hz. Since the actual average weights are exactly the same for both groups, weight estimations are correspondingly caused by the lower F0 value of one group.

Both, male and female participants equally referred from F0 to estimated height. It was shown in the statistical analysis that correlations between estimated and actual height were found. Interestingly, the judgments for this parameter were the only ones, which are consistent between both gender-groups.

Table 8. Height estimations and corresponding F0 values for both listener groups.

	estimated height	actual F0	N
male listeners	tall	194 Hz	10%
	average	198 Hz	62%
	short	202 Hz	28%
female listeners	tall	193 Hz	15%
	average	198 Hz	60%
	short	203 Hz	25%

Looking at the average F0 values it is obvious that the estimated height groups are extremely mixed. This is the only explanation for such low average F0 values for female speaker groups.

It can be summarized, that not only the pleasantness of ones voices but also the assessment of physical characteristics by ones voice is depending on many different factors. The complexity of a voice impression has to be seen in total and cannot be examined in disregard of interacting single factors.

As mentioned in the beginning, the work is still in progress. For instance the statistical analyses for *Acoustic Parameters and Speakers' Characteristics* are currently carried out and of course other examinations about the physical parameters that lead listeners to certain judgments about the owner of a voice will be undertaken. Since the corpus contains both (male and female) voices many more interesting results are to be expected.

References

1. Zuta, V.: Phonetic Criteria of Attractive Male Voices. In: Proc. 16th ICPhS Saarbrücken, pp. 1837--1840, (2007)
2. Zuta, V.: Warum tiefe Männerstimmen doch nicht sexy sind – Das Geheimnis unserer Stimme. Scherz Verlag, Frankfurt (2008)
3. Collins, S.A., Missing, C.: Vocal and visual attractiveness are related in women. Animal Behaviour. 65, 997--1004 (2003)
4. Feinberg, D.R., Jones, B.C., DeBruine, L.M., Moore, F.R., Law Smith, M.J., Cornwell, E., Tiddeman, B.P., Boothroyd, L.G. & Perrett, D.I.: The voice and face of woman: One ornament that signals quality? Evolution and Human Behavior. 26, 398--408 (2005)
5. Hughes, S.M., Harrison, M.A. & Gallup Jr., G.G.: The sound of symmetry: voice as a marker of developmental instability. Evolution and Human Behavior. 23, 173--180 (2002)
6. Hughes, S.M., Dispenza, F. & Gallup Jr., G.G.: Ratings of voice attractiveness predict sexual behavior and body configuration. Evolution and Human Behavior. 25, 295--304 (2004)
7. Evans, S., Neave, N. & Wakelin, D.: Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. Biological Psychology. 72, 160--163 (2006)
8. Fitch, W.T. & Giedd, J.: Morphology and development of the human vocal tract: a study using magnetic resonance imaging. Journal of the Acoustical Society of America. 106, 1511--1522 (1999)
9. Künzel, H. J.: How Well Does Average Fundamental Frequency Correlate with Speaker Height and Weight? Phonetica. 46, 117--125 (1989)
10. Lass, N. J. & Davis, M.: An investigation of speaker height and weight identification. Journal of the Acoustical Society of America. 60 (3), 700--703 (1976)
11. Dommelen, W.A. van & Moxness, B.H.: Acoustic parameters in speaker height and weight identification: sex- specific behaviour. Language and Speech. 38, 267--287 (1995)
12. Bruckert, L., Lienard. J.S., Lacroix A., Kreutzer, M. & Lebourcher, G.: Women use voice parameters to assess men's characteristics. In: Proceedings. Biological sciences, 273, 1582, 83--293. The Royal Society of London (2006)
13. Titze, I.R.: Principles of voice production. Journal of the Acoustical Society of America. 104 (3), 1148 (1994)
14. McComb, K.E.: Female choice for high roaring rate in red deer. *(Cervus elaphus)*. Animal Behaviour. 41, 79--88 (1991)
15. Krauss, R.M., Freyberg, R. & Morsella, E.: Inferring speakers' physical attributes from their voices. Journal of Experimental Social Psychology. 38, 618--625 (2002)
16. Jones, B.C. et al.: Integrating cues of social interest and voice pitch in men's preferences for women's voices. Biology Letters. 4, 192--194 (2008)