F0 and Intensity Distributions of Marsec Speakers: Types of Speaker Prosody

Brigitte Zellner Keller

Dept. Clinical Psychology & Rehabilitative Psychiatry,
University of Bern, Switzerland
IMM, Lettres, University of Lausanne, Switzerland
Brigitte.ZellnerKeller@unil.ch

Abstract Most research on F0 has attempted to model the behaviour of an entire linguistic community (e.g of speakers of US or UK English, French, Japanese etc). In this research, we attempt in two analyses to characterize some prosodic aspects of individual differences within the speaker community. For this, the statistical distributions of F0 and intensity parameters were examined. It was found in the first analysis (34 male speakers, nine speech styles) that F0 distributions showed a number of characteristic patterns while intensity distributions did not pattern in any particular fashion. F0 distributions fell into fours patterns, suggesting four styles of F0 whatever the speech style is. This classification was confirmed in our second analysis (11 male speakers, one speech task). These various patterns of F0 distributions are discussed with regard to the speech task and to the speaker's style.

1. Introduction

It is a common assumption that speakers activate speech components in a similar fashion when they perform the same speech task. Differences among speakers are then interpreted in terms of paralinguistic and/or extralinguistic factors.

A number of current prosodic studies investigate the complex aspects of individual variations, in particular in studies related to the "family of emotions", attitudes and sociolinguistic parameters (see e.g., the proceedings of Speech Prosody 2004). These prosodic variations are mainly characterised by one, or more often several, specific patterns which are superimposed on the "neutral" expected prosodic profile. For example, in the case of fear, the mean values for speech rate and F0 are increased [3].

Other prosodic variations are related to the individual psychological profile (for example [4]). Such variations are distinguished by specific interplays between speech rate, intensity, F0 and pauses.

In this paper, statistical distributions of F0 and intensity are investigated. Possible differences at this level do not seem to have been considered in previous studies [1].

2. Methodology

For this study, the Machine-Readable Spoken English Corpus (MARSEC) see www.rdg.ac.uk/AcaDepts/ll/speechlab/marsec.) was used in two analyses: on the one hand, the intonation and the intensity of 34 male speakers recorded in different speech tasks were investigated. This permitted to examine the effect of speech style. On the other hand, the intonation and the intensity of 11 male speakers giving a commentary on the BBC were analysed which permitted to examine the effect of individual variation within a specific speech style.

For the first study, 23'566 values of F0 for 34 male speakers were extracted from the database obtained from MARSEC by Keller [3]. The pitch periods of voiced sounds were determined by the position of the maximum of the autocorrelation function of the sound. All other parameters were kept at default (time step 10 ms, minimum pitch 70 Hz). Intensity values of these voiced sounds were calculated at a time step of 1 ms using the default values set in the Praat software. Average f0 and intensity values were calculated for each sound segment on the basis of these measures.

Nine styles of speech were represented in this database: address, fiction, lecture, market, news, poetry, religion, report and sports.

F0 values were converted into semitones by the formula given in Fant et al. [2]: 12[Ln(Hz/100)/Ln(2). This formula centers the semitone scale at 100 Hz = 0 semitones, and 67 Hz at about -7 semitones and 318 Hz at about 20 semitones. Frequency distributions were computed for each speaker with a stepsize of 0.5 semitones. The obtained frequency per semitone step was then weighted by converting the values into percentages of total observations per speaker.

Intensity values were squared-root transformed to better approximate a normal distribution. Distributions were computed for each speaker with a stepsize of 0.5 -i.e., 0.25 dB. Then the obtained frequency per intensity level was weighted by converting the values into percentages of total observations per speaker.

For the second study, signals were prealigned with an automatic algorithm written by Prof. Eric Keller (IMM, University of Lausanne), and then manually adjusted. Acoustic analyses were performed with the public software Praat. The pitch period of a sound was determined by the position of the maximum response of the autocorrelation function. The voicing threshold was set to 0.05 and the silence threshold was set to 0.15. All other parameters were kept at default values (time step 10 ms, minimum pitch 75 Hz).

F0 values were automatically extracted thanks to a program written by Eric Keller. It runs the Praat's F0 extraction of the input sound. On the basis of the TextGrids, the ouput provides F0 values for each voiced sound.

In both studies, the statistical analyses were performed with DataDesk 6.1, SPSS 11.0 and XLSTAT 5.1.

3. Results

3.1. First Analysis: F0 results

Histograms computed in DataDesk, with the same window length and the same number of intervals show a considerable variation of distributions among speakers.

Table 1. Statistical central values of F0 values for 34 male speakers in nine speech tasks.

Statistics

	N				
	Valid	Missing	Mean	Median	Mode
AMD	397	4787	1.829037	1.488125	7.4845
BP	901	4283	1.818313	1.454095	.0078[a]
BR	310	4874	3.338160	2.979429	-.2718[a]
CF	110	5074	3.080628	2.807319	-2.5968[a]
CL	1104	4080	4.246489	4.562503	-.9145[a]
CP	146	5038	4.697262	4.573999	2.0074[a]
DH	5181	3	1.396632	.911155	-2.6268[a]
DS	479	4705	5.884802	6.006810	6.8069[a]
GB	443	4741	5.152599	4.959489	4.7805
GF	3318	1866	6.350111	6.270877	7.0162
GL	590	4594	6.861170	6.853301	4.3922[a]
JB1	451	4733	7.779444	7.268890	13.9594
JB2	462	4722	2.637703	2.402813	-.7237[a]
JC	659	4525	6.940815	6.355093	8.9241
JH	345	4839	2.884157	2.526611	-4.2560[a]
JM	450	4734	.903066	.498188	-2.0857[a]
JS	400	4784	5.832542	5.839330	5.6475[a]
KG1	345	4839	7.661769	7.144898	15.0907
KG2	503	4681	6.449524	6.289607	5.1909[a]
LM	587	4597	5.329337	5.291426	3.7066[a]
MC	169	5015	4.391394	4.838427	-3.7520[a]
MF	128	5056	3.203771	2.950964	2.4201[a]
MJ	599	4585	3.920015	4.153611	13.4643
ms1	139	5045	1.817037	1.750320	3.6033
ms6	304	4880	3.765106	3.362408	.6912[a]
ms7	268	4916	.442596	.244106	-2.8346[a]
MW	111	5073	5.622323	5.555810	-1.0841[a]
PD	213	4971	3.833997	3.488348	1.9960[a]
PF	82	5102	.476567	.227052	1.2881[a]
PR	450	4734	8.118200	8.303180	6.8878[a]
RF	3300	1884	3.220368	2.730823	-.0904[a]
RSO	423	4761	.228776	.065835	-1.3390[a]
ST	71	5113	7.660508	7.489997	1.9399
VD	128	5056	5.387478	5.843171	-4.1178[a]

a. Multiple modes exist. The smallest value is shown

This variation was unexpected since the literature does not mention any particuliar issue in this domain. However, noticeable differences are observed by just looking at the graphs: some distributions are multimodal and others are not. Some distributions are strongly left-skewed, some others are rather centered. Some distributions are peaked, and others are rather flat.

The three central values computed in SPSS, mean, median and mode, (see Table 1) show various patterns. For example, the three central values for speaker JS are very close. For many other speakers, the lowest modal value differs considerably from the two other central values.

The computation of weighted frequency distributions permits first the graphical superposition of F0 distributions in order to illustrate these variations among speakers (see Figure 1). Secondly, this table was used for the computation in XLSTAT of an agglomerative hierarchical classification (AHC).

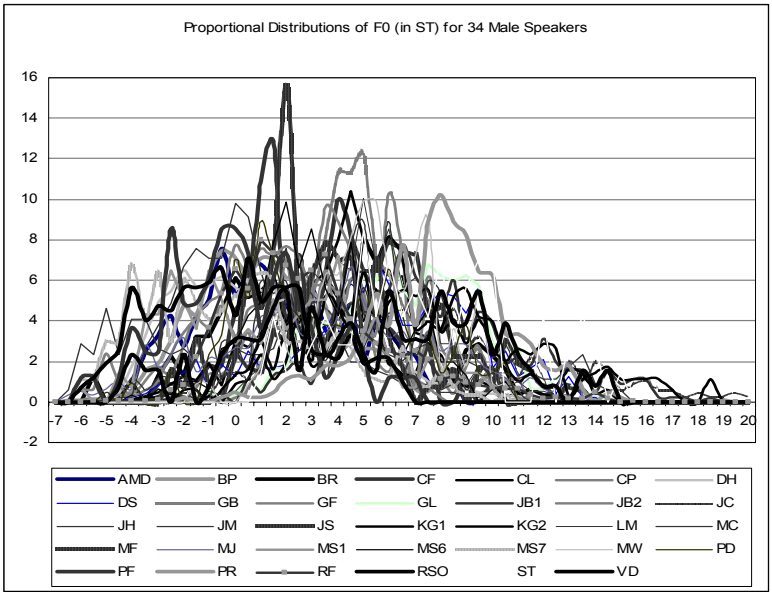

Figure 1. Weighted frequency distributions for 34 male speakers' F0. The x axis represents semi-tones calulated by Fant's equation (2004). This equation sets 100 Hz = 0 st, 70 Hz = 6.17 st, 200 Hz= 12 st. The y axis represents the percentage of samples.

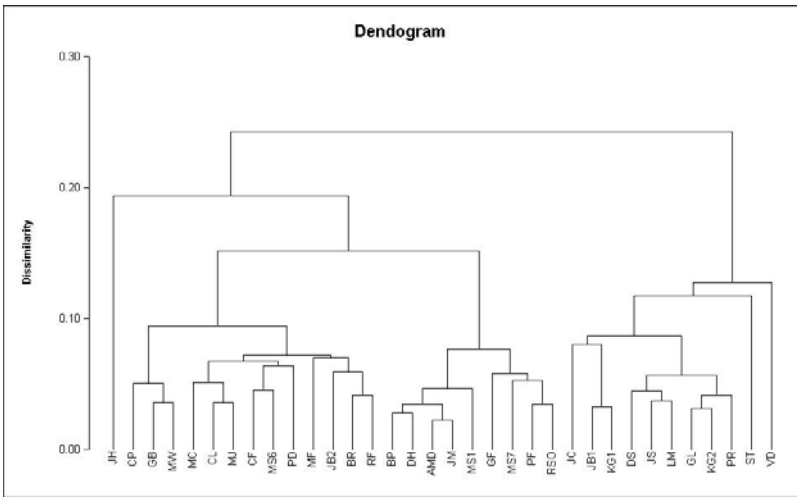

Figure 2. Dendogram of the clustering of F0 distributions for 34 male speakers in nine speech tasks. The clustering suggests 4 groups.

The AHC algorithm gathers the most similar observation pairs, and then progressively aggregates the other observations or observation groups according to their similarity until all observations are in a single group. The settings were as follows: spearman's dissimilarity and complete linkage. In that case, the dissimilarity between objects of A and B is the largest dissimilarity between an object of A and an object of B. Aggregation using complete linkage tends to dilate the data space and to produce compact clusters. The AHC gives a clustering which classify the speakers according to their F0 distributions.It suggests four groups of speakers (Figure 2).

Group 1 is represented by a unique speaker JH in a unique speech style that is fiction. His F0 distribution has a strong left peak and then becomes fairly flat, multimodal and broad. Compared to the mean and median, his lowest modal value is very far-off. The second group (CP, GB, MW, MC, CL, PD, MF, JB2, BR, RF) is characterised by a "normal" height for male speakers with a mean value close to the median value, but both values being distant from the lowest modal value. The skewness is positive. The third group (BP, DH, AMD, JM, MS1, GF, MS7, PF, RSO) is similar to the second group but with a low register of F0. The skewness is positive. In this group, speaker GF is badly classified: his three central values are close and high. In the fourth group (JC, JB1, KG1, DS, JS,LM,GL, KG2, PR, ST, VD), at least two of the three central values are very high and the F0 range is fairly large. The kurtosis is close to 0 or negative. Apart from group 1 with a unique speaker, the three other groups show mixed styles of speech, i.e., these four classes of F0 distributions are not driven by the speech task. When hearing pairs of speech samples according to the leaves of the aggregation, an auditive impression of similarity between speakers emerges.

3.2. First Analysis: Intensity Results

The same procedure for the computation of histograms and weighted distributions were applied to intensity. The superposition of speakers' intensity distributions illustrate a similar pattern irrespective of speaker or speech style (see Figure 3).

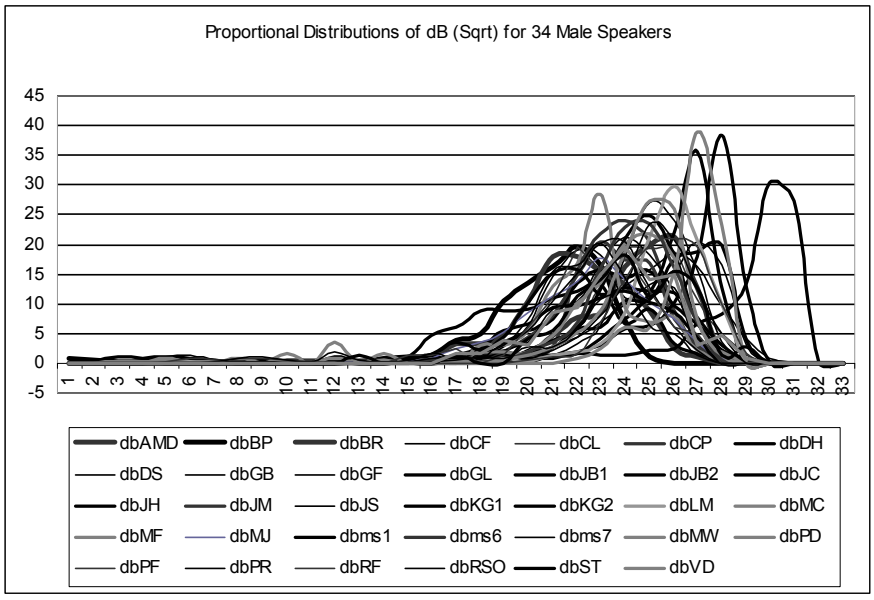

Figure 3. Weighted frequency distributions for 34 male speakers' intensity. The x axis represents squared-root dB values. The y axis represents the percentage of samples.

3.3. Second Analysis: F0 Results

Table 2. Statistical central values of F0 values (in semitones) for 11 male speakers in one speech task.

Statistics

	N						Std. Error of		Std. Error
	Valid	Missing	Mean	Median	Mode	Skewness	Skewness	Kurtosis	of Kurtosis
ST2	340	436	1343696	9328647	3.42468[a]	.616	.132	.903	.264
ST3	440	336	9954147	9240932	5.20582[a]	-.018	.116	.578	.232
ST4	396	380	8925545	2177622	1.14639[a]	.292	.123	-.118	.245
ST5	349	427	0025711	2097615	-.73723	.854	.131	.359	.260
ST6	505	271	8760663	8272467	5.92541[a]	-.400	.109	-.735	.217
ST7	441	335	9693586	6377587	4.22985[a]	.340	.116	.881	.232
ST8	443	333	0122560	7621997	4.28062[a]	.304	.116	-.195	.231
ST9	394	382	0231211	9891787	2.04784[a]	.190	.123	-.212	.245
ST10	517	259	6608246	8251690	1.83327[a]	1.013	.107	1.854	.214
ST11	345	431	8173770	9574606	2.33616[a]	-.352	.131	.390	.262
ST12	402	374	2137634	8857320	4.29279[a]	-.081	.122	.252	.243

a. Multiple modes exist. The smallest value is shown

Although speakers perform the same speech task, the histograms computed in DataDesk, with the same window length and the same number of intervals again show a variation of distributions among speakers. Table 2 shows the central values of the 11 speakers. These variations are visible on the graphical superposition of F0 distributions (see Figure 4). The Kruskall Wallis test confirms that distributions of F0 (in semitones) differ significantly (Chi-square = 1055.890; df=10; p=.000).

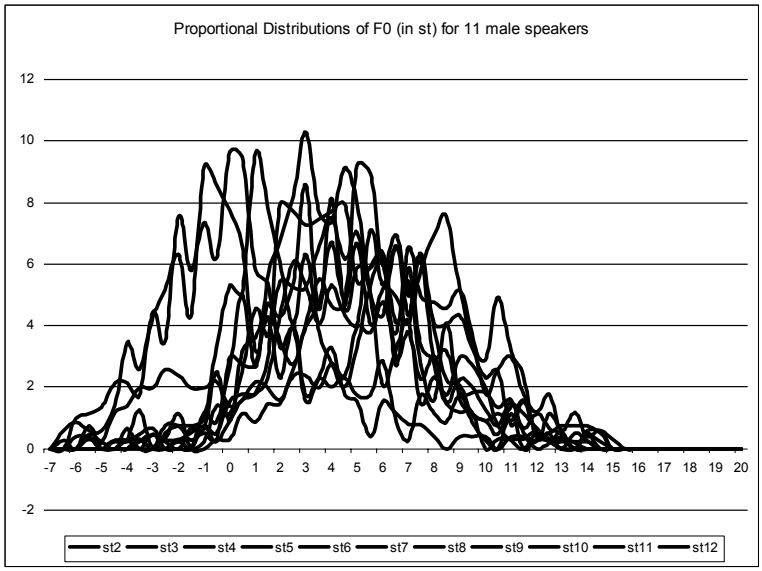

Figure 4. Weighted frequency distributions for 11 male speakers' F0 in a commentary speech task. The x axis represent semi-tones calulated by Fant's equation (2004). The y axis represents the percentage of samples.

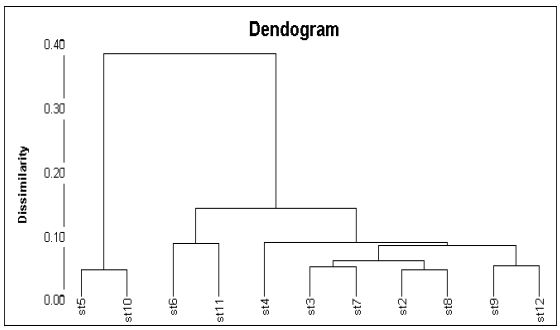

Figure 5. Dendogram of the clustering of F0 distributions for 11 male speakers in a commentary speech task.

The agglomerative hierarchical classification suggests three types of F0 distributions (Figure 5). The first group (ST5, ST10) has the largest positive skewness. Speakers in this group have a low F0 register. Group 2 (ST6, ST11) has the largest negative skewness. The median is higher than the mean and both are far-off the lowest modal value. Speakers in this group have a high register of F0. F0 distributions in the third group (ST4, ST3, ST2, ST8, ST9, ST12) has a skewness close to 0. All the distributions are multimodal. Again, when hearing pairs of speech samples according to the leaves of the aggregation, an auditive impression of similarity between speakers emerges.

3.4. Second Analysis: Intensity Results

Like in the first analysis, the superposition of speakers'intensity distributions illustrate a very similar pattern among the speakers in the commentary task (Figure 6).

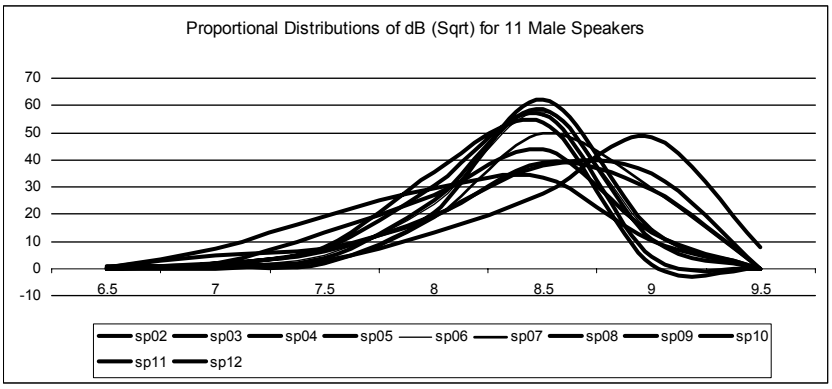

Figure 6. Weighted frequency distributions for 11 male speakers' intensity. The x axis represents squared-root dB values. The y axis represents the percentage of samples.

4. Discussion

Variations in prosody within a linguistic community are triggered by a number of parameters, among others the speech task and the individual style. In this study, it is shown that raw data such as distributions of F0 and dB give interesting information in this area.

Our results show that F0 distributions among the speakers are not similar, whether the speech task is the same or not. The first difference is related to the heigth of the speaker's register and the way speakers use their register. Some speakers present left skewed distributions, meaning that their prefered F0 targets are in the lowest part of

their register. Some other speakers have right-skewed distributions, meaning that they tend to favor F0 targets in the highest part of their register. Beyond the differences in terms of high and low register, we found that some speakers prefer activating their intonation in a multimodal way - with several preferred F0 targets - and some others activate only one preferred F0 target. The preferred F0 target(s) might be close to the two other central values (mean and median) or far-off. These differences sound differently and may characterise styles of intonation which are independant of the speech task.

Conversely, it was found in both analyses that intensity distributions are very similar among the speakers, whether they do perform the same speech task or not. Intensity distributions seem to be less sensitive to the individual characteristics of a speaker - F0 curves are nearly perfectly superposed on each other - and to a certain extent to the speech task. - F0 curves follow the same pattern despite the fact that they are not perfectly superposed on each other.

5. Conclusion

This paper is a contribution to the study of prosodic variations within the same linguistic community. The analyses of F0 and intensity distributions of 34 male speakers in nine speech tasks on the one hand, and the analysis of 11 male speakers in one speech task on the another hand show two interesting facts. Intensity distributions are speaker-independent and to a certain extent are also task-independent. Only one pattern of distribution emerges, whatever the speaker and the speech task are. F0 distributions are task-independent but speaker-dependent. At least fours types of F0 distribution were characterised, suggesting four types of speaker intonation.

Acknowledgments

My special thanks to Eric Keller (Lausanne) for his collaborative support and suggestions. This research is supported by a Swiss OFES grant in support of work performed under COST 277.

References

1. Baken and Orlikoff (2000). Clinical measurement of speech and voice. Singular Publishing Group, San Diego, California.
2. Fant, G., Kruckenberg, A., Gustafson K, & Liljencrants, J. (2002) A new approach to intonation analysis and synthesis of Swedish, Speech Prosody 2002, Aix en Provence.
3. Keller, E. (2003). Voice Characteristics of MARSEC Speakers. VOQUAL: Voice Quality: Functions, Analysis And Synthesis, Geneva - August 27-29, 2003.
4. Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40, 227-256.
5. Zellner Keller, B. (2004). Prosodic Styles and Personality Styles: are the two interrelated? Proceedings of SP2004. (pp.383-386). Nara, Japan.