

RESEARCH REPORT

Effects of speaking style on the shape of fundamental frequency distributions

Pablo ARANTES 

Universidade Federal de São Carlos (UFSCar)

ABSTRACT

The present study has two main goals. The first is to describe the effects of three speaking styles (spontaneous interview, sentence reading and word list reading) on statistical estimators of fundamental frequency (f_0) variability (mean, standard deviation, skewness and kurtosis) in five female and five male speakers of Brazilian Portuguese (BP). Most f_0 contours of word reading are bimodal. Analysis of their time-normalized contours suggests this is caused by the time-compressed realization of fast transitions from low to high or high to low tones aligned with stressed syllables. Considering only unimodal distributions, results show that there are no statistically significant effects in the male data for any of the four variability estimators. Effects show up in female data. Spontaneous style has statistically significant higher mean, SD and skewness than read speech. Findings in the previous literature indicate the reverse pattern, though, for languages other than BP. The second goal of the study is to characterize the statistical properties of f_0 distributions beyond mean and SD. Results confirm previous observations that most f_0 distributions have positive skewness, are left-tailed and have kurtosis values that deviate significantly from the normal because of large deviations from the central or modal value. A distribution fitting procedure tested six distributions. The asymmetric Burr type XII distribution emerges as the one that best fits the data in the corpus. Results show that two of the parameters that determine its shape correlate well with the empirical f_0 distribution values of SD and skewness. Important effects of speaking style on f_0 seen in female speakers can be reproduced by combinations of the Burr distributions' parameters.



OPEN ACCESS

EDITED BY

- Miguel Oliveira Jr. (UFAL)
- Oliver Niebuhr (SDU)

REVIEWED BY

- Ronaldo M. Lima Jr. (UFC)
- Adelaide Silva (UFPR)

DATES

- Received: 30/12/2020
- Accepted: 17/06/2021
- Published: 29/06/2021

HOW TO CITE

Arantes, P. (2021). Effects of speaking style on the shape of fundamental frequency distributions. *Revista da Abralín*, v. 20, n. 1, p. 1-39, 2021.

RESUMO

Este estudo tem dois objetivos principais. O primeiro é descrever o efeito de três estilos de elocução (entrevista espontânea, leitura de frases e leitura de palavras) sobre estimadores da variabilidade de f_0 e também sobre o formato dos histogramas de f_0 de cinco falantes femininos e masculinos do português brasileiro (PB). A maioria dos contornos de leitura de palavras é bimodal. A análise de contornos normalizados temporalmente sugere que a bimodalidade é causada por transições rápidas entre tons altos e baixos. Os resultados das distribuições unimodais mostram que a variável estilo só causa efeitos significativos nos dados das mulheres: o estilo espontâneo apresenta valores maiores de média, desvio-padrão e assimetria em comparação ao estilo leitura de frases. O segundo objetivo do estudo é caracterizar as propriedades estatísticas de f_0 para além da média e do desvio-padrão. Nossos resultados confirmam observações anteriores ao mostrar que distribuições de f_0 no geral têm assimetria positiva e valores de curtose que excedem o que seria esperado para a distribuição normal. Um procedimento para ajuste de distribuições testou seis distribuições de probabilidade teóricas. A distribuição assimétrica Burr tipo XII foi a que teve o melhor ajuste em relação aos dados estudados. Os dois parâmetros que descrevem o formato da distribuição têm boa correlação com os valores de desvio-padrão e assimetria das distribuições empíricas. Os efeitos do estilo de elocução sobre as distribuições de f_0 das mulheres podem ser reproduzidos por combinações dos parâmetros da distribuição Burr.

KEYWORDS

Prosody. Speaking Style. Fundamental Frequency.
Statistical Distributions.

PALAVRAS-CHAVE

Prosódia. Estilos de elocução. Frequência fundamental.
Distribuições estatísticas.

Introduction

In this article we deal with two lines of research that do not cross paths regularly, at least not as we investigate them here. The first is the study of the effects of speaking style on f_0 and the second is the statistical description and modeling of f_0 distributions.

Research on speaking styles revolves around the task of describing or characterizing the many landmarks in a continuum that goes from what can be called spontaneous speech to speech read from previously prepared texts in laboratory conditions. In between these two, it is possible to identify styles that are defined in relation to the content and function of the spoken content, such as news broadcasts, sports narration, theatrical speech and many others. Different dimensions of language and speech are investigated in relation to speaking styles: linguistic stress, speaking rate, vowel reduction, content vs. function words and voice quality, to name a few - see specially Llisterri (1992) for a systematic review of research strategies and results that have accrued around the subject.

Prosodic correlates are consistently researched in regard to the effects of speaking styles; see Llisterri (1992, p. 13-14) for a comprehensive list of suprasegmental acoustic correlates that have already been investigated. Here, we concentrate on the effects of speaking style on overall f_0 variability. The most common research strategy has been to determine how spontaneous and read speech styles affect a number of statistical descriptors of f_0 distribution, mainly f_0 mean and standard deviation. At least six articles systematically review results on this theme (ESKÉNAZI, 1993; HOLLIEN; HOLLIEN; DE JONG, 1997; JESSEN, 2009; KARLSSON *et al.*, 1998; KÜNZEL, 1997; LLISTERRI, 1992). A more recent article (ARANTES; LINHARES, 2017) compares 26 studies mentioned in the review articles listed. In most studies, read speech shows greater f_0 mean than spontaneous speech, although not all studies use the same definition for the latter style. A great number of results show that the two styles do not differ in f_0 standard deviation; studies that show a difference are divided in almost equal numbers between those pointing to spontaneous speech having greater standard deviation and those showing the reverse. Only a few studies report results of inferential statistical tests, and most numerical averages for both mean and standard deviation values presented are close; so, even if differences are statistically significant, the effect sizes are likely to be small. In terms of language diversity, the reviewed studies are dominated by English; and other languages also present in the reviewed papers are Dutch, French, German and Swedish, but in lesser numbers.

Besides reviewing previous work, Arantes and Linhares (2017) present original results from a study that includes seven languages (Brazilian Portuguese, English, Estonian, French, German, Italian and Swedish). The same data collection and analysis procedures were used for all languages. For these data, spontaneous speech was elicited in the form of a semi-directed interview (which the authors classify under the “spontaneous speech” label) and read speech consisted of sentences taken from written transcripts of each participant’s interview and later read by them. Results agree with previous findings¹: all languages considered, read speech has a statistically significant positive difference of 0.83 semitones in relation to spontaneous speech. Breaking the results by language, five of them repeat the overall result, Portuguese reverses the pattern and there is no statistically significant difference in English between the two styles. In regard to standard deviation, no statistically significant difference between spontaneous and read speech is found in the seven languages analyzed separately; collapsing all languages, there is a statistically significant difference of 0.37

1 The authors used ANOVA followed by pairwise tests to investigate differences between groups and speakers.

semitones in favor of spontaneous speech. Although there are no significant differences when languages are considered separately, the values are higher for spontaneous speech in all languages except for Estonian and German. The results are in line with previous findings, at least for the effect of style on mean f_0 value; when it comes to f_0 standard deviation, the results are not more definitive than previously found, although the significant difference in favor of spontaneous speech observed when all language data are collapsed gives an indication that, in larger samples, the small-sized effects observed may yield significance.

The second line of research previously mentioned also has a tradition of its own. Its chief purpose is to establish the main statistical characteristics of f_0 distributions in general. There are theoretical and applied motivations for this line of inquiry. On the theoretical side, one is interested in knowing how to best describe and model f_0 from a statistical point of view and to relate that to the physiology of the voice production mechanism and to linguistic factors that may affect it (FUJISAKI, 1988). On the applied side, there is the development of normalization strategies (JASSEM; KUDELA-DOBROGOWSKA, 1980; MAIDMENT; LECUMBERRI, 1996; ROSE, 1987; 1991) that allow the generation of f_0 contours to abstract away from between-speaker variability and emphasize linguistic-motivated contour movements.

Another main practical reason for the interest in statistical properties of f_0 distributions comes from the potential use of f_0 as an acoustic parameter in forensic speaker comparison. Eriksson (2011, p. 49) mentions that f_0 mean and standard deviation are often suggested as “descriptors of individual differences”. Surveys of common practices in the field (GOLD; FRENCH, 2011; 2019) recognize that f_0 is widely considered by expert practitioners as being useful in speaker comparison tasks. Despite highlighting limitations to the indexical properties of f_0 , Kinoshita and colleagues (2009) suggest that statistical parameters other than f_0 mean and standard deviation may be added as features in voice comparison procedures in order to make f_0 more resistant to within-speaker variability and non-linguistic factors that may affect it and, in turn, make f_0 a more robust factor in forensic speaker comparison. The authors make this claim based on the observation that f_0 histograms generated from audio samples by the same speaker recorded in different occasions “show striking similarities in their shapes” (KINOSHITA; ISHIHARA; ROSE, 2009, p. 93). In the cited article, this observation is corroborated by presenting a selected number of histogram pairs. There is, however, no mention of a systematic study of this behavior to attest its consistency across several speakers.

Research on the general statistical properties of f_0 distributions goes back at least to the 1930s (COWAN, 1936) and review papers (FITCH; HOLBROOK, 1970; HOLLIEN; PAUL, 1969; TRAUNMÜLLER; ERIKSSON, [S.d.]) list numerous studies with similar goals. Most of these studies try to characterize the f_0 distribution by means of f_0 mean and f_0 standard deviation. A few of them try to correlate differences in those estimators with different speaking styles and, in some cases, with speaker sex and physical traits such as body height and weight (HOLLIEN; PAUL, 1969). It is much rarer for statistical descriptors beyond mean and standard deviation to be reported, maybe because it is assumed that f_0 data can be modeled as a normal distribution and as such could be wholly characterized by the distribution mean and standard deviation. Cowan (1936 *apud* HORII, 1975, p. 197), for instance,

states that f_0 contours of “stage speech” are “more or less normally distributed”, although no further technical details are provided about this statement. More recent studies have not corroborated this claim (JASSEM; STEFFEN-BATÓG; CZAJKA, 1973; JASSEM, 1971). The studies reported by the authors applied χ^2 goodness-of-fit tests to f_0 distributions of samples of read speech with about one minute in duration. They concluded that about 90% of them differ significantly from a normal distribution (the existence of one bimodal distribution is also reported). The authors attribute non-normality to deviations in skewness and kurtosis. Measuring skewness by means of Walker’s measure and using a table of critical values taken from a collection of statistical tables, they conclude that, in most cases, the skewness values in their samples differ significantly from what would be expected for a normal distribution. Positive skewness (asymmetry to the left of the central value) is the typical case, with some cases of negative or no skewness. Later studies (HORII, 1975; 1982) have corroborated the finding that f_0 distributions are characterized by positive skewness deviation with reference to the normal, both for read and spontaneous speech, although no information is provided about tests performed to identify statistical significance. Positive deviation seems the most common occurrence, although evidence of typical negative skewness values is also found (ZEMPLIN, 1968 *apud* HORII, 1975). The articles by Jassem and colleagues seem to be the only ones to present systematic data on kurtosis. They measure it using Walker’s measure of kurtosis and use a table of critical values to determine if there is significant deviation from the normal distribution. They find that most distributions have a kurtosis value that is larger than the expected: only two in 20 samples, both from the same speaker, have non-significant values.

Given the fact that no deviation in skewness and kurtosis is the exception rather than the rule, the authors point out that this “may be helpful in the classification of voices for purposes of identification” (JASSEM; STEFFEN-BATÓG; CZAJKA, 1973, p. 219). This observation ties in well with the suggestions made by Kinoshita and colleagues (2009) mentioned earlier in this section that estimators that contribute to more fine-grained detail in f_0 histograms may strengthen its usefulness in speaker comparison tasks.

Evidence of significant deviations in skewness and kurtosis being the typical case in f_0 is mostly based on data collected in English, with the exception of Mikheev (1971 *apud* HORII, 1975), which reports distributions of fundamental period (the reciprocal of f_0) of Russian speech to be positively skewed. It would be desirable that a more diverse language pool be studied to claim that this finding is truly cross-linguistic.

At the start, we stated that one of the goals in this article is to bring together the two lines of research we described. This is done by reanalyzing a speech *corpus* that has already been studied by Arantes and Linhares (2017) with the purpose of finding evidence for the effect of speaking style on f_0 . In that study, the authors found the effects by looking at differences in a set of measures of central tendency and variability of f_0 . Here we advance the analysis by also looking for differences in skewness and kurtosis.

Motivated by previous evidence concerning the presence of many cases of bimodality in f_0 histograms in the *corpus* analyzed by Arantes and Linhares (2017), a methodology to identify and analyze

such cases is presented in this study. We also go beyond the previous literature on the statistical characterization of f_0 and perform a distribution fitting analysis including other distributions other than the normal. Lastly, we try to model differences due to speaking style as changes in values of the distribution that was found to be the best one to model f_0 data.

1. Material and Methods

In this section we present the speech material analyzed (1.1); the phonetic methods used to extract f_0 contours (1.2); and the procedures used to analyze the f_0 distributions and to fit statistical distributions to f_0 data (1.3). In order to conform to open science principles, data files and R scripts used to analyze the data are made available at <https://osf.io/7ms46/>. Praat scripts used for acoustic measurements are available at the author's GitHub profile (<https://github.com/parantes>).

1.1 Speech Material

The speech material analyzed here is the Brazilian Portuguese subset of a database of recordings called “A typology for word stress and speech rhythm based on acoustic and perceptual considerations”,² designed to study lexical stress in a number of languages. The corpus was designed to elicit three different speaking styles: spontaneous speech, read phrases and read words. Spontaneous speech was elicited by way of informal semi-directed interviews with participants conducted by a native speaker that worked for the project. These recordings were then transcribed and used to produce material for the other two speaking styles. Phrases containing suitable target words for lexical stress experiments were selected from the interview recordings from stretches of fluent speech that had no speech errors. Because of this restriction, the number of selected phrases was not uniform for all participants and the total duration of sentence recordings was shorter than interviews. At the next stage, the speakers were called back and asked to read the phrases and words they had produced in their spontaneous speech. This way it was possible to obtain identical linguistic content in all three speaking styles. Speakers were selected in such a way as to minimize variation due to linguistic regional variation and age. All recruited participants spoke a well-defined regional standard. In the case of the Brazilian Portuguese branch of the corpus, participants were five female and five male university students from cities around the Campinas city region, located to the north-west of São Paulo state's capital city. Speaker age variation was the same for all languages within narrow margins. Female speakers in the BP sample ranged from 18 to 32 years of age with a mean of 23; male speakers ranged from 20 to 30 years of age with a mean of 23.

2 See <https://wordstress.ling.su.se/> for more information on the goals of the research, a description of data collection procedures and demographic information of the participants. The project is headed by Anders Eriksson from Stockholm University, in Sweden.

1.2 Phonetic analysis

In the first step of the phonetic analysis, audio samples were segmented into units defined as a function of speaking style. Interview samples were segmented into phonetic utterances, defined by Kendall (2013) as stretches of speech delimited by silent pauses. Each complete sentence sample was segmented into individual sentences and each complete word reading sample was segmented into individual words. The net duration of all units combined, the number of units and the mean duration of units are shown in table 1 for each sample in the corpus. The table also shows frame duration (in milliseconds) and the number of voiced frames per audio sample. The reciprocal of frame duration indicates the rate at which the f_0 extraction algorithm tries to estimate values in the voiced portions of the speech sample (more information on f_0 extraction procedure ahead). Number of frames per sample corresponds to the number of f_0 observations taken from each audio sample.

Speaker	Style	Duration (s)	Units	Mean (s)	Frame period (ms)	Frames	Frames
Female 1	Interview	365	234	1.56	5.36	42,257	42,257
	Sentences	152	46	3.29	5.36	19,047	19,047
	Words	28.3	42	0.67	5.77	3,480	3,480
Female 2	Interview	357	236	1.51	5	3,787	3,787
	Sentences	143	51	2.81	5.77	17,358	17,358
	Words	30.2	45	0.67	6.56	3,672	3,672
Female 3	Interview	921	349	2.64	5.36	49,405	49,405
	Sentences	206	43	4.78	5.77	23,895	23,895
	Words	50.7	45	1.13	5.77	6,150	6,150
Female 4	Interview	649	239	2,71	6.25	48,804	48,804
	Sentences	184	64	2.88	6.25	18,686	18,686
	Words	49	46	1.06	6.82	5,563	5,563
Female 5	Interview	489	272	1.8	5.77	48,162	48,162
	Sentences	148	53	2.8	5.36	17,942	17,942
	Words	38.5	48	0.8	12.5	2,197	2,197
Male 1	Interview	671	279	2.41	10.7	21,282	21,282
	Sentences	279	113	2.47	10.7	13,646	13,646
	Words	38.6	45	0.86	10.7	2,130	2,130
Male 2	Interview	506	307	1.65	8.33	31,828	31,828
	Sentences	178	87	2.05	7.5	15,273	15,273
	Words	25.8	47	0.55	9.38	1,418	1,418
Male 3	Interview	494	277	1.78	8.33	31,996	31,996
	Sentences	202	98	2.06	8.33	13,748	13,748
	Words	29.6	48	0.62	8.33	2,328	2,328
Male 4	Interview	569	312	1.82	10.7	32,432	32,432
	Sentences	207	74	2.8	10.7	13,266	13,266
	Words	49.9	60	0.83	10.7	2,417	2,417
Male 5	Interview	632	235	2.69	8.33	52,777	52,777

	Sentences	212	70	3.02	9.38	16,649	16,649
	Words	37.6	46	0.82	10.7	2,475	2,475

TABLE 1 - Net duration (in seconds) of all units in each audio sample, number of units (see the text for details on what the units are for each of the speaking styles), mean unit duration (in seconds), frame duration (in milliseconds) used by Praat's auto-correlation f_0 extraction algorithm, and number of voiced frames in each audio sample.

Source: the author.

Mean net duration of participant speech per audio sample (in seconds, standard deviation in parentheses) is 565 (166) for interviews, 191 (40.4) for sentence readings and 37.8 (9.42) for word readings. Median number of units (median absolute deviation in parentheses) is 274 (54.1) for interviews, 67 (26.7) for sentence readings and 46 (1.48) for word readings. Mean unit duration (in seconds, SD in parentheses) is 2.06 (0.49) for interviews, 2.9 (0.77) for sentence readings and 0.8 (0.19) for word readings. Frame duration (in milliseconds, SD in parentheses) has median values of 6.22 (1.8) for the female group and 9.52 (1.22) for the male group.

Before the f_0 extraction phase, stretches of audio files that contained the speech of the experimenter, overlap between speaker and experimenter, and non-speech events were silenced to minimize f_0 extraction errors. Extraction of f_0 contours was done with the help of a Praat script (ARANTES, 2019) that implements a heuristic suggested by Hirst (2011) to optimize the values passed to the *floor* and *ceiling* parameters used by Praat's *To Pitch (ac)* autocorrelation-based extraction function (BOERSMA, 1993). The heuristic consists of a two-pass procedure. In the first pass, the *Pitch* object is extracted using default values of 50 and 700 Hz as *floor* and *ceiling* estimates. In the second pass, another *Pitch* object is extracted using optimal values for the two parameters, estimated from the voiced samples in the first *Pitch* object. The optimal *floor* and *ceiling* values are obtained by the expressions $0.7 \cdot q_1$ and $1.5 \cdot q_3$, where q_1 and q_3 are the first and third quartiles of the voiced samples in the first *Pitch* object. The *floor* value defines the rate at which the algorithm tries to estimate values in the voiced portions of the speech sample. The analysis frame period (in seconds) is defined as the result of $0.75 / \text{floor}$. Table 1 shows frame period as a function of speaker and speaking style. Mean values (in seconds, SD in parenthesis) are 6.22 (1.8) for the female group and 9.52 (1.22) for the male group.

Each f_0 contour obtained through the script was then checked individually and remaining f_0 extraction errors were hand-corrected by two analysts trained to perform the task. Most errors commonly detected by this procedure were octave halving or doubling and incorrect voicing detection, usually in fricatives or transient noise in plosive releases. Cases such as incorrect devoicing of frames, which can occur during glottalized or creaky phonation, had to be found by the analyst by comparing the f_0 contour with both the respective oscillogram and spectrogram.

1.3 Statistical analysis

In order to characterize f_0 contours in our sample in terms of their statistical properties, a series of statistical estimators were calculated for each of them.

- Measures of central tendency: arithmetic mean;
- Measure of dispersion: standard deviation;
- Measure of asymmetry: Pearson's moment coefficient of skewness;
- Measure of kurtosis: Pearson's kurtosis.

All statistical estimators were calculated for values in Hertz and log-Hertz (hereafter referred to as logHz)³. Log-transformation was applied to f_0 data for two separate reasons. From a purely statistical point of view, it is well known that it can reduce the skewness of many types of data and help them become more normal-like (LIMPERT; STAHEL; ABBT, 2001). Besides that, log-transformation can be justified on physiological and linguistic grounds, as pointed by Fujisaki and colleagues (FUJISAKI, 1988; FUJISAKI; HIROSE, 1984; FUJISAKI, HIROYA; OHNO; GU, 2004) and explained below.

Fujisaki advocates for log-transforming f_0 values based on the observation that the relation between f_0 and vocal folds elongation can be described linearly if f_0 is measured on the log-scale. From a linguistic point of view, Fujisaki puts forward the idea that the surface f_0 contour of an utterance can be conceived as the result of the superposition of two separate components – local and relatively fast rise-fall movements, and a global and relatively slow declining baseline. The first of the two components roughly corresponds to pitch accents connected to prosodic words and the second corresponds to larger units, such as clauses, phrases, or sentences. If the f_0 contour is expressed on a log scale, it is possible to treat the superposition of both components mathematically as an addition operation, simplifying the formulation of a model of the interaction between the accent and phrase components, such as Fujisaki's namesake model. The fact that there are two independent laryngeal mechanisms to elongate the vocal folds – rotation on the thyroid cartilage around the cricothyroid joint, associated with the accent component; and forward translation of thyroid cartilage, associated with the phrase component – further justifies treating their combined action as additive in nature.

D'Agostino test of skewness (D'AGOSTINO, 1970) and Anscombe-Glynn test of kurtosis (ANSCOMBE; GLYNN, 1983), available through the *moments* R package (KOMSTA; NOVOMESTKY, 2015), were performed to determine if values for each f_0 sample in the corpus differ significantly from what would be expected for the Gaussian or normal distribution. The normal distribution is

³ The natural (base e) logarithm was used in the transformation.

symmetric, meaning that data are about equally distributed around the central value, with skewness equals to zero. A sample with negative skewness is said to be left-skewed or left-tailed, meaning that the mass of the distribution is concentrated to the right of the center; whereas one with positive skewness is said to be right-skewed or right-tailed, meaning that the mass of the distribution is concentrated to the left of the center. The normal distribution has a kurtosis value of 3. Kurtosis is traditionally presented as a peak feature, describing the shape of the center of a distribution: either more flat-topped (platykurtic) or more pointed (leptokurtic) relative to the normal distribution. A number of authors (see Westfall (2014) and references therein) argue that the correct interpretation of kurtosis is to consider it a measure of the propensity of a distribution to be heavy-tailed, that is, to generate extreme values, or values far from the central tendency.

Given prior experience with the data set we analyze in the present study (ARANTES; ERIKSSON, 2019; ARANTES; LINHARES, 2017), we knew that there were cases of distributions that showed evidence of bimodality, i.e., histograms with more than one modal value. In order to identify these cases in a more objective way, Hartigan's *dip* test for unimodality (HARTIGAN; HARTIGAN, 1985) was performed for all sample f_0 distributions (both in Hz and logHz) in the corpus – the test yields significance when distributions are non-unimodal. Since in bimodal distributions the mean value of the overall distribution is likely not to be representative of any of the distributions that can be assumed to be mixed together, bimodal cases were identified and treated separately according to the procedures described in section 1.3.2. Unimodal samples were subject to a distribution fitting procedure described in the following section.

An α level of 5% was adopted for all statistical analysis conducted in the present study and they were all carried out using the R statistical computing environment (R CORE TEAM, 2020).

1.3.1 Distribution fitting

With the aim of establishing which univariate parametric distributions best describe the unimodal f_0 samples in our corpus and whether speaking styles have an effect on this, we used the R package called *fitdistrplus* (DELIGNETTE-MULLER; DUTANG, 2015) to fit six theoretical probability distributions, listed below. Fitting was carried out on the log-transformed values for the reasons outlined in section 1.3.

Symmetric distributions:

- Normal or Gaussian;
- Logistic.

Asymmetric or skewed distributions:

- Burr type XII, also known as Singh–Maddala or generalized log-logistic distribution;
- Gumbel or Generalized Extreme Value distribution Type-I;
- Gamma;
- Weibull.

As will be reported in the Results section, most f_0 distributions are right-skewed, and for this reason we tested more asymmetric than symmetric distributions. Of the tested distributions, Burr type XII is a heavy-tail distribution, meaning that it has “a larger probability of getting very large values” (WOLFRAM RESEARCH, 2020). The others are considered thin-tail distributions, meaning that “the PDF [probability density function] decreases exponentially for large values” (WOLFRAM RESEARCH, 2020) of the variable. Weibull can have both kinds of tails depending on the values of its parameters. Distributions included in the fitting analysis were chosen consulting a compendium of probability distributions (MCLAUGHLIN, 2016) and considering their availability within R, either as part of the *base* library or add-on libraries. Some of the skewed distributions were available through a package called *actuar* (DUTANG; GOULET; PIGEON, 2008), specialized in actuarial science, as heavily skewed distributions are common for that kind of application.

All six distributions listed above were tested against each f_0 sample in our corpus, and the best fit was the distribution that yielded the smallest Anderson-Darling goodness-of-fit statistic (abbreviated A^2).

1.3.2 Bimodality analysis

Following the identification of bimodality with the dip test, we did a visual inspection of the histograms that yielded significant p -values to check if the non-unimodality was clearly visible. As it will be reported in detail in the Results section, there were a small number of cases that we considered to be false positives, that is, the test turned out significant although visual inspection of the histogram did not show more than one obvious mode. Because of that, we decided to visually inspect all histograms to check for the occurrence of false negatives, that is, histograms that showed more than one mode but were not registered by the statistical test. There were a small number of those as well, also reported in section 2.3. Considering the sample of 30 f_0 distributions, two cases of false positives and three of false negatives were found. In all cases, non-unimodal distributions were bimodal, i.e., when more than one mode was detected, only two were present in the histogram. Distributions identified as non-unimodal were further analyzed in two ways. First by submitting them to a distribution mixture analysis and then to time-normalization.

We now discuss the procedure to identify bimodality. The smoothed histogram shown in figure 1 shows evidence of two distinct modes, almost one octave apart, suggesting that there is not just one source of systematic f_0 variability in play, but possibly two. The distribution shown in figure 1 could be, then, the result of the superposition of two more basic (unimodal) distributions. In situations like this, a technique called distribution mixture analysis (BISHOP, 2006) can be used to estimate the parameters of the putative underlying distributions. One of the simplest and most tractable cases is when component distributions assumed to underlie a given distribution are normal or Gaussian and this is referred to as a Gaussian Mixture Model or GMM. In order to carry out a GMM analysis, it is necessary to give estimates for the mean, the standard deviation and a parameter called λ , which corresponds to the relative weight of each normal distribution in the joint distribution. We used the *mixtools* R package (BENAGLIA *et al.*, 2009) to carry out the analysis. Figure 1 illustrates the procedure we used to obtain these values. First, we generated smooth histograms of the distributions identified as non-unimodal with the help of the *kernSmooth* R package (WAND, 2015). Peaks in the histogram contour (points identified by the solid vertical lines in figure 1) were used to estimate the location of the modal values of the component distributions, used as a proxy for their mean values. Then, we identified the local minimum between the two peaks and considered that to be a rough estimation of the boundary between the two component distributions. Standard deviation estimates for the two distributions were calculated by taking the standard deviation of values ranging from the lowest to the boundary values and then the standard deviation of values from the boundary to the maximum value in the sample. In the example shown in figure 1, these two intervals correspond to values ranging from the minimum to 5.03 logHz for the component centered around 4.8 logHz, and from 5.03 logHz to the maximum value for the component centered around 5.4 logHz. Given the rough estimates provided by the user, the function *normalmixEM* from the *mixtools* package returns estimates for the three parameters (mean, SD and λ) for the two components, based on an expectation maximization (EM) procedure. For the example in figure 1, the values returned by the function are 4.74 logHz and 5.38 logHz for the mean, 0.21 logHz and 0.13 logHz for SD, and 0.21 and 0.79 for λ . All mixture analyses were run on the log-transformed distributions.

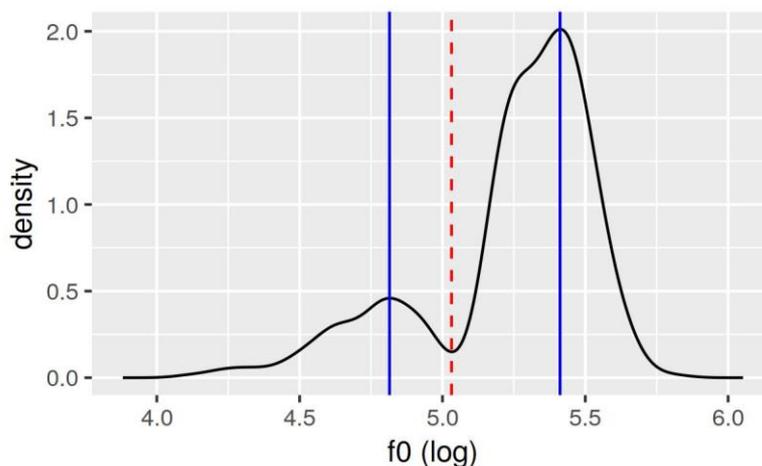


FIGURE 1 - Smoothed histogram of a word reading distribution (female speaker 5). Modes, indicated by solid vertical blue lines, are located at 4.81 and 5.41 (123 and 224 Hz, respectively). The dip point at 5.03 (153 Hz), indicated by the dashed red vertical line, was identified as a boundary between the two component distributions centered at the modes.

Source: the author.

Word reading was the style that yielded the most cases of bimodal distributions (more details in the Results section). In order to better understand what might be driving this behavior, we submitted the word reading samples to a time-normalization analysis. Time-normalization allows the comparison on the same temporal scale of the f_0 contour applied to different words with the same size and repetitions of the same word within word reading samples across different speakers. See Arantes (2015) for a more detailed explanation of the time-normalization technique. The interval over which time-normalization was done was the duration of each word in word reading samples. Start and end points for each word were marked in *TextGrid* files within Praat. The stress pattern of each word was labeled using the symbols “W” for pre- and post-stressed syllables and “S” for stressed syllables, such that a word like “trabalho” (*work*, stressed syllable in bold face) was labeled “WSW”. Time-normalization was done with the help of a Praat script (ARANTES, 2018). A fixed number of 30 equally-spaced f_0 samples were taken from each marked interval. The f_0 contour was smoothed (the *Bandwidth* parameter of Praat’s *Smooth* function was set to 4 Hz) prior to sample collection. Plots of time-normalized contours as a function of word stress patterns were generated and visually analyzed. The main features we looked for in the plots were correlations between downward and upward movements and their alignment with the stressed syllable or the initial or final word boundary.

2. Results

This section presents descriptive statistics and distribution fitting analysis for unimodal and bimodal f_0 distributions. As explained in section 1.3, we decided to report unimodal and bimodal cases separately, based on the reasoning that values of statistical estimators taken from the overall sample in

bimodal distributions are likely not to be representative of any of the distributions that are mixed together in the ensemble.

By applying the procedure described in section 1.3 to identify bimodal distributions, it was possible to determine that 21 out of the 30 distributions in the study sample are unimodal and are going to be analyzed in more detail in sections 2.1 and 2.2. The other nine are bimodal and a complete analysis of those distributions is presented in section 2.3.

To provide the reader with an overview of the variability encountered in our corpus, figure 2 presents smoothed histograms of all the 30 f_0 distributions that make this study's corpus. The smoothed histograms in the figure were generated using the procedure described in 2.3.2 that uses kernel density estimates to obtain a continuous density curve out of raw histograms.

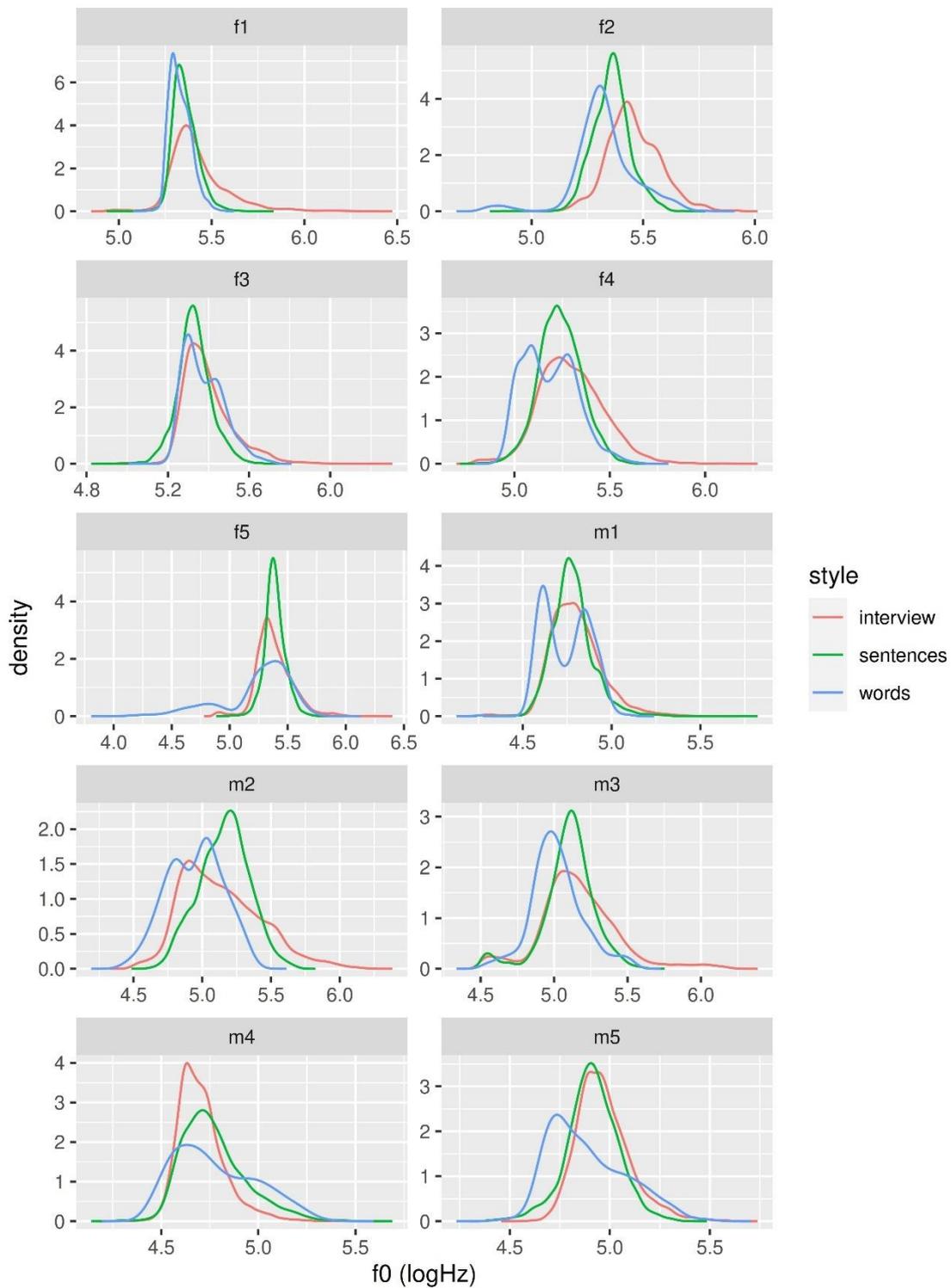


FIGURE 2 - Smoothed histograms of all 30 f_0 distributions in the study corpus presented as a function of speakers and speaking styles. Female speakers are identified by the labels $f1$ to $f5$ and male speakers by labels $m1$ to $m5$. Different speaking styles have different colors. In all cases, f_0 values are in log-scale.

Source: the author.

2.1 Unimodal distributions statistics

This section presents descriptive and inferential statistics regarding unimodal f_0 distributions. The breakdown of the 21 unimodal distributions by speaking style (and speaker sex) is: interview (5 female, 4 male), sentence reading (5 female, 4 male), and word reading (1 female, 2 male). Values of the four statistical estimators mentioned in section 1.3 for each of the 21 unimodal distributions are shown in Figures 3 (Hertz scale) and 4 (logHz scale).

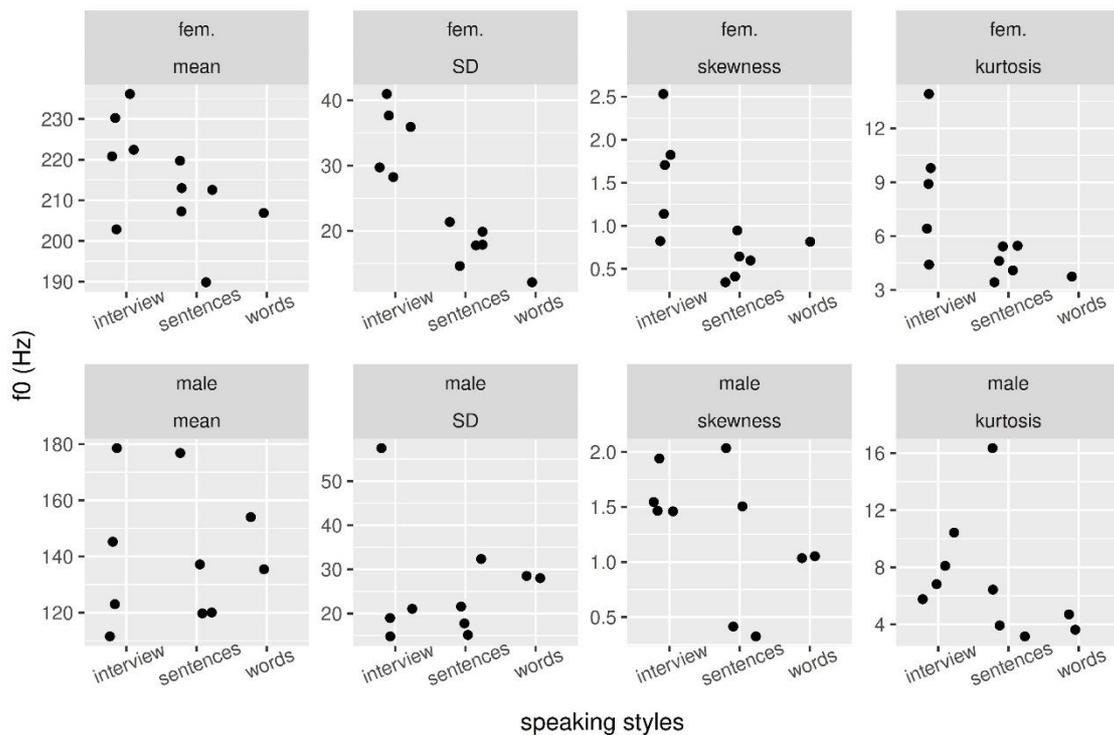


FIGURE 3 - Values of four statistical estimators of f_0 taken from unimodal distributions (in Hz scale) as a function of speaking style and speaker sex (female on top, male at the bottom).

Source: the author.

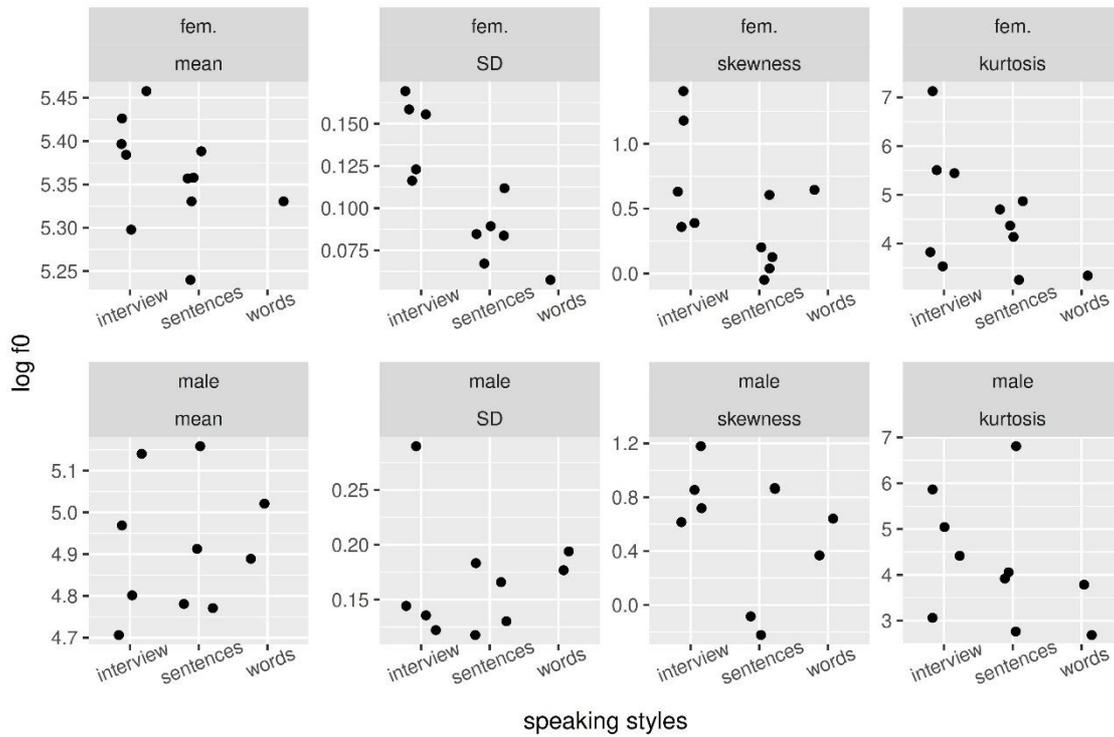


FIGURE 4 - Values of four statistical estimators of f_0 taken from unimodal distributions (in log scale) as a function of speaking style and speaker sex (female on top, male at the bottom).

Source: the author.

Summary statistics for the four estimators will be presented below, in both scales, followed by statistical comparisons. Since the number of cases in the word reading style is low for both female and male speakers, paired t-tests were used to test for a difference between the interview and sentence reading styles.

	Interview	Sentences	Words
Female	223 (12.6)	208 (11.3)	207
Male	140 (29.5)	138 (26.9)	145 (13.1)

TABLE 2 - Mean f_0 value (SD in parentheses) in Hertz scale as a function of speaker sex and speaking styles.

Source: the author.

	Interview	Sentences	Words
Female	5.39 (0.06)	5.33 (0.06)	5.33
Male	4.9 (0.12)	4.91 (0.01)	4.95 (0.01)

TABLE 3 - Mean f_0 value (SD in parentheses) in log scale as a function of speaker sex and speaking styles.

Source: the author.

For female speakers, the interview style has a greater mean average f_0 in both scales; effect sizes are large [Hz: $t(4) = 3.84, p = 0.018, d = 1.72$; logHz: $t(4) = 3.4, p = 0.027, d = 1.52$]. No significant difference is found for male speakers; effect sizes are negligible [Hz: $t(3) = 0.33 \text{ ns}, d = 0.17$; logHz: $t(3) = -0.55 \text{ ns}, d = -0.028$].

	Interview	Sentences	Words
Female	34.5 (5.38)	18.3 (2.54)	12.1
Male	28.1 (19.8)	21.7 (7.57)	28.3 (0.36)

TABLE 4 - Mean f_0 standard deviation value (SD in parentheses) in Hertz scale as a function of speaker sex and speaking styles. Source: the author.

	Interview	Sentences	Words
Female	0.14 (0.023)	0.087 (0.016)	0.058
Male	0.17 (0.078)	0.15 (0.03)	0.18 (0.012)

TABLE 5 - Mean f_0 standard deviation value (SD in parentheses) in log scale as a function of speaker sex and speaking styles. Source: the author.

For female speakers, the interview style has greater f_0 standard deviation values in both scales; effect sizes are large [Hz: $t(4) = 5.7, p < 0.005, d = 2.56$; logHz: $t(4) = 5.6, p < 0.005, d = 2.51$]. No significant difference is found for male speakers; effect sizes are small [Hz: $t(3) = 0.95 \text{ ns}, d = 0.48$; logHz: $t(3) = 0.76 \text{ ns}, d = 0.38$]. Regarding the effect of speaker sex, different patterns emerge: male speakers show no difference between interview and sentence reading in mean and standard deviation; for female speakers, there are significant differences.

Similar results regarding mean and standard deviation were reported in Arantes and Nascimento (2017) for the same BP speech material, although there the authors did not separate unimodal and bimodal distributions. Here the analysis is taken further by also looking at skewness and kurtosis. Regarding skewness, the test used to detect it turned significant results for all 21 unimodal distributions, regardless of measurement scale. Skewness is positive for all distributions measured in Hz scale and ranges from 0.32 to 2.53. When measured in the log scale, 18 distributions have positive values and 3 negative, ranging from -0.22 to 1.40. These findings point to a strong tendency towards positive skewness, confirming previous observations found in the literature that f_0 distributions tend to be right-skewed. Tables 6 and 7 present mean skewness values as a function of speaking style and speaker sex in the Hertz and log scales, respectively. As expected, log-transformation has the effect of bringing skewness values down, sometimes close to zero, although, even in this case, statistical tests for skewness turned out to be significant in all cases.

	Interview	Sentences	Words
Female	1.61 (0.66)	0.59 (0.23)	0.82
Male	1.6 (0.23)	1.07 (0.84)	1.04 (0.012)

TABLE 6 - Mean f_0 distribution skewness (SD in parentheses) in Hertz scale as a function of speaker sex and speaking styles. Source: the author.

	Interview	Sentences	Words
Female	0.79 (0.48)	0.18 (0.25)	0.64
Male	0.84 (0.25)	0.36 (0.59)	0.50 (0.19)

TABLE 7 - Mean f_0 distribution skewness (SD in parentheses) in log scale as a function of speaker sex and speaking styles. Source: the author.

For female speakers, the interview style has greater skewness values in both scales [Hz: $t(4) = 5.27$, $p = 0.006$, $d = 2.36$; logHz: $t(4) = 5.01$, $p = 0.007$, $d = 2.24$]. For male speakers the pattern is the same, although the difference is not statistically significant [Hz: $t(3) = 1.42$, ns , $d = 0.71$; logHz: $t(3) = 1.86$, ns , $d = 0.93$]. Effect sizes are large for female speakers and moderate and large for male speakers.

Regarding kurtosis, the statistical test to detect excess kurtosis yielded significant values for all distributions, regardless of measurement scale: all values are greater than 3 for f_0 values in Hz (range: 3.15-16.3); for values in logHz, 19 out of 21 values are greater than 3 (range: 2.69-7.13). As with skewness, conversion to log-scale lowers excess kurtosis, although not enough to bring it to the level of the normal distribution. The fact that most samples have kurtosis values that deviate from what would be expected were they normally distributed can thus be seen as evidence that f_0 distributions are best characterized as heavy-tail distributions. The results seen for skewness corroborate this and point to the fact that the heaviness tends to concentrate on the left tails, or towards higher f_0 values.

Tables 8 and 9 present mean kurtosis values as a function of speaking style and speaker sex in the Hertz and log scale respectively. For female speakers, the interview style has a higher kurtosis level than sentence reading; effect sizes are large and moderate for Hz and logHz scales respectively; the difference is significant in Hz scale but no on logHz. [Hz: $t(4) = 3.09$, $p = 0.037$, $d = 1.38$; logHz: $t(4) = 1.54$, ns , $d = 0.69$]. For male speakers, there is no significant effect of speaking style on kurtosis levels; effect sizes are negligible [Hz: $t(3) = 0.11$, ns , $d = 0.055$; logHz: $t(3) = 0.28$, ns , $d = 0.14$].

	Interview	Sentences	Words
Female	8.69 (3.61)	4.60 (0.87)	3.75
Male	7.78 (2.01)	7.46 (6.09)	4.16 (0.76)

TABLE 8 - Mean f_0 distribution kurtosis (SD in parentheses) in Hertz scale as a function of speaker sex and speaking styles. Source: the author.

	Interview	Sentences	Words
Female	5.09 (1.46)	4.26 (0.63)	3.34
Male	4.60 (1.18)	4.39 (1.72)	3.24 (0.78)

TABLE 9 - Mean f_0 distribution kurtosis (SD in parentheses) in log scale as a function of speaker sex and speaking styles.
Source: the author.

Summing up the results of style effect on distribution parameters, the takeaway is that, for female speakers, sentence reading has a lower mean level, less variability (SD), less asymmetry (skewness) and less extreme values (kurtosis) compared to the interview style. This is true regardless of scale. For male speakers, there is no significant effect of style on any of the parameters, regardless of scale. For both sexes, we do not include word reading in the statistical comparisons, because of sample size for this style: for female speakers, four out of five word reading distributions are bimodal and for male speakers, three out of five.

2.2 Unimodal distribution fitting

In this section we report the results of the distribution fitting procedure described in section 1.3.1. Overall, the theoretical distributions that best describe the f_0 samples are: Burr 76% (N = 16), Gumbel 14% (N = 3), Gamma 5% (N = 1) and Logistic 5% (N = 1). Breaking the results as a function of style, we have:

- *Interview*: Burr 89% (N = 8), Gumbel 11% (N = 1);
- *Sentence reading*: Burr 78% (N = 7), Gamma 11% (N = 1), Logistic 11% (N = 1);
- *Word list reading*: Gumbel 67% (N = 2), Burr 33% (N = 1).

First, we can note that asymmetric or skewed distributions are those that best fit f_0 data with except for one sentence reading sample, for which the Logistic distribution is best. This could be expected, given that the results presented in section 2.1 show that, regardless of speaking style, most distributions have significant positive skewness and significant excess kurtosis.

Numerically, Burr type XII distribution is the best fit for interview and sentence reading styles. The Gumbel distribution, which is similar to Burr's in the sense of being right-skewed, gets first place in the unimodal cases for two of the word reading style samples. As mentioned in section 1.3.1, the best fit was determined by ranking the candidates by their goodness-of-fit values as estimated by the Anderson-Darling statistic. Since Burr type XII is so prevalent, we decided to investigate by how much it lost the first place in the goodness-of-fit procedure, especially in cases where Gumbel was the best candidate, as it is similar to Burr type XII.

In all cases where Gumbel is considered the best fit (f1 words, m2 interview and m5 words), Burr is a close second: the A^2 statistic for Burr being from 1.2 to 2.5 times bigger than Gumbel's. In the one case where the symmetric Logistic distribution is the best fit (m5 sentences), Burr is second with an A^2 value that is 1.3 times greater than the one for the best fit. In the case where Gamma is the best fit (f4 sentences), Normal is the second best (A^2 1.3 times bigger), followed by Burr (A^2 1.9 times bigger). Considering all five samples, the candidates ranked third or lower have an average A^2 value that is 17.8 times bigger than the best fit, with A^2 values ranging from 2 to 107 larger than the best fit. The results suggest that even when Burr type XII is not the best fit, it is a close second (or third in one single case), even when the best fit is a symmetric distribution. Thus, the data derived from our corpus of samples seem to support the proposal that Burr type XII is a reasonable candidate for the purpose of modeling most unimodal f_0 distributions.

2.2.1 The Burr type XII distribution

Burr type XII is a distribution that is defined by three parameters: shape 1, shape 2, and rate, all being positive real numbers. Figure 5 shows how the shape of a Burr type XII distribution changes when we keep two parameters fixed and change the values of the third (values in the figure were chosen in a range plausible for f_0 data). As we can see in panel (a), lowering the value of shape 1 makes the distribution more right-skewed and heavy-tailed; higher values make the distribution more symmetric. Panel (b) shows that shape 2 controls the height of modal density and tail weight; increasing the value heightens the modal density and thins the distribution's tails. Panel (c) shows that the rate parameter (or location, the reciprocal of rate) locates the distribution's center in the x -axis; the distribution is translated to the right as rate decreases (or location increases); there is also a small degree of widening and modal height decreasing as rate value decreases (or location increases).

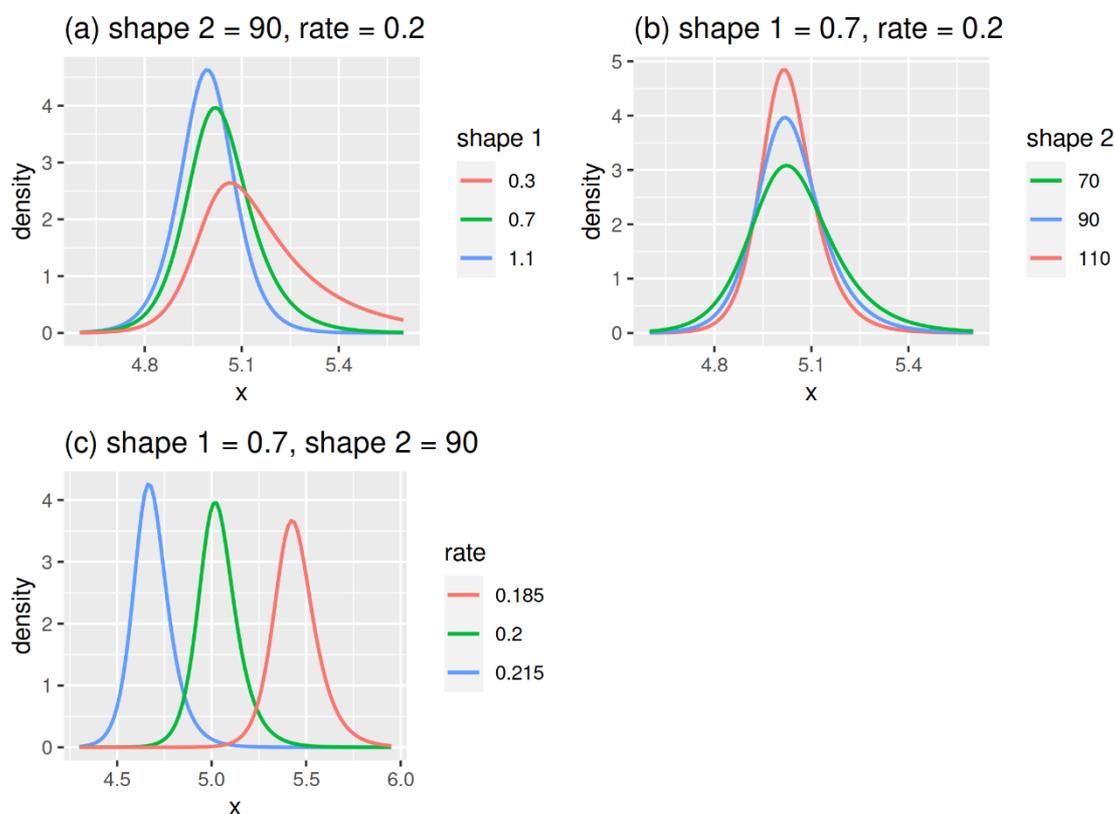


FIGURE 5 - Effect of changes in the three parameters that define the shape of a Burr type XII distribution.
Source: the author.

We used the *fitdist* function from the R package *fitdistrplus* (DELIGNETTE-MULLER; DUTANG, 2015) to estimate the values of the three parameters of the Burr type XII distribution from the 21 unimodal f_0 samples in our corpus. Then, for each of the 21 unimodal f_0 contours we correlated the estimated values for the three parameters to the statistical descriptors (mean, SD, skewness and kurtosis) reviewed in section 2.1. Applying the Pearson product-moment correlation test, we found three significant correlations: skewness and shape 1: -0.82 [$t(19) = 6.33$, $p < 0.001$]; standard deviation and shape 2: -0.64 [$t(19) = 3.66$, $p = 0.002$]; mean and rate: -0.97 [$t(19) = -16.8$, $p < 0.001$]. These results corroborate what can be observed in Figure 5: panel (a) shows that lowering shape 1 value results in increased asymmetry in the overall distribution shape and a heavier right tail; also, panel (b) in the same figure shows that a lowering in shape 2 value results in a widening of the distribution's central section without a noticeable change in the degree of symmetry; finally, and more obvious, panel (c) shows that an increase in mean value causes a decrease in the rate parameter (or increase in location). The results show that the Burr type XII distribution does a good job in capturing the influence sample standard deviation and skewness have in determining the overall shape of f_0 distributions in a way that other distributions do not. The role of the sample mean in locating the distribution on the x -axis can reasonably be modeled by other distributions as well. There is a significant positive

correlation of 0.62 [$t(19) = 3.44, p = 0.003$] between skewness and kurtosis, although that does not translate into a significant correlation between kurtosis and shape 1 [$-0.33; t(19) = -1.53, p = 0.14$].

2.2.2 The effect of speaking style

In order to uncover a possible effect of speaking style on the shape of the f_0 data as described by a Burr type XII distribution, we calculated the mean values of the three distribution parameters as a function of the three speaking styles. Figure 6 shows the probability density functions of the three speaking styles for female (panel a) and male speakers (panel b), generated from mean parameter values.

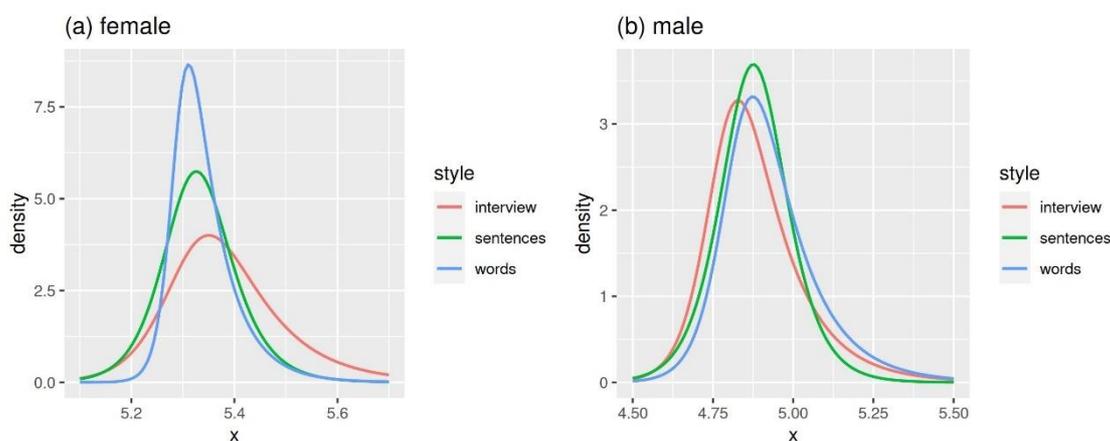


FIGURE 6 - Effect of speaker sex and speaking style on the shape of Burr type XII distribution.
Source: the author.

There is a noticeable effect of style on the distribution shape for female speakers. Modal value shifts towards higher values in the following order: word reading < sentence reading < interview. Values are 5.3, 5.325 and 5.35 or approximately 204, 205 and 210 Hz, converting the values back to the Hertz scale. The most visible change is seen in the wideness of the distribution, following the same direction of the shift in modal value. All three distributions are right-skewed. Sentence reading and interview have much heavier tails on both sides. These observations corroborate the significant differences in standard deviation and skewness presented in section 2.1 for female speakers. No relevant difference in distribution shape motivated by speaking style is seen for male speakers, also corroborating the lack of statistical significance in standard deviation and skewness also presented in section 2.1.

It seems unlikely that the lack of effect of speaking style on male speakers could be explained on physiological grounds. One possible explanation for the lack of effect in male data could be that it is due to the relatively small sample size in each group. Since the effect sizes reported in female data were larger than the ones seen in the male data, significant results were likelier to be found in

that group even in a small sample. Future work with larger sample sizes may help understand if the sex difference is reliable or an artifact caused by an insufficient sample size.

The pattern seen in the female data suggests that the interview style generates more asymmetric distributions, with larger f_0 excursions and overall greater variability than the other two styles. This is compatible with a view that spontaneous or semi-spontaneous speaking styles are livelier, as speakers tend to be more engaged and involved in what they are saying than in read speech, that has a lesser degree of f_0 modulation and therefore could be perceived as having a relatively more level intonation pattern. Arantes and Eriksson (2019) present data that may be interpreted as indication that this is not a fixed pattern, but something that can, at least in part, be language-specific or culture-specific. In their study, the authors studied the multilingual corpus that includes the BP data analyzed here and developed a methodology to measure similarity between f_0 contours. Applying the methodology to investigate inter-speaker variation in contour similarity, the authors report that for a group of languages consisting of English, French, Italian, Brazilian Portuguese and Swedish, spontaneous interview was the style which yielded the highest levels of variation. A closer inspection of the pairs of contours with the largest dissimilarity index values revealed that most of them included rare cases of interview samples with a more level f_0 profile. The few atypical interview contours, then, generated most of the inter-speaker variation. For Estonian and German, on the other hand, sentence reading was the style yielding more inter-speaker variation. The crucial difference, though, was that most sentence reading contours showed a good deal of f_0 modulation and extended ranges and the pairs with the largest dissimilarity values included, in most cases, the few atypical cases of sentence reading contours with less modulation and narrower ranges. One conclusion that could be drawn from the results is that speaking styles differ in terms of f_0 variation, but language communities may vary in terms of which style has a more lively profile, with larger excursions and overall more modulation, and which ones show a more level profile.

Going back to the results in the present study, it seems that male speakers adhere to the pattern show in Arantes and Eriksson (2019) to a lesser extent than the female speakers. It is possible to see that, for the sentence reading style, males and females present similar SD and skewness values. For the interview style, on the other hand, female speakers show an increase in SD and skewness that is not present to the same degree in the male sample (cf. tables 5 and 7). This could result in the impression that females sound livelier or more expressive than males in the interview style, although this hypothesis should be confirmed by future experiments. If it is shown that female spontaneous contours are reliably perceived as livelier than male's contours, that may be a sociophonetic effect that results from the different expectations regarding the behavior of females and males in social interactions.

The coarse-grained dynamics revealed by the distinct statistical characteristics of spontaneous and read speech in the female group is illustrated in figure 7. It shows stretches of two contours taken from the central 20 seconds of speaker f_1 interview sample (green dots) and speaker f_2 sentence reading sample (blue dots). The interview distribution has skewness and kurtosis values of 1.4 and 7.1, respectively, while the sentence distribution has skewness and kurtosis values of -0.04 and 4.4, respectively. As expected, the phonetic effects of these differences are: much wider f_0 excursions

and an expanded f_0 range in the case of the highly positive-skewed green contour and the lack of extreme f_0 excursions and a more symmetrical variation around the median f_0 value in the blue contour. Follow-up studies should look in more detail how the differences play out in the small scale of individual sentences.

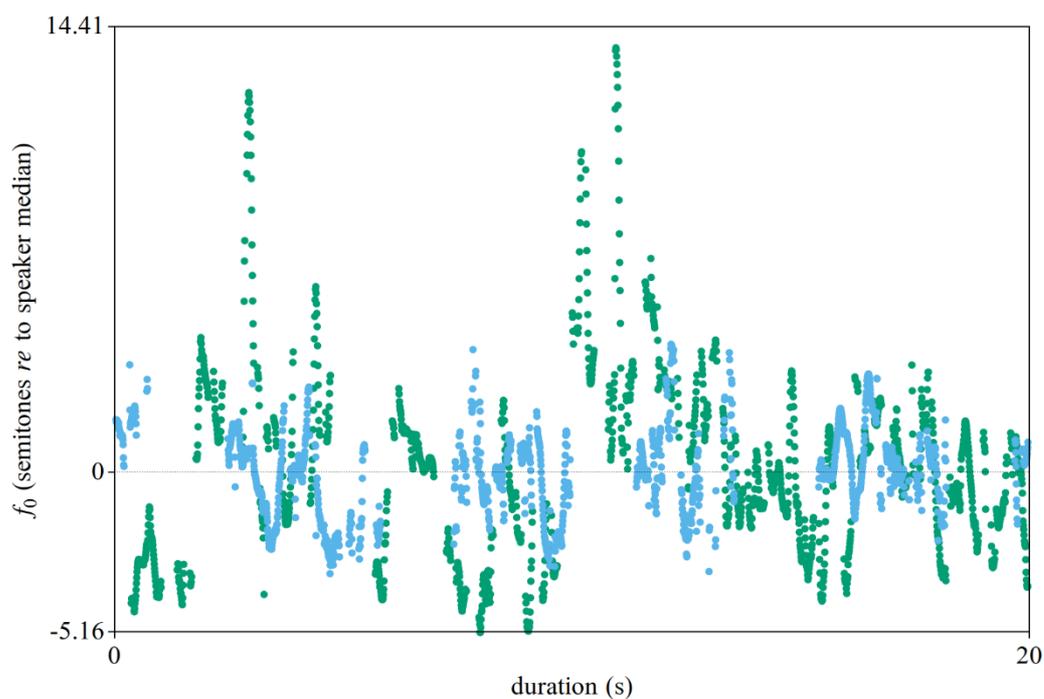


FIGURE 7 - Representative stretches of a typical female spontaneous f_0 contour (green dots) and a female sentence reading f_0 contour (blue dots). Hertz values are converted to semitones relative to the speaker median f_0 value.

Source: the author.

2.3 Bimodal distributions

In this section we present the results of the analysis that was carried out on bimodal distributions as described in section 1.3.2. First, we review the effect of speaking style and speaker sex on the occurrence of f_0 bimodality. Then we present the results of the Gaussian Mixture Models analysis. Last, we apply time-normalization to a set of f_0 contours to better understand what might be motivating the emergence of bimodality.

2.3.1 Effect of speaking style and speaker sex

Hartigan's dip test turned out significant p -values for eight samples. As noted in section 1.3.2, in order to avoid false positives, we visually inspected the histograms of these eight distributions and decided that two of them presented no sign of a second prominent mode and reclassified them as unimodal cases. To also avoid false negatives, histograms of the distributions that did not yield significant results in the dip test were visually inspected. In doing so, we identified three cases where there was evidence for bimodality. The net result is that a fair number of f_0 distributions are not unimodal: nine out of the 30 distributions, or 30% of the sample, according to the procedures described in section 1.3 and irrespective of scale (Hz or logHz). Speaking style has a major effect on this distribution: almost 78% of those occurrences (7 out of 9) correspond to the word reading style – the other two are from the sentence reading and the interview sample of the same male speaker (labeled m3). Out of all word reading samples (10), 70% are bimodal. Both sexes are roughly as likely to generate bimodal f_0 distributions: four out of five female speakers did it, all of them in the word reading style; among male speakers, four out of five also did it, three when reading words, and one when reading sentences and in the interview.

2.3.2 Mixture analysis

As noted in the Material and Methods section, all distributions classified as non-unimodal are bimodal, i.e., there are two prominent peaks in their histograms. The values obtained by the GMM analysis for the three parameters (mean, standard deviation and λ) that characterize the two component distributions present in each of the nine bimodal samples in the corpus are shown in Figure 7 as a function of speaker sex.

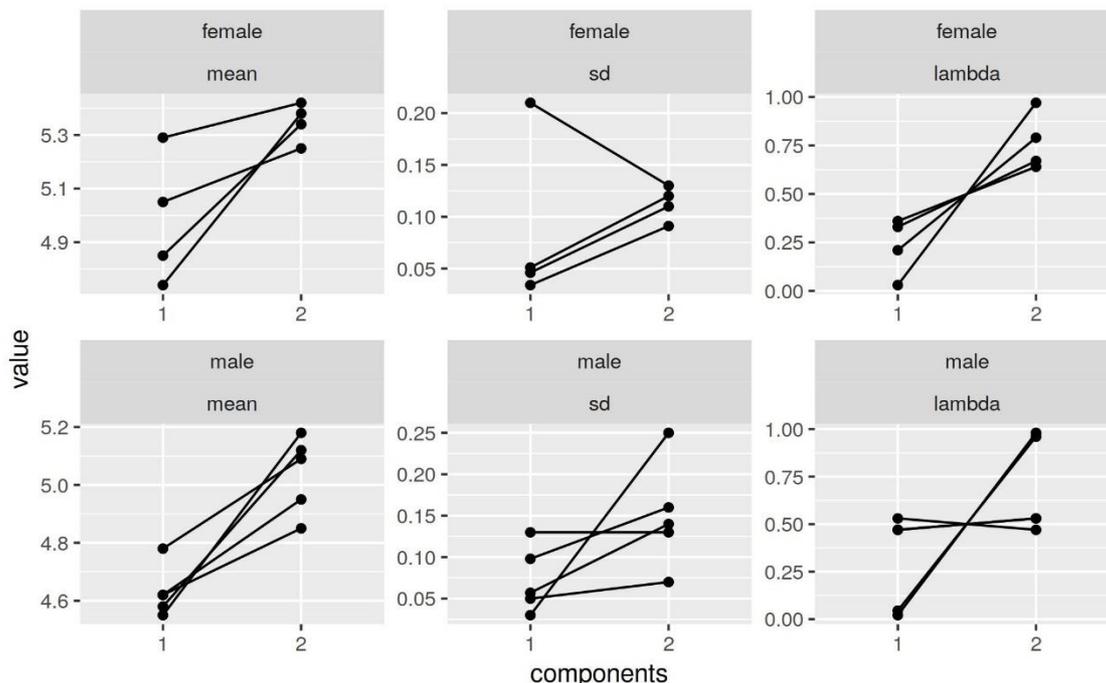


FIGURE 8 - Mean, SD and lambda values estimated by the GMM analysis of the two components (1 and 2 in the x-axis) present in the joint distribution of female and male speakers. Source: the author.

In terms of the mean values of the two component distributions, the mean difference is 0.36 logHz for the female speakers and 0.41 logHz for the male speakers, indicating that the lower component distribution is centered around values that are around 50% below the center of the distribution with a higher mean value. For 4 out of the 9 samples (2 females and 2 males), the lower component mean is 70% to 90% below the higher component mean, suggesting that at least part of this lower distribution can be in the non-modal voicing register.

There is a negative correlation of -0.74 for the difference between the second and first components' mean and the λ of the first component distribution, meaning that the further apart the means of the two components are, the less the lower component contributes to the overall distribution. The histogram in figure 1 is an example of this pattern: the mean of the first component's distribution is located at a point almost one octave below the mean of the second component's distribution and it accounts for around 20% of the overall distribution.

The variation coefficients of the mean parameter estimated by the GMM analysis of the two components are 1.72% and 2.13% for the female speakers and 1.56% and 2.94% for the male speakers. The component with the lower mean also tends to have a lower standard deviation and this pattern tends to be slightly stronger for the male speakers.

2.3.3 Time-normalized contours

Time-normalized contours of all words in word reading style for three speakers, m4, f5 and f1, are shown in figures 9, 10 and 11. They were chosen because they illustrate three different patterns, detailed below.

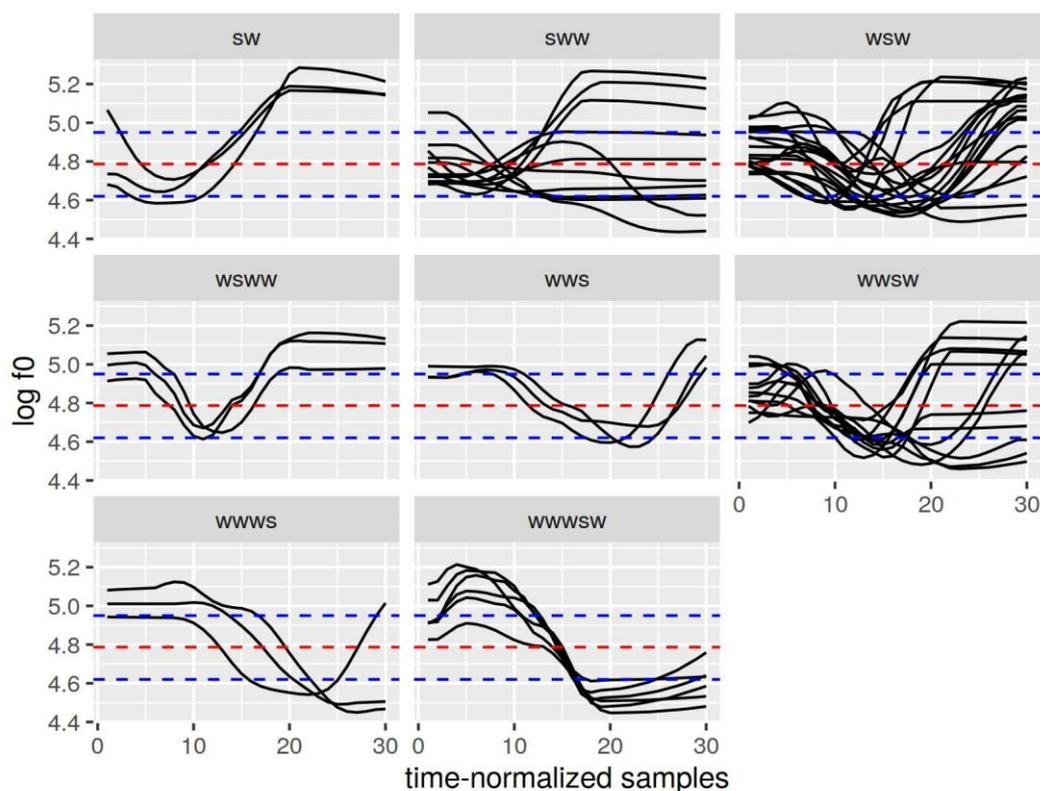


FIGURE 9 - Time-normalized contours of all words in a male speaker (m4) word reading sample. Different panels correspond to words with different stress patterns. Red dotted lines correspond to the overall median f_0 value. Blue dotted lines correspond to the mean values of the two component distributions identified by the GMM analysis.

Source: the author.

Speaker m4's overall f_0 distribution is bimodal, but the mean value of the two component distributions is not far apart (4.62 and 4.95 logHz, approximately half octave apart) and both account more or less equally for the overall distribution (53% and 47%). The range spanned by word contours is around 1.2 octaves. Blue dashed lines indicate the locations of the mean values of the two component distributions. A visual examination of the contours indicate that the two means approximately coincide with the rises and falls of the contours and that there is a balance in terms of the contribution of rises and falls to each word contour.

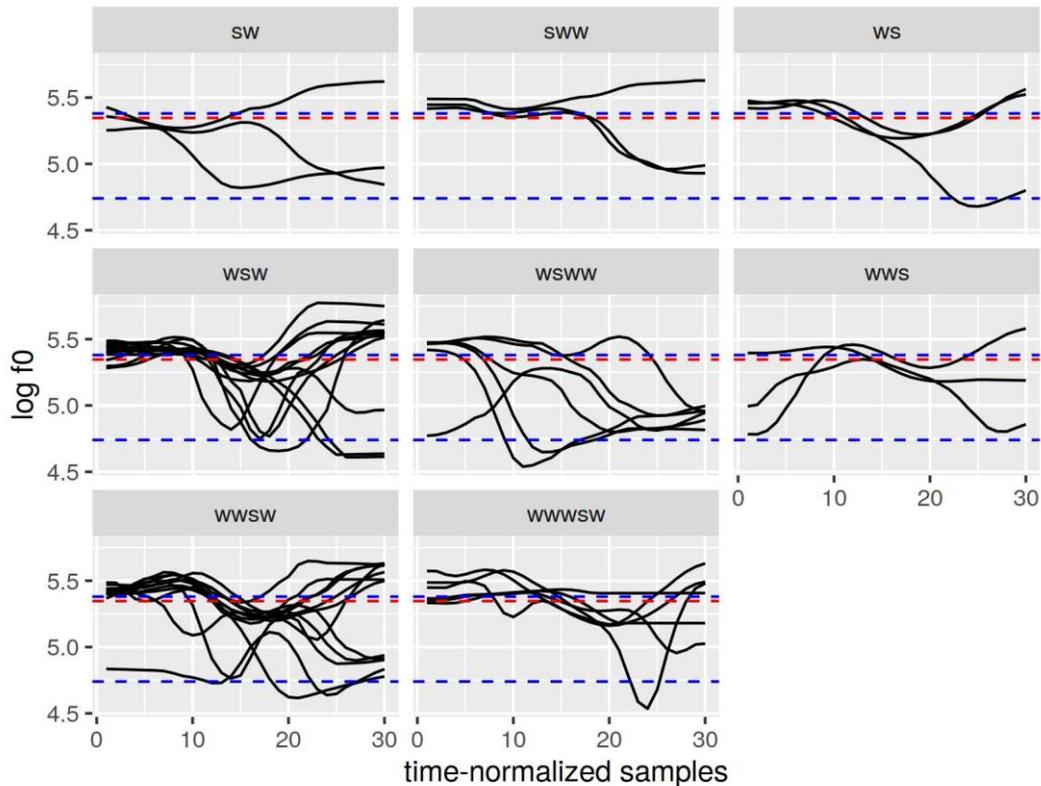


FIGURE 10 - Time-normalized contours of all words in a female speaker (f5) word reading sample. Different panels correspond to words with different stress patterns. Red dotted lines correspond to the overall median f_0 value. Blue dotted lines correspond to the mean values of the two component distributions identified by the GMM analysis.

Source: the author.

Speaker f5's overall f_0 distribution is bimodal, but the component distributions' means are located farther apart (4.74 and 5.38 logHz, approximately 0.9 octave) and the lower mean component distribution accounts for approximately 21% of the overall f_0 distribution. The range spanned by the word contours is around 1.8 octaves. Blue dashed lines indicate the location of the two component distributions' means. For this speaker, we can see that the overall median is very close to the mean of one of the component distributions. The other component has a lower mean value that is at the same level of most falls in the word contours; most falls reach deeper levels when compared to speaker m4 and the contour stays there for less time; this could explain why the lower component distribution contributes less to the overall f_0 distribution.

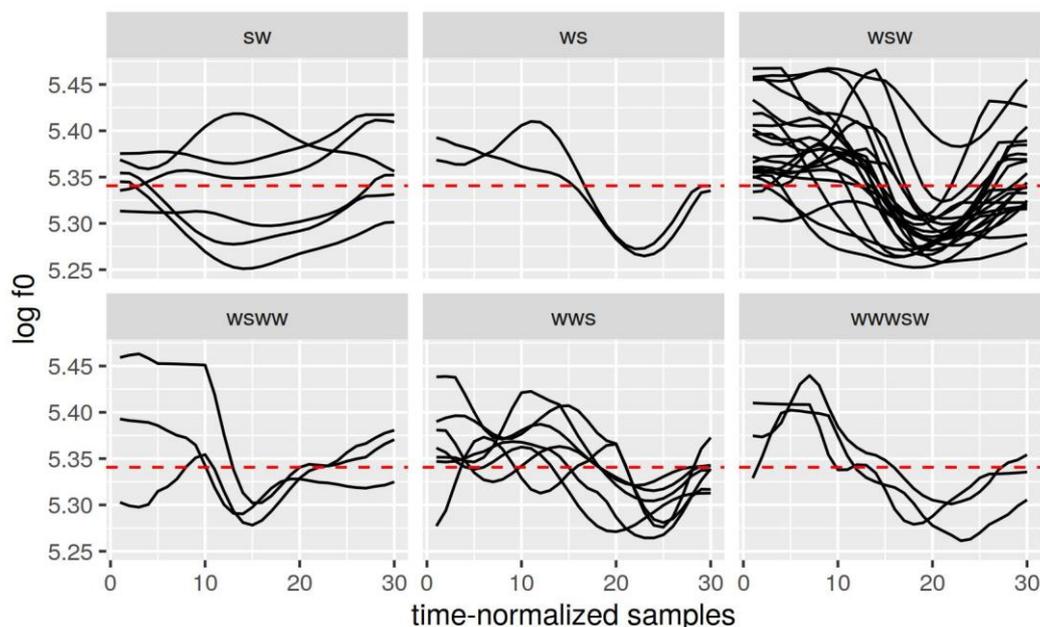


FIGURE 11 - Time-normalized contours of all words in a female speaker (f1) word reading unimodal sample. Different panels correspond to words with different stress patterns. Red dotted line corresponds to the overall median f_0 value. Notice how much narrower the range of values spanned by the contours show here in comparison to the ones in figures 8 and 9.

Source: the author.

As a comparison, speaker f1 overall f_0 distribution has a unimodal histogram and it is apparent from the word contours in figure 10 that there is not a great deal of intonation modulation – word contours span around 0.3 octave.

In terms of contour dynamics, word reading seems to be characterized by a good amount of f_0 movement in a relatively short time interval – mean word duration is 800 ms (SD 186) all speakers polled; see the mean word duration as a function of speaker in table 1.

In a language like BP, in which word stress is not fixed and there can be lexical contrasts defined by stress position within a word, the patterns observed seem to suggest that speakers tend to single out the stressed syllable against the background of unstressed syllables by aligning an upward or downward f_0 movement with it.

In general terms, the time-compressed nature of the contours in isolated word reading seems to be what causes the bimodality in f_0 histograms: contour rises contribute to an f_0 distribution with a relatively higher mean value and contour falls generate a distribution with a lower mean value; the ensemble distribution that results, then, is bimodal.

3. Discussion and conclusions

Taken as a whole, the results presented here corroborate tendencies already suggested in the previous literature about speaking styles with the added effort to be as explicit and complete as possible in the description of the phonetic and statistics analysis procedures to ensure an adequate level of reproducibility. Besides contributing to a well-established stream of previous studies on speaking styles, we also introduced something new by going a step further in the statistical modeling of f_0 distributions by trying a distribution fitting analysis that included a number of theoretical probability distributions other than the Normal or Gaussian.

We will start the general discussion by presenting the most important findings regarding the effect of speaking style on the shape of f_0 distributions. First, word reading style strongly favors the emergence of bimodal distributions regardless of speaker sex. In contrast, the other two styles seldom generate bimodal distributions. The mixture analysis shows that the two components that make up the bimodal distribution can be either closer or farther apart. When they are farther apart, the lower component is usually low enough that it is consistent with being in the non-modal register. Visual inspection of time-normalized f_0 contours coming from bimodal distribution cases suggests that the two components in the mixture can be associated with the time-compressed instantiation of intonational rises and falls that are linked with the signaling of stressed syllables. Recognizing the existence of a nontrivial amount of bimodal f_0 distributions, its prevalence in a particular speaking style and giving it a proper treatment is a new approach in the literature. Previous mentions to bimodality are limited to: (1) Jassem (1971), which reports the occurrence of one case of bimodal distribution (1 out of 10 speakers); Jassem *et al.* (1973, p. 215, fig. 3.4), which shows histograms of two f_0 samples by the same speaker that suggest bimodality; and (2) Kinoshita *et al.* (2010, p. 50), which states that “bimodal distribution was found to be very common due to creaky phonation” in a corpus comprised by samples of 201 speakers, although no quantification of its prevalence is provided or an objective procedure to detect bimodality is described. Kinoshita and colleagues (2010) attribute bimodality to creak phonation and the histogram presented in figure 1 in their paper seems to support this hypothesis - the lower of the two peaks is low enough to be compatible with non-modal phonation. In our data, a number of the cases could be attributed to non-modal phonation as well for the same reason: one of the peaks is located in the very low range of f_0 . In other cases, bimodality seems to arise from the fact that the typical contours in isolated word reading are characterized by lower and higher levels and brief transitions between them. Because of the time-compressed nature of the contours, the transition is so brief that the two peaks in the histogram are associated with the lower and higher levels with only a few data points between these levels. In some cases, the lower level is in non-modal phonation territory, but not in all of them. Regarding the occurrence of non-modal phonation in the corpus, we refer the reader to Silveira and Arantes (2017). In that study, the authors report the results of an auditory analysis of non-modal phonation occurrences in the same corpus analyzed here. The authors conducted auditory analysis of stretches of very low f_0 regions in contours with bimodal histograms and found that most instances of very low f_0 resulted in the

perception of non-modal phonation, although there was no attempt to quantify this association. This finding corroborates the frequent observation made in the voice quality literature that lowered f_0 levels are usually associated with the perception of laryngealized phonation (EDMONDSON; ESLING, 2006; ESLING; HARRIS, 2005; GORDON; LADEFOGED, 2001; HANSON; CHUANG, 2001; LAVER, 1980; REDI; SHATTUCK-HUFNAGEL, 2001).

Results concerning the effects caused by sentence reading and interview styles presented here indicate they are not uniform among male and female speakers. Female speakers show significant differences between styles: interview tends to have higher mean, standard deviation, skewness and kurtosis than sentence reading. Differences are statistically significant with large effect sizes for all estimators, except for kurtosis in log-scale (non-significant, moderate effect size). Male speakers show differences in the same direction (interview > sentence reading) for all four statistical estimators, although none of them is statistically significant. Effect sizes are smaller than the ones observed for female speakers for mean, standard deviation and kurtosis, but moderate for skewness. The present results are in line with what was reported by Arantes and Nascimento (2017) for the same data. In the earlier study, the difference in mean value in favor of the interview style in comparison to the reading style is greater in female speakers (1 semitone) than in male speakers (0.15 st), although no significant test is reported. For standard deviation, previous results showed that males present greater values than females (3 st vs. 2 st) irrespective of style; in terms of styles, interview presents a greater value than reading (2.92 st vs. 2.1 st), but values as a function of both sex and style are not presented. The results of the present study also show this tendency of male speakers to have larger standard deviation values than female speakers in both styles. Considering the results in the perspective of the broader literature on speaking style, they seem to confirm that, while for the majority of other languages studied read speech style presents larger f_0 mean and standard deviation than read speech, the opposite is the case for Brazilian Portuguese.

Now we turn our attention to the statistical characterization of f_0 distributions. The statistical tests reported in section 2.1 show that most distributions have skewness and kurtosis values that significantly deviate from what would be expected from normally distributed samples, regardless of measurement scale. The current results corroborate what the scarce previous literature (HORII, 1975; 1982; JASSEM; STEFFEN-BATÓG; CZAJKA, 1973; JASSEM, 1971) has reported: regardless of speaker sex or speaking style, f_0 distributions tend to have positive skewness and kurtosis values above 3. These results point to the fact that these distributions are asymmetric and heavy-tailed, strongly hinting that the normal distribution is not the best theoretical statistical distribution to model empirical f_0 distributions. Eriksson (ERIKSSON, 2011, p. 49–50) offers one possible explanation for why positive skewness arises in f_0 data: “positive skewing occurs primarily because there is much more room for fundamental frequency variation upwards than downwards”. Downward movement range is limited, according to this explanation, because going lower than a certain threshold “will normally result in creak which speakers tend to avoid” (*id.*, p. 57), whereas “there is, in principle, no corresponding upper limit, however, resulting in a distribution bias towards higher frequencies” (*ibid.*, p. 57).

Given the empirical results on statistical estimators of f_0 variability and possible explanations for this behavior, there is a clear gap in the literature. Until now, the available studies (JASSEM; STEFFEN-BATÓG; CZAJKA, 1973) only demonstrate that empirical f_0 data deviate significantly from normal distributions. No previous study that we are aware of tried to go beyond that and test the fitness of other theoretical distributions to f_0 data. The results of the distribution fitting analysis reported in section 2.2 try to fill this gap. As expected, none of the distributions in our corpus can be adequately modeled by a normal distribution. Even in the one case of an empirical distribution that was best modeled by a symmetric distribution, this distribution was the Logistic, not the normal. For all other 95% of unimodal empirical f_0 distributions, the best fit were right-skewed theoretical distributions, especially Burr type XII, a three-parameter heavy-tailed distribution, and in second place a two-parameter thin-tailed distribution called Gumbel. Both are used to model real-world phenomena such as survival data, insurance losses and income distribution that are characterized by the presence of events with extreme deviations from a central value. Given Eriksson's (2011) account of why f_0 data has pervasive positive skewness, one could associate the presence of relatively unbounded upwards excursions as a source of extreme values. The results presented in section 2.2.1 provide evidence that the three Burr distribution parameters estimated from the empirical f_0 distributions have a good amount of correlation with the empirical distribution's mean, standard deviation and skewness (and kurtosis indirectly, given its significant positive correlation with skewness). This result shows that the Burr distribution does a good job of capturing important information provided by three (four, indirectly) estimators that define the shape of a unimodal f_0 distribution.

Bringing together both the speaking style and the distribution fitting themes, section 2.2.2 shows that the effect of speaking styles on the shape of f_0 distributions can be represented by different combinations of the three Burr distribution parameters. The probability density plots in figure 6 generated from the parameter combinations for female and male speakers and the three styles show that they are able to represent the lack of a significant style effect in male speakers and adequately capture the larger effects seen in the female speakers, especially the fact that the interview style has greater mean, standard deviation and skewness than the sentence reading style. This result gives additional evidence for the usefulness of the Burr distribution to model the effect of a relevant paralinguistic phenomenon on the shape of f_0 distributions.

The results concerning the application of the so called "extreme value" probability distributions to f_0 data are encouraging since they are able to capture important features documented in empirical f_0 samples. Burr type XII stands out as the distribution that best fits the data in the speech material analyzed here. Future work should improve the present results by enlarging the number of speakers and including other languages as well to test if language has an effect on what distribution comes out as the best fit. Further suggestions for future studies include exploring how assuming that f_0 data follows an underlying distribution such as Burr type XII can be useful both in explaining linguistic phenomena (as we did here with speaking styles) and in practical applications such as speaker comparison in forensic contexts. In the latter case, we suggest checking if the distribution parameters are useful at capturing possible invariant features in f_0 distributions coming from the same

speaker in non-contemporaneous recordings such as those that Kinoshita and colleagues (2009; 2010) noted in unsystematic observations. Follow-up studies with a larger number of participants could test if individual speakers can be identified in a big pool on the basis of the values of the parameters that describe their f_0 distribution following the methodology used by Kinoshita and colleagues. A larger number of speakers can help estimate the degree of between-speaker variation in these parameters. These studies should also obtain the distribution parameter values for different speech samples by the same speaker to estimate the within-speaker variability. If the between-speaker variability is larger than within-speaker variability, then the parameter values are useful in speaker comparison tasks (NOLAN, 1993; ROSE, 2002).

A final suggestion for future work is to explore ways of adequately modeling bimodal f_0 distributions. Here we used Gaussian mixture models to analyze the data as a first approximation, but, given the evidence against f_0 being normally distributed, other mixture possibilities should be explored. Two important questions to be answered are whether the individual component distributions in bimodal cases have the same statistical characteristics as unimodal ones and whether the two underlying component distributions in bimodal cases of the same type, especially in the case where one of them is mostly comprised of creaky phonation.

On the whole, the results reported in this article show how speaking styles have an important role in shaping the overall shape of f_0 distributions. The patterns observed suggest a robust tendency for word list reading to generate bimodal distributions. Read sentences and the interview style are associated with unimodal but right-skewed distributions. The distribution fitting analysis corroborates previous suggestions in the literature that the normal distribution is not the best theoretical distribution to model f_0 distributions. We report results of an initial analysis showing that Burr type XII, an extreme value distribution, is the best statistical distribution to model empirical right-skewed f_0 distributions.

Acknowledgments

The author would like to thank Professor Anders Eriksson (Stockholm University) for generously granting access to the sound files of his project “A typology for word stress and speech rhythm based on acoustic and perceptual considerations”. The author acknowledges the work of Maria Érica Linhares (FAPESP grant 2014/21161-5), Suska Gutzeit (PIBIC-CNPq/UFSCar grant 2014–2015) and Isabela Silveira (FAPESP grant 2016/16544-8) in the processing of the f_0 contours analyzed here, done as part of undergraduate research projects carried under the author’s supervision. The author also thanks the reviewers for their careful reading and for their comments, which greatly improved the paper. Finally, the author thanks Julia Arantes and Leonardo Oliveira for proofreading the manuscript.

REFERÊNCIAS

- ANSCOMBE, F. J.; GLYNN, William J. Distribution of the kurtosis statistic b_2 for normal samples. *Biometrika*, v. 70, n. 1, p. 227–234, 1983. <https://doi.org/10.1093/biomet/70.1.227>.
- ARANTES, P. *better_f0: A Praat script for better f0 extraction*. [S.l.]: Zenodo, 2019. Available from: <https://zenodo.org/record/3470108>. Date accessed: dec. 25th, 2020.
- ARANTES, P. *Time-Normalized-F0: Praat script to perform time-normalization of F0 contours*. [S.l.]: Zenodo, 2018. Available from: <https://zenodo.org/record/1217159>. Date accessed: dec. 25th, 2020.
- ARANTES, P. Time-normalization of fundamental frequency contours: a hands-on tutorial. In: MEIRELES, A. R. (Org.). *Courses on Speech Prosody*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2015. p. 98–123.
- ARANTES, P.; ERIKSSON, A. Quantifying fundamental frequency modulation as a function of language, speaking style and speaker. In: INTERSPEECH 2019, 2019, Graz. *Anais... Graz: ISCA*, 2019. p. 1716–1720. <https://doi.org/10.21437/interspeech.2019-2857>.
- ARANTES, P.; LINHARES, M. E. N. Efeito da língua, estilo de elocução e sexo do falante sobre medidas globais da frequência fundamental. *Letras de Hoje*, v. 52, n. 1, p. 26–39, 2017. <https://doi.org/10.15448/1984-7726.2017.1.25419>.
- BENAGLIA, T.; CHAUVEAU, D.; HUNTER, D.R.; YOUNG, D. mixtools: An R Package for Analyzing Mixture Models. *Journal of Statistical Software*, v. 32, n. 1, p. 1–29, 2009. <https://doi.org/10.18637/jss.v032.i06>.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- BOERSMA, P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, v. 17, p. 97–110, 1993.
- COWAN, M. J. *Pitch and Intensity Characteristics of Stage Speech*. Iowa City: Department of Speech, University of Iowa, 1936.
- D'AGOSTINO, R. B. Transformation to normality of the null distribution of g_1 . *Biometrika*, v. 57, n. 3, p. 679–681, 1970. <https://doi.org/10.1093/biomet/57.3.679>.
- DELIGNETTE-MULLER, M. L.; DUTANG, C. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, v. 64, n. 4, p. 1–34, 2015.
- DUTANG, C.; GOULET, V.; PIGEON, M. actuar: An R Package for Actuarial Science. *Journal of Statistical Software*, v. 25, n. 7, p. 1–37, 2008.
- EDMONDSON, J. A.; ESLING, J. H. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, v. 23, p. 157–191, 2006.
- ERIKSSON, A. Aural/acoustic vs. automatic methods in forensic phonetic case work. In: NEUSTEIN, A.; PATIL, H. A. (Org.). *Forensic Speaker Recognition: Law Enforcement and Counter-terrorism*. [S.l.]: Springer, 2011. p. 41–70.
- ESKÉNAZI, M. Trends in Speaking Styles Research. In: EUROSPEECH'93, 1993, Berlin. *Anais... Berlin: [s.n.]*, 1993. p. 19–23.

ESLING, J. H.; HARRIS, J. G. States of the glottis: an articulatory phonetic model based on laryngoscopic observations. In: HARDCASTLE, W. J.; BECK, J. (Org.). *A festschrift for John Laver*. Mahwah, NJ: Lawrence Erlbaum Associates, 2005. p. 347–383.

FITCH, J. L.; HOLBROOK, A. Modal vocal fundamental frequency of young adults. *Archives of Otolaryngology*, v. 92, n. 4, p. 379–382, Outubro 1970.

FUJISAKI, H. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In: FUJIMURA, O. (Org.). *Vocal Fold Physiology: Voice Production, Mechanisms and Functions*. New York: Raven, 1988.

FUJISAKI, H.; HIROSE, K. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustic Society of Japan*, v. 5, n. 4, p. 233–242, 1984.

FUJISAKI, H.; OHNO, S.; GU, W. Physiological and Physical Mechanisms for Fundamental Frequency Control in Some Tone Languages and a Command-Response Model for Generation of their F0 Contours. In: INTERNATIONAL SYMPOSIUM ON TONAL ASPECTS OF LANGUAGES: WITH EMPHASIS ON TONE LANGUAGES, 2004, Beijing. *Anais...* Beijing: [s.n.], 2004. p. 1–4.

GOLD, E.; FRENCH, P. International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law*, v. 18, n. 2, p. 293–307, 2011.

GOLD, E.; FRENCH, P. International practices in forensic speaker comparisons: second survey. *International Journal of Speech Language and the Law*, v. 26, n. 1, p. 1–20, jun. 2019. <https://doi.org/10.1558/ijsl.38028>

GORDON, M.; LADEFOGED, P. Phonation types: a cross-linguistic overview. *Journal of Phonetics*, v. 29, n. 4, p. 383–406, 2001. <https://doi.org/10.1006/jpho.2001.0147>

HANSON, H. M.; CHUANG, E. S. Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of The Acoustical Society of America*, v. 106, n. 2, p. 1064–1077, 2001.

HARTIGAN, J. A.; HARTIGAN, P. M. The Dip Test of Unimodality. *The Annals of Statistics*, v. 13, n. 1, p. 70–84, 1985.

HIRST, D. J. The Analysis by Synthesis of Speech Melody: from Data to Models. *Journal of Speech Sciences*, v. 1, n. 1, p. 55–83, 2011.

HOLLIEN, H.; HOLLIEN, P.; DE JONG, G. Effects of three parameters on speaking fundamental frequency. *Journal of the Acoustical Society of America*, v. 102, n. 5, p. 2984–2992, 1997.

HOLLIEN, H.; PAUL, P. A second evaluation of the speaking fundamental frequency characteristics of post-adolescent girls. *Language and Speech*, v. 12, n. 2, p. 119–124, Abril 1969.

HORII, Y. Some statistical characteristics of voice fundamental frequency. *Journal of Speech and Hearing Research*, v. 18, n. 1, p. 192–201, 1975.

HORII, Y. Some voice fundamental frequency characteristics of oral reading and spontaneous speech by hard-of-hearing young women. *Journal of Speech and Hearing Research*, v. 25, p. 608–610, 1982.

JASSEM, W.; KUDELA-DOBROGOWSKA, A. Speaker-independent intonation curves. In: WAUGH, L.; VAN SCHOONEVELD, C. H. (Org.). *The Melody of Language*. Baltimore: University Park Press, 1980. p. 135–148.

JASSEM, W.; STEFFEN-BATÓG, M.; CZAJKA, S. Statistical characteristics of short-term average F0 distributions as personal voice features. In: JASSEM, W. (Org.). *Speech Analysis and Synthesis*. Warsaw: Panstwowe Wydawnictwo Naukowe, 1973. v. 3. p. 209–225.

JASSEM, W. Pitch and compass of the speaking voice. *Journal of the International Phonetic Association*, v. 1, p. 59–68, 1971.

JESSEN, M. Forensic phonetics and the influence of speaking style on global measures of fundamental frequency. In: GREWENDORF, G.; RATHERT, M. (Org.). *Formal linguistics and law*. Berlin: Mouton de Gruyter, 2009. p. 115–139.

KARLSSON, I.; BANZIGER, T.; DANKOVICOVÁ, J.; JOHNSTONE, T.; MELIN, H.; NOLAN, F.; SCHERER, K. Within-speaker variability due to speaking manners. 1998, Sydney, Australia. *Anais...* Sydney, Australia: [s.n.], 1998. p. 2379–2382.

KENDALL, T. *Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics*. London: Palgrave Macmillan, 2013.

KINOSHITA, Y; ISHIHARA, S.; ROSE, P. Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *The International Journal of Speech, Language and the Law*, v. 16, n. 1, p. 91–111, 2009.

KINOSHITA, Y; SHUNICHI, I. F0 can tell us more: speaker verification using the long term distribution. In: AUSTRALASIAN INTERNATIONAL CONFERENCE ON SPEECH SCIENCE AND TECHNOLOGY, 2010. *Anais...* Melbourne, Australia: [s.n.], 2010. p. 50–53.

KOMSTA, L; NOVOMESTKY, F. *moments: Moments, cumulants, skewness, kurtosis and related tests*. [S.l: s.n.], 2015. Available from: <https://CRAN.R-project.org/package=moments>.

KÜNZEL, H. Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, v. 4, n. 1, p. 48–83, 1997.

LAVER, J. *The phonetic description of voice quality*. New York: Cambridge University Press, 1980.

LIMPERT, E; STAHEL, W. A.; ABBT, M. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, v. 51, n. 5, p. 341, 2001.

LLISTERRI, J. Speaking styles in speech research. In: ELSNET/ESCA/SALT WORKSHOP ON INTEGRATING SPEECH AND NATURAL LANGUAGE, 1992, Dublin, Ireland. *Anais...* Dublin, Ireland: [s.n.], 1992. p. 1–28.

MAIDMENT, J. A.; LECUMBERRI, M. L. Pitch analysis methods for cross-speaker comparison. In: ICSLP 96, 1996, Delaware. *Anais...* Delaware: [s.n.], 1996.

MCLAUGHLIN, M. P. *Compendium of Common Probability Distributions*. 2016. Available from: https://www.causascientia.org/math_stat/Dists/Compendium.pdf. Date accessed: jul. 30th, 2019.

MIKHEEV, Y. Statistical distribution of the periods of the fundamental tone of Russian speech. *Soviet Physics-Acoustics*, v. 16, p. 474–477, 1971.

NOLAN, F. *The Phonetic Bases of Speaker Recognition*. Cambridge, UK: Cambridge University Press, 1993.

R CORE TEAM. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020. Available from: <https://www.R-project.org/>.

REDI, L.; SHATTUCK-HUFNAGEL, S. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, v. 29, n. 4, p. 407–29, 2001.

ROSE, P. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, v. 6, n. 4, p. 343–352, 1987.

ROSE, P. How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication*, v. 10, n. 3, p. 229–247, 1991.

ROSE, P. *Forensic speaker identification*. London: New York: Taylor & Francis, 2002.

SILVEIRA, I. J.; ARANTES, P. Efeitos do vozeamento não modal na descrição estatística de amostras de frequência fundamental. In: IV COLÓQUIO BRASILEIRO DE PROSÓDIA DA FALA, 2017, Mariana. *Anais...* Mariana: Associação Luso-Brasileira de Ciências da Fala, 2017. p. 119–122.

TRAUNMÜLLER, H; ERIKSSON, A. *The frequency range of the voice fundamental in the speech of male and female adults*. [S.d.]. Available from: http://www2.ling.su.se/staff/hartmut/fo_m&f.pdf. Date accessed: dec. 25th, 2020.

WAND, M. *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. [S.l: s.n.], 2015. Available from: <https://CRAN.R-project.org/package=KernSmooth>.

WESTFALL, P. H. Kurtosis as Peakedness, 1905–2014. R.I.P. *The American Statistician*, v. 68, n. 3, p. 191–195, 2014.

WOLFRAM RESEARCH. *Heavy Tail Distributions*. Available from: <https://reference.wolfram.com/language/guide/HeavyTailDistributions.html>. Date accessed: dec. 25th, 2020.

WOLFRAM RESEARCH. *NormalDistribution*. Available from: <https://reference.wolfram.com/language/ref/NormalDistribution.html>. Date accessed: dec. 25th, 2020.

ZEMLIN, W. *Speech and Hearing Science*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.