Functional Heirarchy in Vocal Production:
Relationships Between Sex Assigned At Birth, Height, and Frequency in Pitch-Modulated Speech.

Ashlae Blum, November 29, 2024.

Introduction

The hierarchical nature of speech in human communications remains a subject of great depth and fascination. From linguistics, we affirm considerable discourse on the origins of form and function[1]; a rose by any other name (*Shakespeare)*, and the second blue ball (*Fig.1*), are but two illustrative examples of this phenomenon [1]. From a physical as well as psychoacoustic lens, the mechanics of consonant and vowel production further highlight the perceptual consequences of vocal production [2,3]. With respect to empirically measurable acoustic parameters of speech, vocal characteristics such as fundamental frequency and formant dispersal of are high importance in determining emotive content in speech [2,3,4,5]. They are also indicative of the physical characteristics of the source and filter of a sound producer, and can tell us a variety of measurable properties about that individual, such as vocal tract length [3,4,5]. For example, formants are closely related to front-back vowel production, and higher-order formants are associated with articulation and speech rate [3,5]. Using a variety of techniques from spectral analysis, these patterns may be qualitatively as well as quantitatively measured (*Figs.2,3*). The authors of the study are herein motivated to examine the implications of physical characteristics (sex assigned at birth and size) on acoustic phenomena (fundamental frequency, f0, and formant spacing) in both natural and pitch-modulated vocal production.

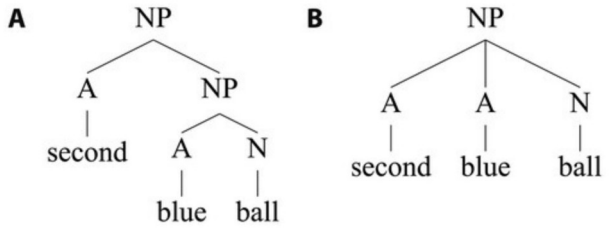

Fig 1. Hierarchical (a) and linear (b) representations for the phrase "second blue ball".

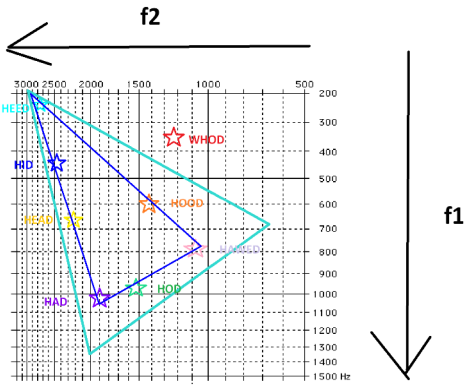

Fig 2. First- and second-order formant pitch diagram for an AFAB English-speaker. This illustrates supralaryngeal characteristics in vowel production.

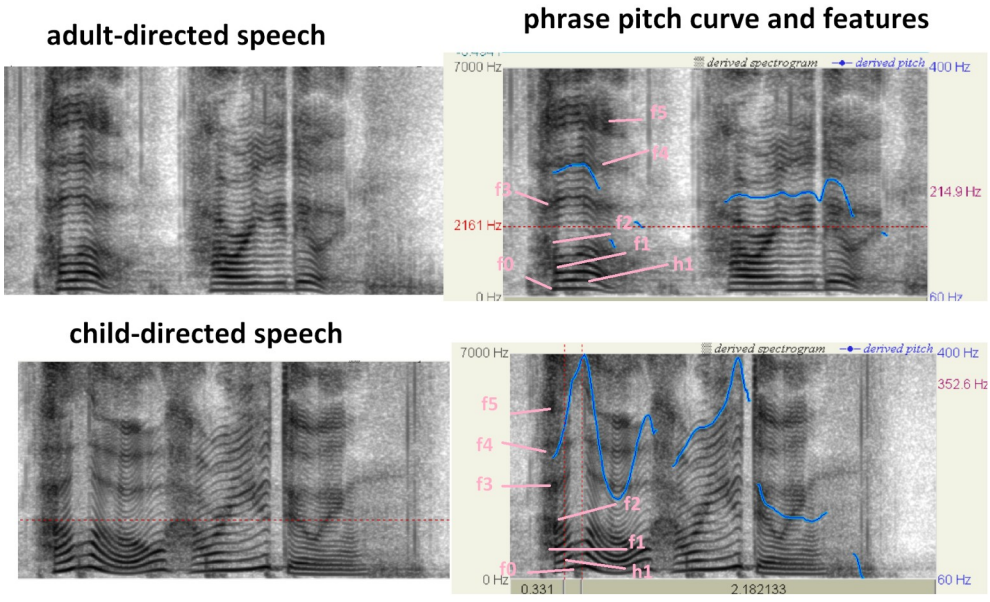

Fig 3. Spectrograms of adult- and child- directed speech. Here, the adult-directed speech was produced with a "natural" speaking voice, and child-directed speech was produced with a higher-pitched voice, with fundamental frequency varying widely over the duration of the signal. Formant structures of vowel production may be perceived as dark curves overlaid on top of the harmonics of the signal.

Experimental Methods

To determine the effect of physical characteristics on vocal production in natural and pitch-modulated speech, we analyzed a dataset of 198 vocalizations collected from 66 individuals. The recordings included 3 sets of vocalizations: one natural, and two pitch-modulated. Natural vocalizations from each speaker were recorded and used to determine the fundamental frequency and formant structure of their everyday voices. Pitch-modulated vocalizations were used to determine the same prior vocal characteristics as speakers attempted to sound as small and as large as possible, respectively. Additional data concerning binary-valued sex assigned at birth and height was collected for each speaker. The presence of transgendered individuals was not indicated in this study.

Analysis Methods

Data was processed and analyzed using R language. Since the experiment included a variety of non-parametric variables, we compared both single- and multivariate analysis methods to ascertain relationships between variables. Both sex and age were unevenly represented in the data, so weights were constructed, and the data was cleaned for uniformity. Due to known differences in vocal production between AMAB and AFAB speakers [3], binary sex value was indicated in our comparisons.

We determined the fit of the data by constructing statistical plots of f0 and formant spacing using height and age, while preserving a binary sex structure to comparatively visualize the data. The dataset was cleaned, filtered, and smoothed. Upon inspection, the data seemed to be right skewed, though values were truncated due to a lack of datapoints (*Figs.4,5*). A LOESS fit was used for analysis because of the small dataset size and skewness (*Fig.6*). As such, we were able to visualize finer nonlinear relationships within the data. Further computational details may be seen in the codebase, linked here and below.

Results

The distributions for AMAB and AFAB f0 as well as formant spacing data were both right skewed (*Figs.4,5*), though their tails were cut off due to both physical limitations of the participants as well as a lack of datapoints. A wavelet-like microstructure in the tail was observed, indicating that there may be fine-grained information encoded in the higher-order formant structures (*Figs.4,5*). (May be related to the "even formant pairs" observed previously in the practical.) We observe that AFAB vocalizations tend to have a wider or broader spread in frequency distribution (*Figs.4,5*).

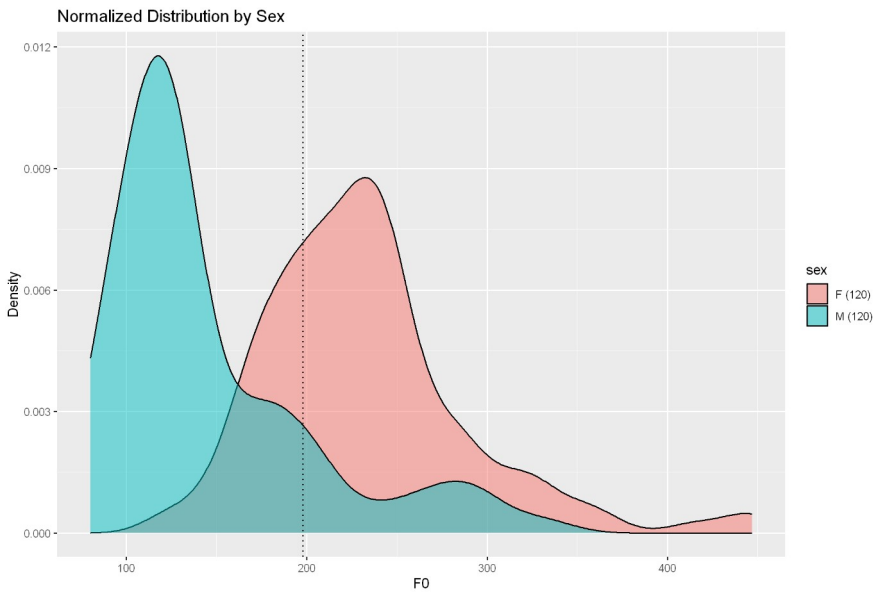

Fig. 4 Distribution of f0 vs. density for AFAB and AMAB vocalizations.

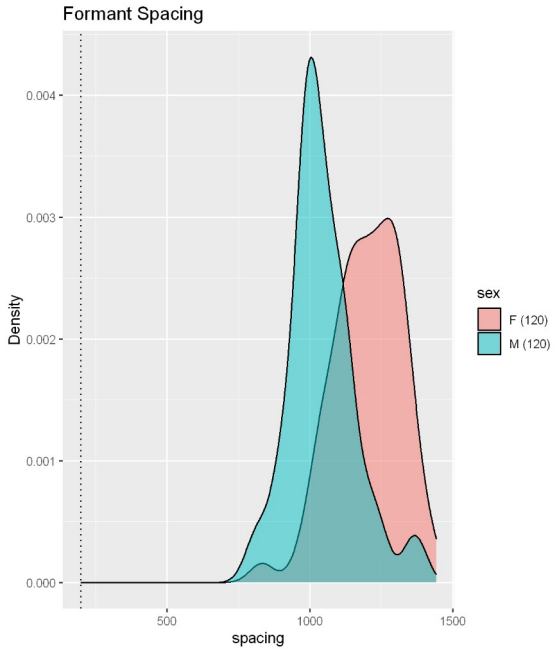

Fig. 5 Distribution of formant spacing density for AFAB and AMAB vocalizations.

Fig. 6 Comparison of LOESS fits for f0 and formant spacing by height, age, and sex. Shadowy grey areas indicate areas of overlap between sex in respective categories.

Discussion

From our study, we conclude that both fundamental frequency and formant spacing are right-skewed with respect to their distributions (*Figs.4,5*). This indicates that the relative center of pitch and formants are centered towards the lower end of the theoretical range of our distribution. The theoretical maximum stretching of the voice into higher-pitched regions may be considered a higher limit into which only certain outliers stretch.

We can also see that the plots of f0 and formant spacing with respect to age have significant overlapping regions across binary sex categories (shadowy grey areas) for big, normal, and small size representations (*Fig.6*). That means that in a sense, the gender lines are indeed blurred: AFAB and AMAB individuals occupy the same vocal space in a variety of pitch-constrained vocal production instances. This could also be reflective of the physical limitation, or pitch ceiling / floor, as people try to raise their voices higher or lower. Of note is that the 'small' sizing had the most overlap area between sexes, with formant spacing and age overlaps being the most highly correlated (*Fig.6*). It might be interesting to explore the effects of these same characteristics on pitch ceiling and floor as theoretical maximum and minimum limitations of the human voice, and to consider the structural dispersal of higher-order formants within these limits.

Extensions

For future studies, we would like to highlight a variety of data-driven considerations. First, equivalent representation is needed from across sex- and age- based categories to determine the validity of further analyses. In addition, it may be useful to consider higher-order nonparametric density estimations such as a gaussian psi-angle kernel, or multidimensional weights. For example, if one takes height as a function of sex or formants as a function of fundamentals into account, there may be interesting emergent properties in the data that will require nonlinear analysis to determine. Also of interest is the use of wavelet transformations in curve approximation. We saw here that the tail of our distribution exhibited fluctuations, which may also be indicated in the studies of the practical from this week (see Blum, Week 12 Practical for reference). Especially in the context of their extensive experience with modes of professional vocal production (in voice-adjacent industries such as music, theatre, film, and transgender speech dynamics) the authors feel that it merits consideration to explore the microfluctuations of higher-order formants in the context of pitch-limited vocal production. For example, referring to the discussion of pitch floor/ceiling above, it could be interesting to conduct a study in which people with relatively good pitch control naturally modulate their voices (speak while keeping voices at the same frequency) so as to explore the hierarchical relationships between higher order formant structures as the voices near the physical limits of the vocal range.

Codebase

https://github.com/laelume/bioacoustics/tree/main/vocalproduction/practical

Citations

[1] Coopmans, C. W., de Hoop, H., Kaushik, K., Hagoort, P., & Martin, A. E. (2021). Hierarchy in language interpretation: evidence from behavioural experiments and computational modelling. *Language, Cognition and Neuroscience*, *37*(4), 420–439. https://doi.org/10.1080/23273798.2021.1980595

[2] Zhang Z. (2016). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, *140*(4), 2614. https://doi.org/10.1121/1.4964509

[3] Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

[4] Pisanski, K., & Reby, D. (2021). Efficacy in deceptive vocal exaggeration of human body size. *Nature Communications*, 12, 968. https://doi.org/10.1038/s41467-021-21008-7

[5] Anikin A, Pisanski K, Reby D. Static and dynamic formant scaling conveys body size and aggression. R. Soc. Open Sci. 9: 211496 (2022). https://doi.org/10.1098/rsos.211496