Published in final edited form as:

J Voice. 2017 May ; 31(3): 382.e9–382.e13. doi:10.1016/j.jvoice.2016.08.014.

A FORMANT RANGE PROFILE FOR SINGERS

Ingo R. Titze,

National Center for Voice and Speech, Salt Lake City, Utah, 84101 USA, Department of Communication Sciences and Disorders, University of Iowa, Iowa City, Iowa, 52242 USA

Lynn Maxfield, and

National Center for Voice and Speech, Salt Lake City, Utah 84101 USA

Megan Walker

National Center for Voice and Speech, Salt Lake City, Utah 84101 USA

Abstract

Vowel selection is important in differentiating between singing styles. The timbre of the vocal instrument, which is related to its frequency spectrum, is governed by both the glottal sound source and the vowel choices made by singers. Consequently, the ability to modify the vowel space is a measure of how successfully a singer can maintain a desired timbre across a range of pitches. Formant range profiles (FRPs) were produced as a means of quantifying this ability. 77 subjects (including trained and untrained vocalists) participated, producing vowels with three intended mouth shapes, (1) neutral or speech-like, (2) megaphone-shaped (wide open mouth), and (3) inverted-megaphone-shaped (widened oropharynx with moderate mouth opening). The first and second formant frequencies (F_1 and F_2) were estimated with fry phonation for each shape and values were plotted in F_1 – F_2 space. By taking four vowels of a quadrangle /i, æ, a, u/, the resulting area was quantified in kHz^2 (kHz squared) as a measure of the subject's ability to modify their vocal tract for spectral differences.

Keywords

singing; formants; formant range; singing vowels; belting

I. INTRODUCTION

Singing styles are in part differentiated by the timbre of the voice. In some instances, the use of a particular timbre may be a simple artistic choice by the vocalist, but it can also be an acoustic requirement in many cases, determined by the size of the hall (if singing is unamplified), the instrumental accompaniment, or the dramatic content of the song. For example, if strings and flutes are the primary accompaniment, a “brassy” voice timbre produces a contrast in timbre. On the other hand, if brass and percussion dominate the

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

accompaniment, a “flute-like” voice timbre produces a contrast. As instrumental accompaniment, performance spaces, and electronic enhancement vary, vocal timbre may change in response.

According to the American Standards Institute, timbre is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar (ANSI, 1994). Acoustically, timbre is thus largely related to the frequency spectrum in the voice. Regulation of the spectrum involves adjusting the relative amplitudes of individual harmonics, which can be accomplished at the source or by the choice of vowels and allophonic variations of these vowels. Specifically, recent investigations have shown that the strength of the second harmonic relative to the first harmonic is a major difference between vocal styles (Bjorkner, 2008; Titze et al, 2009, 2011; Bourne & Garnier, 2012; Sundberg & Thalen, 2015). For example, to maintain a female *belt* or a *call* timbre at medium-high pitches (G5 - C5 and occasionally up to F5), as is frequently required in music theatre styles, the first resonance frequency of the vocal tract f_{R1} is often raised to its upper limit (much higher than in conversational speech) to reinforce a second harmonic ($2f_o$) that is already strong at the source with modal (chest) registration. However, because there is a limit in how high this resonance frequency can be raised, music theatre singers are limited in their f_o range if they want to retain the dominance of $2f_o$ in the spectrum. Conversely, in operatic styles in the C4 to C5 range (males and females), the first and second resonances are often lowered by widening the oropharynx (Sundberg et al. 1993; Stone et al. 2003; Miller, 2008) to balance the harmonic energies of f_o , $2f_o$ and $3f_o$. Mixed registration at the source can also be used along with the vocal tract modifications. Sometimes the energy of $3f_o$ is reinforced with the second resonance (Miller, 2008).

[It is noted parenthetically that in the current study, as in many other studies involving resonances of the vocal tract, a direct measurement of the resonance frequencies f_{Rn} is not attempted. Rather, the assumption is made that the resonance frequencies closely mirror the formant frequencies F_n , measured as peaks in the output spectrum at the mouth. Hence, the approximation $f_{Rn} \approx F_n$ is assumed in subsequent discussions and the symbolic notation F_n is used (Titze et al, 2015) with the full understanding that F_n is not exactly equal to f_{Rn} .]

Continuing with classical operatic singing, lowering F_1 and F_2 also has the effect of creating a spectral distance between the vowel-determining formants (F_1 and F_2) and those formants often referred to as the singer’s formant cluster (F_3 , F_4 , and F_5). This distance allows the listener to more clearly distinguish *chiaroscuro* (bright-dark) timbre in a singer’s voice. One form of brightness, also referred to as vocal ring (Bartholomew, 1932), is related to the increased energy in the singer’s formant cluster (Sundberg, 1977), a hallmark of unamplified singing in the operatic style. However, lowering F_1 and F_2 by simply using a small mouth opening (as in the case of a speech-like /u/ vowel) would reduce the radiated output power too much for unamplified singing. Hence, stretching of the oropharyngeal space beyond what is typical for speech allows a moderate (/ɔ/-like) mouth opening to be maintained while F_1 and F_2 are lowered.

The purpose of this investigation was to determine the degree to which singers with limited access to sophisticated laboratory instrumentation can develop a personalized F_1 – F_2 formant

plot to establish the f_0 range over which contrasting singing styles can be performed. It is known that vocal tract resonances are affected by vocal tract length and cross-sectional area variation (Fant, Section 2.3) and that F_1 and F_2 largely determine vowel perception (Peterson & Barney, 1952). Therefore, the degree to which an individual can stretch the dimensions of the vocal tract determines the range of vowel modification or adjustment available to that individual (de Boer, 2010). We propose that providing an understanding of an individual's formant range profile, in a manner similar to a voice range profile for pitch and loudness, can help the individual better determine appropriate use of vowels for various stylistic vocal needs.

II. METHODS

Vowels were divided into three categories, based on stylized vocal tract shapes: neutral (speech-like), megaphone (belt or call), and inverted megaphone (opera-like), following shape descriptions of Titze and Worley (2009) for males and Titze, Worley and Story (2011) for females. Table I lists the vowels we placed in each category. The first (neutral) category consists of vowels that are consistent with an individual's general speech gestures. The second category, the megaphone, or divergent, mouth shape, is based on the above studies and other pedagogical literature (Estill, 1992; Bozeman, 2013). It is known to produce a timbre that is often described as open, yell-like, or brassy. This shape facilitates the $2f_0$ proximity to F_1 . It is created with a wide opening at the lips, a narrow pharynx, and sometimes an elevated larynx (Bozeman, p. 22). It is characteristic of musical theater singing, some jazz and pop singing, and high-energy gospel singing. The third category, the inverted megaphone or convergent mouth shape, produces a timbre described as full and deep if higher harmonics are sufficiently present, but also a "hooty" or "fluty" quality when first harmonic (f_0) energy dominates (Bozeman, p. 23). To create a dark "hooty" timbre, one must have a wide oropharyngeal space and a narrower space toward the front of the mouth. Lessac (1996) has called this timbre the "Y-buzz" in speech, in contrast to the "call" timbre from which belting is derived. The corresponding vocal tract shapes are not unique to humans. Chimpanzees are known to produce a "pant-hoot" and a "scream", two sounds that have a similar dichotomy in vocal tract shape and related sound quality.

A. Subjects

Data were gathered as a regular classroom activity during the Summer Vocology Institute conducted at the University of Utah. The class was entitled Voice for Performers, but only some registrants were performers; others were speech-language pathologists interested in studying the singing voice. Students ranged in age from their early 20s to over 50 years old and experience in vocal training, singing and/or acting voice, similarly varied widely across the subject pool. In total, 77 subjects (54 females and 23 males) participated in this study. Data were collected over a period of three years and subsequently compiled by the authors for final analysis.

B. Equipment and Procedures

Students sustained each vowel shape in vocal fry (as opposed to high-pitched phonation) in order to energize the vocal tract with a dense frequency spectrum, thereby obtaining the best

resolution of formant peaks. It was anticipated that recording fidelity would not be an issue for obtaining formant frequencies; hence, a variety of microphones, including Shure SM48s (Niles, IL) and internal microphones on students' laptops, were used to obtain the acoustic signal. The project was conducted in a space with minimal background noise (below 65 dB) to reduce any interference of the acoustic signal, but a sound isolation booth was not deemed necessary for frequency analysis.

Students were instructed to produce five speech vowels (/i,e,ɑ,o,u) in a manner that felt like their “normal” speech production, but in vocal fry instead of high-pitched phonation. Next, they were instructed to produce the “megaphone” vowels (æ, a, ɔ) by widening the mouth opening as much as possible while maintain the intelligibility of the vowel. Finally, they were instructed to produce the “inverted megaphone” vowels (ɪ, ɛ, ʌ, ʊ) with a moderate mouth opening, but widening the oropharynx as much as possible while still maintaining the intelligibility of the vowel. As the students carried out these tasks, one or more of the authors monitored the outputs perceptually and provided further instruction when necessary to encourage individual subjects to widen their pharynx or mouth more. Some subjects expressed difficulty in sustaining vocal fry on all vowel configurations, but with practice, all were able to complete the task.

Once the tokens had been recorded, students used either Praat (Amsterdam, The Netherlands) or Voce Vista (Princeton, NJ) to measure and record approximate frequencies of F_1 and F_2 . In a formal research setting, vocal tract resonance frequencies f_{Rn} would have been obtained (Epps, Smith, and Wolf, 1997) instead of formant frequencies, but an important objective was for students to be able to obtain these measurements by themselves, quickly and economically. For students using Voce Vista, the default settings (10 Hz bandwidth, and power spectrum frequency scale of 0–5000 Hz) were preserved. A time point was selected in the middle of a steady portion of the vocal fry audio recording. The power spectrum was then visually inspected and frequencies for the first two peaks, F_1 and F_2 , were recorded. For students using Praat, the spectrogram was viewed with a range of 0–5000 Hz, window length of 0.025 s and a dynamic range of 90 dB. A steady segment of the audio signal, 250 milliseconds in length, was selected, and the first two formants were retrieved using the following settings: Maximum formant of 5000 Hz, window length of 0.025 s, and dynamic range of 90 dB. As the vocal tract configurations changed and formant frequencies shifted accordingly, the “number of formants” setting was adjusted manually for each token, usually between five and seven, until the formant contours reasonably matched the intensity maximums visible in the spectrogram.

Using one of these two analysis tools (Voce Vista or Praat), F_1 and F_2 were extracted by the students for the following sequences: Speaking Vowels - /i, e, a, o, u/; inverted-megaphone - mouth shape - /ɪ, ɛ, ʌ, ʊ/; and megaphone-mouth-shape - /æ, a, ɔ/. Students created their own personalized $F_1 - F_2$ chart as a reference tool for future use in helping determine vowel choices on certain pitches in various singing styles. Recordings of 24 of the subjects were reviewed by the authors, using the Praat software, to check for validity of the students' measurements. In addition to the personalized $F_1 - F_2$ charts produced by the students, the authors used the F_1 and F_2 values measured by the students and calculated the area of the quadrangle created by plotting F_1 against F_2 for the vowels /i/, /æ/, /a/, and /u/. These vowels

were chosen as they generally presented the extreme high and low values for the two formant frequencies. Area for the quadrangle was calculated from these points as

$$A = \left| \frac{(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_4 - x_4y_3) + (x_4y_1 - x_1y_4)}{2} \right|$$

where x_n and y_n are the frequency values of F_1 and F_2 respectively for each vowel.

III. RESULTS

Figure 1 shows the F_1 – F_2 vowel space for the five selected speech-like vowels across all subjects, females on the left and males on the right. The vowels /æ/ and /a/ are excluded because they are in a special class for singing in a belt style. The /i/ and /u/ vowels best define the range of F_2 , ranging from 600 Hz to 2900 Hz in females and 500 Hz to 2600 Hz in males.

As two single subject examples, consider a female singer and a male singer characterized by the average of the formant frequencies shown in Figure 1. Further, consider that an attempt is made to sing these vowels on a G4 musical note, the middle of the C4–C5 octave where most of the timbre adjustments occur in singing. Figure 2 shows the average female and male formant frequencies for the five speaking vowels. Superimposed are grid lines that represent the first three harmonics for the note G4 (392 Hz). Note that F_1 of the vowels /u/ and /i/ is close to f_0 . The timbre for /i/ and /u/ would be flute-like because of the dominance of first harmonic energy. Similarly, F_1 for the /a/ vowel is close to $2f_0$, which means that the timbre for /a/ would be brass-like. (Brass instruments generally have a strong second harmonic). The timbre with these corner vowels can change dramatically with small changes in f_0 . In some singing styles a sudden timbral change is desirable, whereas in other styles a sudden timbral change is avoided by choosing vowels that do not have formants close to a specific harmonic. The vowels /o/ and /e/, for example, would appear to produce a stable timbre for a range of notes below and above G4.

Figure 3 shows the F_1 – F_2 space for style-specific singing vowels, the megaphone shaped vowels /æ/ and /a/ and the inverted megaphone shaped vowels /ɪ/, /e/, /ʌ/, /ɔ/ and /ʊ/. Note that the inverted megaphone shaped vowels have F_1 frequencies that lie between f_0 and $2f_0$, indicating a timbre balance that could likely be maintained over several notes in either direction from G4. Classically trained singers use these vowels in the *passaggio* region of their voice, between D4 and A4. To the contrary, the megaphone shaped vowels have F_1 above $2f_0$, favoring this second harmonic in the spectrum. For the males, there is almost a tuning of F_1 to $2f_0$ on this one note, but exact tuning to any single harmonic is not required for stylistic distinctions. Females experience the $F_1 - 2f_0$ proximity (or near-tuning) at slightly higher notes, around A4 – B4, or 440 – 490 Hz. For /æ/ and /a/, F_1 is about 100 Hz higher than reported by Peterson and Barney (1952) for males and females, indicating that a formant stretch took place. However, on average the students did not come close to the 1000 – 1200 Hz F_1 values achieved by top-rank professional theatre singers (Titze, Worley, and Story, 2011).

Figure 4 shows FRPs for four individuals, two males and two females who showed the greatest difference in the area of the vowel quadrangle /i/-/æ/-/a/-/u/. If vowels are produced according to the procedure outlined earlier, this quadrangle measures the degree of articulatory stretch that the individual is capable of producing. The vowels /æ/ and /a/ stretch F_1 and the vowels /i/ and /u/ stretch F_2 . Since formant frequencies can be measured in kHz, the area of the quadrangle here is expressed in kHz^2 . The two females produced formant ranges of 0.195 kHz^2 and 0.911 kHz^2 , while the two males produced formant ranges of 0.337 kHz^2 and 0.625 kHz^2 .

IV. DISCUSSION AND CONCLUSIONS

A primary objective of this study was to determine if singers with limited access to laboratory instrumentation could personalize an F_1 – F_2 vowel chart for the purpose of understanding voice timbre in singing styles. The original intent was to produce a maximum formant range profile for each individual. This was not accomplished with our set of instructions. To produce such a maximum range, it is likely that an articulatory model will first be needed to determine which configurational stretches will produce the ultimate formant stretches. With that determination, a better set of stretch-instructions may result, ultimately leading to a formant range profile (FRP) that can be considered maximum for vowel availability toward a given style and a given pitch range. This study was only a first step in this direction.

Vocal timbre is in part predictable on the basis of the proximity of specific harmonics to specific formants. Unlike in most man-made musical instruments, in which an attempt is made to preserve a distinct timbre of the instrument by simultaneously tuning multiple harmonics to resonances, the biological vocal instrument relies on highly variable vocal tract shapes to adjust the output spectrum. Results here showed that the area of the vowel quadrangle (in kHz^2) could be more than a factor of 2 across individuals. Future studies could address the effect that voice training can have on the ability to stretch the formant range profile for a given individual. Also of interest would be the effect that native language, race, or body size could have on the FRP.

Quantification of the degree of *reduction* of formant space was not addressed well in this study. Future studies could address a quantification of the oropharyngeal stretch for the centralized vowels /I/, /e/, /Λ/, /ɔ/ and /ʊ/ with mouth openings that allow maximum sound radiation from the lips. Here the question would be: how small can the area in the $F_1 - F_2$ space be to preserve both the timbre and vowel intelligibility when the objective is to get the most sound to the listener. Measurement of lip opening and MRI or CT scanning would be the obvious methodology to quantify the stretch.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grant R01 DC012045, awarded to Ingo R. Titze, Principal Investigator.

References

- American national standard acoustical terminology. American National Standards Institute. 1994 **ANSI S1-1-1994** (R1999).
- Bjorkner E. Musical theater and opera singing – why so different? A study of subglottal pressure, voice source, and formant frequency characteristics. *J Voice*. 2008; 22(5):533–540. [PubMed: 17485197]
- Bourne T, Garnier M. Physiological and acoustic characteristics of the female music theater voice. *J Acoust Soc Am*. 2012; 131(2):1586–1594. [PubMed: 22352528]
- Bozeman, KW. Practical vocal acoustics: Pedagogic applications for teachers and singers. Hillsdale, NY: Pendragon Press; 2013.
- De Boer B. Investigating the acoustic effect of the descended larynx with articulatory Models. *J Phon*. 2010; 38(4):679–686.
- Epps J, Smith JR, Wolfe J. A novel instrument to measure acoustic resonances of the vocal tract during phonation. *Meas Sci Technol*. 1997; 8(10):1112.
- Estill, J. Basic Figures and Exercises Manual. New York City, NY: Imagetech; 1992.
- Fant, G. Acoustic Theory of Speech Production. Mouton: The Hague; 1970.
- Lessac, A. The use and training of the human voice: a bio-dynamic approach to vocal life. Mayfield Pub; 1967. reprint 1996
- Miller, D. Resonance in Singing: Voice building through acoustic feedback. Princeton, NJ: Inside View Press; 2008.
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *J Acoust Soc Am*. 1952; 24(2):175–184.
- Stone RE Jr, Cleveland TF, Sundberg PJ, Prokop J. Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in a professional female singer. *J Voice*. 2003; 17(3):283–297. [PubMed: 14513952]
- Sundberg J, Gramming P, Lovetri J. Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing. *J Voice*. 1993; 7(4):301–310. [PubMed: 8293062]
- Sundberg J, Thalen M. Respiratory and acoustical differences between belt and neutral style of singing. *J Voice*. 2015; 19(4):418–425.
- Titze IR, Worley AS. Modeling source-filter interaction in belting and high-pitched operatic male singing. *J Acoust Soc Am*. 2009; 126(3):1530–1540. [PubMed: 19739766]
- Titze IR, Worley AS, Story BH. Source-vocal tract interaction in female operatic singing and theater belting. *J Singing*. 2011; 67(5):561–575.
- Titze IR, Baken RJ, Bozeman K, et al. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *J Acoust Soc Am*. 2015; 137(5):3005–3007. [PubMed: 25994732]

Figure 1.
 F_1 - F_2 chart for five selected speaking vowels, females left and males right.

Figure 2.
 F_1 - F_2 formant space for five averaged speech vowels with harmonic grid lines superimposed for the note G4 (392 Hz).

Figure 3.
 F_1 - F_2 formant space for the averaged specialized singing vowels with harmonic 299 grid lines superimposed for the note G4 (392 Hz).

Figure 4.
 F_1 - F_2 quadrangle for individuals with the smallest and the greatest FRP

Table I

Vowels used for three mouth shape categories

Mouth Shape	Neutral	Megaphone	Inverted Megaphone
Vowels	/i, e, a, o, u/	/æ, ʌ, ɔ/	/ɪ, e, ʌ, ʊ/