OVERLAP OF HEARING AND VOICING RANGES IN SINGING

Eric J. Hunter, PhD[+]

Ingo R. Titze, PhD[+*]

[+]The National Center for Voice and Speech

at The Denver Center for the Performing Arts,

Denver, Colorado

[*]The National Center for Voice and Speech

at Department of Speech Pathology and Audiology,

The University of Iowa; Iowa City, Iowa

CORRESPONDING AUTHOR:

Eric Hunter, PhD

National Center for Voice and Speech

The Denver Center for the Performing Arts

1245 Champa St; Denver, CO 80204; USA

Phone: 303-446-4834   Fax: 303-893-6487

Email address: ehunter@dcpa.org

Running head: Overlap of hearing and voicing ranges

Word Count: 2600; Table Count: 1; Figure Count: 3

INTRODUCTION

Trained singers adjust their vocal output to be heard in a variety of performance environments. In a study of a well-known operatic tenor, Sundberg [1] reported that the singer enhanced his vocal output between 3 and 4 kHz so that he could be heard above an orchestra; Sundberg called this enhancement the *singer's formant*. As the possible acoustical source of this enhancement, he later pointed to the clustering of Formants 3, 4 and 5 [2], which results from a separate resonator known as the epilaryngeal tube [3;4]. More recently, Oliveira-Barrichelo *et al*. [5] examined long-term average spectra to confirm that trained singers used the singer's formant more than untrained singers. Although it is likely that such singing enhancements are learned in order to exploit various auditory sensitivities, few studies have addressed specifically how singing enhancements relate to the auditory system or how the differences between the use of these enhancements by trained and untrained singers translate to the hearing range.

One study which shed light on the relation between vocalization and perception was conducted by Rossing *et al*. [6]. This study demonstrated that trained singers produce different frequency and dB SPL ranges in a choral setting verses in a solo setting; specifically, a trained performer uses a singer's formant in a solo setting but not in a choral setting. Although there may be many explanations for this observation (e.g., stylistic preference of overall choral sound), it is possible that the close proximity of trained singers in a choral setting could be a deterrent to using spectral enhancements like the singer's formant. The question of proximity between performers affecting vocal output could be approached by the overlap of the vocal ranges and auditory ranges, assuming room contributions to be of secondary importance.

Klingholz [7], a primary researcher in using the phonetogram or voice range profile (VRP) for voice classification, provided an important tutorial showing how to measure and interpret the

VRP of various singers. Although only used as a brief illustration, the notion of overlapping

Stimmfeld (voice field or VRP) and the *Hörfeld* (hearing field) was introduced. This idea

became the motivation for the current study, which also overlaps the VRP onto the hearing

range. However, because the VRP displays only sound levels and fundamental frequency, while

the voice contains many harmonics that can be perceived by the ear, this study also maps the

higher spectral components (overtones) of the VRP onto the hearing range.

Therefore, this study begins an investigation into the relationship between production

frequencies and perception frequencies. Specifically, the questions of interest were: 1) At a

distance of one meter, how near are trained singers' dB SPL levels to hearing thresholds? 2) At

what frequencies is the proximity between the level and the thresholds the greatest? 3) What

percentage of the human hearing range is used at 1 m and 30 cm? and 4) How does this use

change with different vowels?

METHODS

The current study examined the vocal output produced by four trained singers, one from

each of the categories of soprano, mezzo-soprano, tenor, and bass-baritone; four untrained

vocalists were used as controls. Voice range profiles (VRPs) were measured for all subjects. In

addition, third-octave band spectrum analysis (in dB SPL) was conducted on each subject's vocal

output and compared to data published for normal hearing by superimposing the output onto the

hearing range. To directly compare the human hearing range to the vocal range, all analysis was

done in dB SPL.

Subjects

Four professional vocalists trained in western opera and concert performance were

recruited: two male vocalists (a bass-baritone and a tenor, self-classified) and two female

vocalists (a soprano and a mezzo-soprano, self-classified) with an average age of 40. These

vocalists, from the faculty of the School of Music at the University of Iowa were selected for

their ability to produce performance-like sounds in an acoustically dead room (an anechoic

chamber). In addition, four untrained vocalists were recruited with similar voice classifications

as the professional singers: two male vocalists (a bass-baritone and a tenor, self-classified and

verified by fundamental frequency range judgment) and two female vocalists (a soprano and a

mezzo-soprano, again classified as stated above) with an average age of 31. The untrained

vocalists reported no formal training in singing or speaking performance; however, because the

study required that they be able to reproduce a pitch, all subjects had some informal singing

experience. All subjects reported normal hearing. Subjects were also asked about vocal history

and no vocal pathologies were reported. At the time of recording, all subjects reported they were

in good health.

Instrumentation

Acoustic recordings were conducted in an anechoic chamber at the Wendell Johnson

Speech and Hearing Center at the University of Iowa. The chamber was structurally isolated

from the main building and had fiberglass wedge sound treatment on the inner walls. The room

was a cube, with internal dimensions of 6.33 m to each side and a total free space volume of 253

m^3. The room was rated as anechoic for frequencies above 60 Hz.

The recording microphone (AKG Acoustics CK22, pressure gradient, C460B preamp:

20-20,000 Hz +/- 1dB) was mounted with a Quest Technologies Model 2700 sound-level meter,

whose microphone was positioned parallel to the recording microphone (approximately 5 cm

apart). Both microphones faced the sound source at a distance of 1 m, chosen because the AKG

microphone frequency response was known at this distance, which corresponded to a plausible

distance between singers in a vocal duet or members in a choir. Thus, measured loudness could

be related to these source-receiver distances. The AKG microphone signal was amplified from

microphone to line level with a Symmetrix Microphone preamplifier and recorded with a

Panasonic SV-3700 DAT Recorder at a 48 kHz sampling frequency.

Recording Session

Recorded vowels were based on the recommendation of Schutte and Seidner [8] for making

a VRP (i.e., subjects produce the vowels /i/, /a/, and /u/). Subjects were asked to produce each

vowel by singing it at their lowest and highest comfortable loudness at multiple pitches,

sustaining each production for at least 1.5 seconds. The order of the vowel sequence was unique

and randomly chosen for each of the eight subjects, with nine possible sequences.

A typical recording session for a subject lasted 30 to 40 minutes. Subjects were provided

with water and were given a break any time they felt it was necessary; no subject took more than

a five-minute break. For a given vowel, a subject was first asked to produce the vowel at a

comfortable pitch and loudness. Starting at the nearest keyboard whole note that matched this

comfortable pitch, the subjects were then asked to sing the vowel at their lowest and then their

highest loudness (*pp* and *ff*). A tone was played a whole step lower than the starting pitch, and

the subject was asked to match the pitch again and sustain the same vowel at low and high

loudness. This pattern was repeated with progressively lowered pitches, in whole step

increments, until the lowest producible pitch was reached. Next, the subject was asked to

produce a tone that was a whole step above the starting pitch as demonstrated by the keyboard,

followed by successively higher pitches until the upper limit of the subject's pitch range was

reached. This same procedure was then repeated for the remaining two vowels.

Before and after each recording session, a calibration tone was produced and presented through a single loud speaker at a distance of 1 m from the microphone at the position of the subject's mouth. This tone was recorded by the acquisition system (i.e., the microphones, amplifiers, and DAT deck described above). The maximum dB SPL of the tone was measured simultaneously using the sound level meter (fast response, linear weighting) for later calibration. This process was necessary because the sound level meter output was not recordable on the DAT deck in this study.

Analysis

Recorded tokens were played back from the DAT into the line input of a Larson-Davis System 824 analyzer. The analyzer, with settings identical to the sound level meter for dB SPL measurement, was controlled by a computer to obtain the maximum level of a production; simultaneously, the third-octave band levels at the time of this maximum were obtained. Token levels were converted to dB SPL using an intensity calibration constant, which was calculated from the calibration tone level (recorded by the analyzer) and the tone's known dB SPL (recorded by the sound level meter at the time of recording). One of these conversion constants was calculated for each subject's recording session.

Voice range profile (VRP) plots across all of the vowels were created by taking the maximum overall dB SPL for the loud rendition and the minimum overall dB SPL for the soft rendition at a given pitch. Spectral Level Profiles (SLPs), which contain all harmonic energy, were also produced. To obtain the SLPs, third-octave band spectral analysis was conducted on each rendition of a subject. The minimum level within individual third-octave bands during the soft renditions constituted the lower edge of the SLP.

Percentage of the hearing range area overlapped by the voice SLP area was calculated, where the overlap area was defined as a dB range multiplied by the respective third-octave bandwidth. In this study, the hearing range was defined as the spectral range between 40 and 16000 Hz and between 4.2 and 120 phon [9], approximately the thresholds of both audibility and discomfort (120 phon is an arbitrary boundary; yet, since it is a known standard, any other boundary can be computed). Given that the auditory system's pressure level sensitivity is related more to the dB domain rather than the linear pressure domain, calculating the overlap in terms of dB was a reasonable step. Although all recordings were made at a distance of 1 m, a second version of these percentages was calculated for the distance of 30 cm, representing the estimated nearest proximity between two singers in a choir performance.

RESULTS AND DISCUSSION

Voice range profile (VRP) plots were created for each subject (white areas in Figure 1; left column for the untrained singer, right column for the trained). Recall that the left-to-right spread of this small region represents the pitch range. Subjects produced an average VRP scale of 19 whole steps from low pitch to high. But now consider the range of all the harmonic frequencies in the tones. The maximum spectral level profile (SLP) across all vowels (diagonal cross-hatch) was plotted overlapping the hearing range (the SLP border, covered by the VRP, is depicted by a dashed line). Because the SLP area extended below the lower hearing threshold, only the area within the hearing range was considered when calculating the overlap percentage. Percentage overlap values are shown in the figures.

In general, the untrained vocalists' SLPs were in closest proximity to the discomfort level of hearing (top boarder) at approximately 600 Hz, 30 to 40 dB from the top. This is the spectral

region of F_1, the first formant frequency. While the trained vocalists' SLPs also had an F_1 peak,

they had a second peak that nearer to the discomfort level at 3150 Hz, with a difference of

between 20 to 25 dB. This is the spectral region of the singer's formant. Therefore, the singer's

formant provides not only a spectral boost, but is perceptually enhanced because of its proximity

to the discomfort level.

SLPs for individual vowels were created for each subject (Figure 2 shows data of only

the trained and untrained soprano). Percentage overlap of the individual vowel SLP on the

hearing range was calculated and listed in Table I. The average overlap percentages across

trained and untrained singers for the given vowels are listed in the table and plotted in Figure 3.

Some differences in the SLP of each vowel can be seen. In particular, three findings to note are:

1) the trained singers produced the /a/ vowel with the largest overlap percentage; 2) the untrained

singers produced the /i/ vowel with the largest overlap percentage (although the difference

between /a/ and /i/ was not significant); and 3) both sets of subjects produced the /u/ vowel with

the smallest overlap percentage. This implies that the /a/ vowel would be the loudest among

listeners, which is probably why it is often the preferred vowels of trained singers.

The second version of these overlap percentages (for the distance of 30 cm) was

calculated using the inverse square law, so that all dB SPL values would be increased by 10.4 dB

if the measuring distance were decreased to 30 cm. This increase (10.4 dB) was applied to all dB

SPL levels and resulted in an average overlap increase of 9 ($\pm$ 0.03) percent across all subjects.

Because of the limited number of subjects, further statistical analysis was not prudent.

Generally, the trained vocalists in this study produced a higher pitch at a higher intensity

level (from the VRP) with a larger overlap percentage than the untrained. There was little

difference in the maximum SLP data under 1000 Hz for the trained and untrained vocalists, with

the biggest difference occurring around the area of the singer's formant (3 to 5 kHz). However, the trained and untrained bass-baritone had the most comparable ranges (overall maximum level of approximately 89 dB SPL and overlap of 42.4 and 43.1 percent respectively) (Figure 1). Nevertheless, the differences in the SLP for these two subjects show that the trained bass-baritone had a spectral boost at the most sensitive region of the ear, likely causing a comparatively larger perceptual difference. In fact, using the singer's formant, all trained vocalists were able to produce a spectral boost in the ear's most sensitive region, a boost that was not seen in the corresponding untrained singers.

CONCLUSIONS

The trained voice has long been studied from a production point of view, but its close tie to perception has not been investigated abundantly. Sundberg [1] stated that the singer's formant was a necessary attribute to allow a singer to be distinguished from the spectrum of an orchestra. By comparing a singer's voice spectrum to the hearing thresholds, the current study suggests that the trained singer's voice is not only enhanced acoustically by the singer's formant, but also perceptually because it occurs at the most sensitive region in the auditory spectrum. Conversely, all vocalists appeared to be perceptually weak at the lower end of the spectrum, which is not only where there is less harmonic energy and where the ear is less sensitive. Therefore, singers have learned to maximize perception level of a listener by altering the level of spectral peaks.

In this study, the trained tenor had the highest spectral output, with a maximum third-octave level of 95 dB SPL at 1500 and 3000 Hz (105 dB SPL at 30 cm). As all of the trained singers had a similar spectral boost at the most sensitive region of the ear, the question of why a singer drops the singer's formant in choral settings [6] might be at least partially explained: it

would probably be perceptually unpleasant to be surrounded, in a choral setting, by several vocalists, each with a 95-105 dB SPL at the sensitive 3 to 4 kHz region. Note, however, that it is possible that a larger subject sample might yield different results.

This discussion of the audio-vocal system highlights an interesting question: Did the eventual training to produce the singer's formant occur specifically to exploit the sensitivity region of the ear? Expanding on this topic, further questions concerning the depth of connectivity between the auditory and vocal systems, especially in light of evolutionary processes, should include animal vocalization.

ACKNOWLEDGEMENT

Reference List

1. J. Sundberg, "An Articulatory Interpretation of the Singing Formant," *Speech Transmission Laboratory/Quarterly Progress and Status Report, Stockholm* 1 (1972): 45-53.

2. J. Sundberg, "Synthesis of Singing," *Swedish Journal of Musicology* 60, no. 1 (1978): 107-112.

3. J. Sundberg, "Articulatory Interpretation of the "Singing Formant"," *Journal of the Acoustical Society of America* 55, no. 4 (1974): 838-844.

4. I. R. Titze, "Acoustic Interpretation of Resonant Voice," *Journal of Voice* 15, no. 4 (2001): 519-528.

5. V Oliveira Barrichelo et al., "Comparison of Singer's Formant, Speaker's Ring, and LTA Spectrum Among Classical Singers and Untrained Normal Speakers," *Journal of Voice* 15, no. 3 (2001): 344-350.

6. T. D. Rossing, J. Sundberg, and S. Ternstrom, "Acoustic Comparison of Voice Use in Solo and Choir Singing," *Journal of the Acoustical Society of America* 79, no. 6 (1986): 1975-1981.

7. Fritz Klingholz, *The Voice Field: A practical guide for measurement and evaluation (in German)* (Munchen: Verlag J. Peperny, 1990).

8. H. K. Schutte and W. Seidner, "Recommendation by the Union of European Phoniatricians (UEP): Standardizing Voice Area Measurement/Phonetography," *Folia Phoniatr.(Basel)* 35, no. 6 (1983): 286-288.

9. ISO 226, "Acoustics -- Normal Equal-Loudness-Level Contours," *International Organization for Standardization* (2003).

Table I. Percentages of spectral level profile (SLP) overlap of the hearing range given for three vowels, as well as the overall maximum SLP across all vowels (as shown in Figure 3).

	Untrained				Trained			
	/i/	/a/	/u/	**overall**	/i/	/a/	/u/	**overall**
Soprano	36.9	35.3	31.4	37.8	36.8	43.5	40.1	44.0
Mezzo-Soprano	36.9	32.8	34.9	38.1	41.6	45.6	37.3	46.2
Tenor	35.2	29.9	23.2	35.8	45.4	43.0	42.9	46.4
Bass-Baritone	36.5	42.9	34.9	43.1	41.1	39.7	34.2	42.4
Average	36.4	35.2	31.1	38.7	41.2	42.9	38.6	44.8

FIGURE CAPTIONS

Figure 1. Hearing region with overall maximum spectral level profile (SLP) (diagonal cross-hatch) and voice range profile (VRP) (white area) plotted on top. The percentage of the hearing range overlapped by the spectral level profile was shown in a box.

Figure 2. Hearing region with individual vowel spectral level profile (SLP) for the (a) untrained and (b) trained soprano.

Figure 3. The average overlap percentage of the vocal and hearing range is shown for each vowel at 1 meter distance. The increase of all average overlap percentages if the ear-to-mouth distance were decreased from 1 m to 30 cm is 9%.

(a)

(b)