Overtone Focusing in Tuvan Throat Singing

• Christopher Bergevin (York University)
• Brad Story (University of Arizona)
• Joy Williams (York University)
• Jennifer Steeves (York University)
• Natasha Mhatre (Western University)
• Chandan Narayan (York University)

Background

Figure 1 – Feynman & Tuva (or bust!)

The otherworldly sound of Tuvan throat singing was in part introduced to Western audiences by Richard Feynman, who himself first heard the acoustically mesmerizing phenomenon in 1981 (while with his friend & biographer Ralph Leighton) having had received a record from fellow physicist Kip Thorne.

Tuva, a small Russian republic located in the geographic center of Asia (Fig.2), is renowned and celebrated for its unique style of song. There are deep cultural aspects to the music, such as *sound mimesis*.

Figure 2 – Tuva

A particularly striking feature of Tuvan singing, characteristic of the ***Sygyt*** style, can best be described as "*the simultaneous performance by one singer of a held pitch in the lower register and a melody (composed of overtones) in the higher register* [...] *similar at times to a cello playing ponticello*" [Aksenov, 1973].

It has been noted that throat singing does "*not involve any physiology unique to Turco-Mongol peoples; anybody can, given the effort, learn to throat-sing*" [Levin & Edgerton,1999].

However, while this salient vocal signature has been described impressionistically, the precise biomechanics and resulting acoustic characteristics remain less understood.

For example, Levin & Edgerton (1999) concluded that three components at play: ``*tuning a harmonic in the middle of a very narrow and sharply peaked formant; lengthening the closing phase of the opening-and-closing cycle of the vocal folds; and narrowing the range of frequencies over which the formant will affect harmonics*". However, little empirical justification was provided, leaving unclear the ``*question how this enormous enhancement or reinforcement in the spectrum works*" [Grawunder, 2009].

Figure 3 – Nonlinear phonation

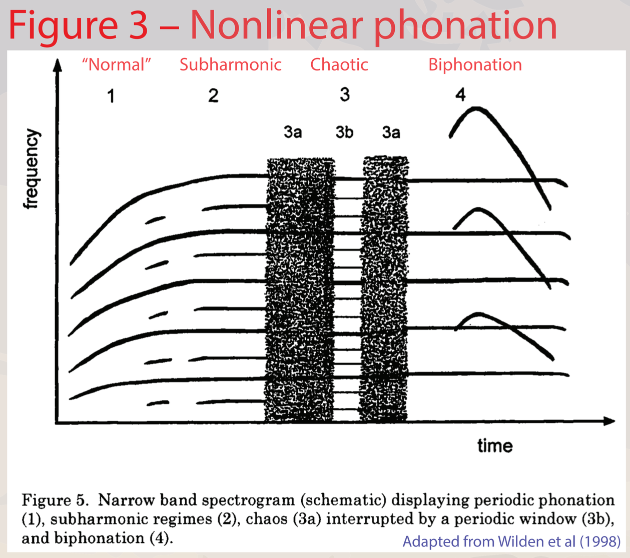


Figure 5. Narrow band spectrogram (schematic) displaying periodic phonation (1), subharmonic regimes (2), chaos (3a) interrupted by a periodic window (3b), and biphonation (4).
Adapted from Wilden et al (1998)

Further, there is evidence of an inherently nonlinear nature of vocal production (Fig.3), such as *biphonation*, motivating the question as to what role nonlinearities play in Tuvan song.

⇒ To address these questions, this study combined acoustical and magnetic resonance imaging data with theoretical modeling of vocal tract biomechanics. **We show how singers can merge resonances of their vocal tract to effectively focus certain overtones** (i.e., harmonics of their vocal fold vibrations).

Methods

Figure 4 – Data collection

We collected data from a group of Tuvan professional musicians (**Huun Huur Tu**; Fig.1).

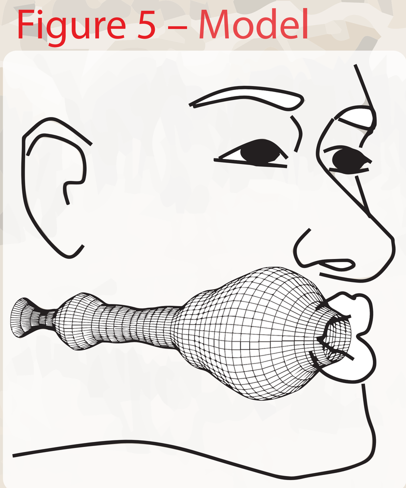
Two types of data were collected (Fig.4):
- *Sound booth recordings* (to characterize spectral properties & transitions into Sygyt-style song)
- *MRI* (while singing to determine vocal tract shape)

For the MRI, data were collected from only one singer (RT). Two methods were employed:
- *Dynamic* single-slice mid-sagittal 2-D recordings [Fig.11] at a rate of ~3.6 Hz)
- *Volumetric* 3-D recordings [Fig.12] while in steady-state Sygyt-style song

The scanned subject had a dental implant that affected the NMR signal. However, it was lateralized to the side (see Fig.12) and reliable 2-D and 3-D data could still be obtained.

Note: An ***overtone*** is generally defined as any frequency higher than the fundamental (f_0) of the source. Here we take overtone to be synonymous with those harmonic frequencies.

Figure 5 – Model



Adapted from Bunton et al (2013)

A modeling framework was employed that assumed 1-D wave propagation along the vocal tract [Story et al., 1996]. Thus the vocal tract is treated as a set on concatenated circular tubes of varying radius extending from larynx to the lips (Fig.5).

The key parametric input to the model is thereby an *area function*, which can be determined from either the 2-D* or 3-D MRI data [Figs.11&12].

* – Results here considered only the 2-D MRI data

Results

Figure 6 – Overtone focusing

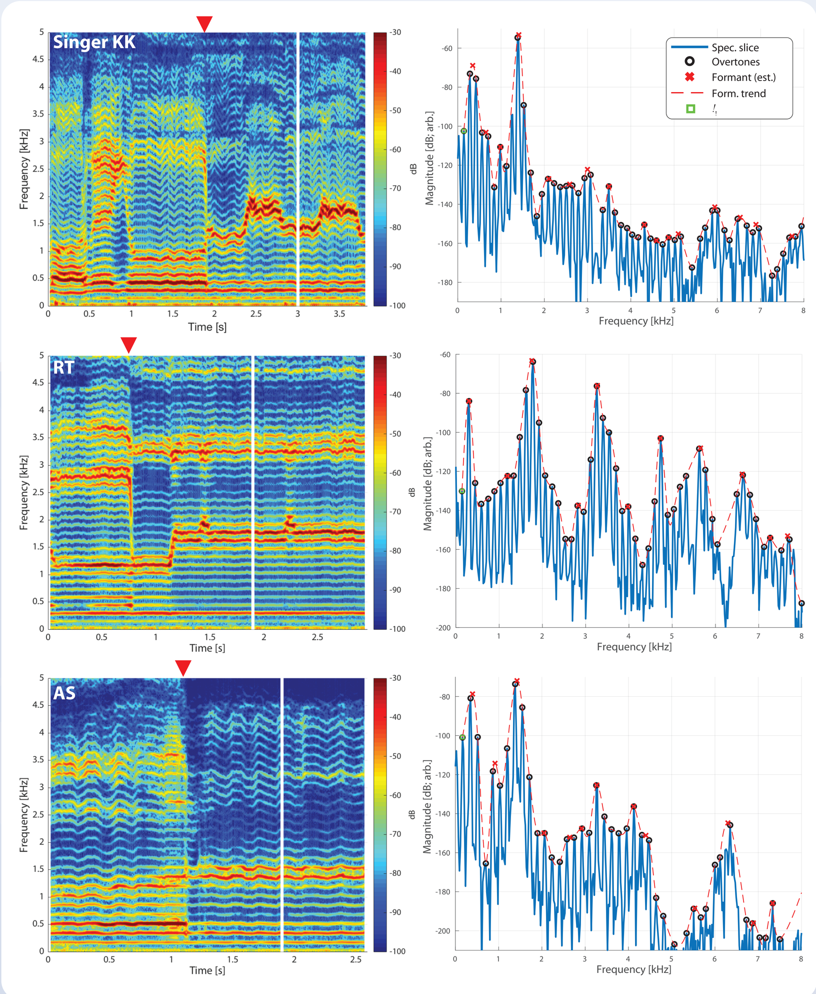


Figure 7 – Formant merging

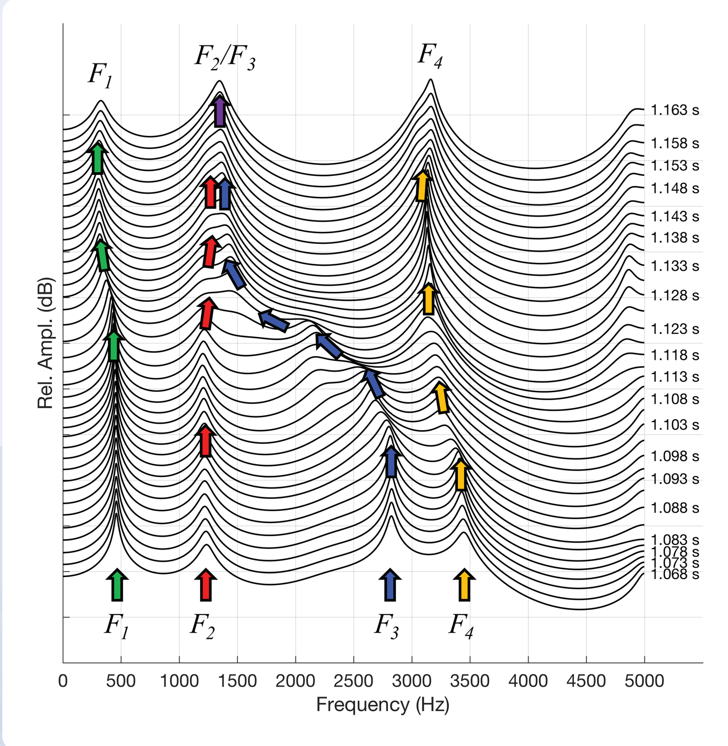


Figure 8 – Multiple focused states

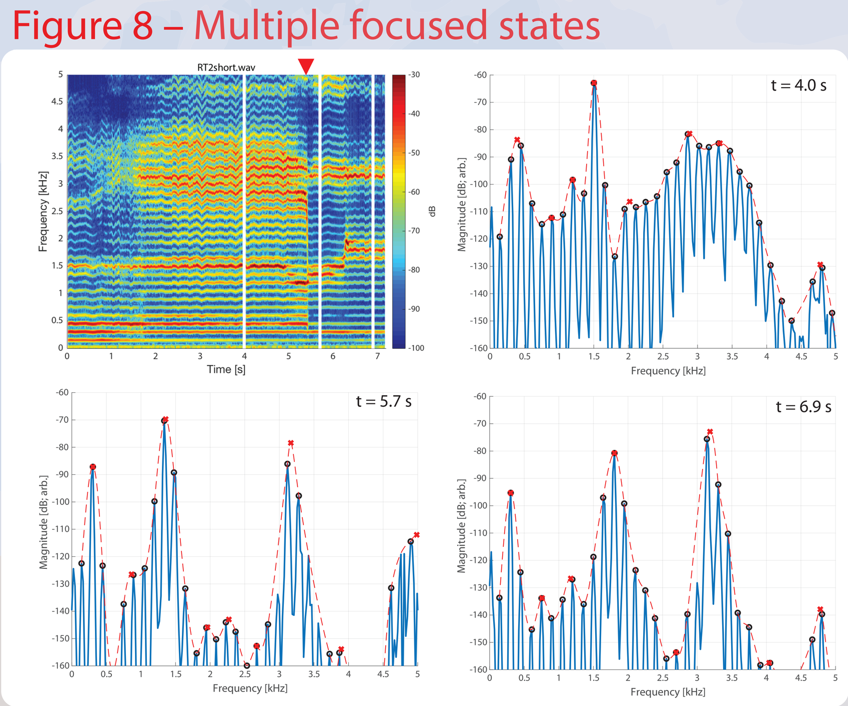
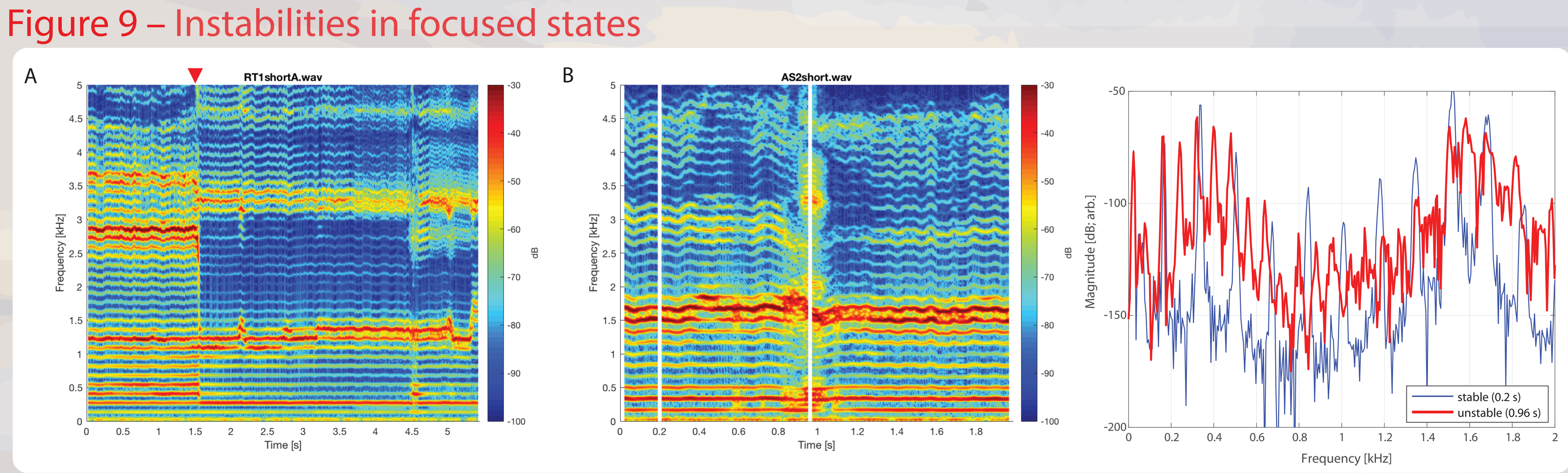


Figure 9 – Instabilities in focused states



Spectral Analysis

Overtone focusing – As shown in Fig.6, overtone structure appears as the horizontal bands and is relatively constant throughout the time courses. Note the lack of subharmonics. However, the formant structure (which appears as a superimposed color map) varies significantly in both time and frequency about the transition (▼). Once in the Sygyt-style, all three singers demonstrate that overtone energy is effectively ``focused" into a narrow spectral band(s). Not only is just a single (or small group of) overtones accentuated, but also that nearby ones are greatly attenuated.

Rapid transitions & Merged Formants – The transition into such is rapid (50-60 ms; Fig.6) and shows the merging of formants F2 & F3 (Fig.7)

Multiple focused states – One of the singers (RT) demonstrated multiple focused states. A second state was present at higher frequencies that was not explicitly dependent (e.g., harmonically) upon the first state (Fig.8). Note that a single sharply defined harmonic alone is not sufficient to get the salient perception of a focused state: It is not until the cluster of overtones about 3–3.5 kHz (Fig.8) is brought into focus that the perceptual effect becomes salient.

Instabilities – Several instances of brief transient instabilities in the focused state were observed (Fig.9), including frequency doubling (Fig.9 right). These observations suggest that the focus mechanism is actively maintained.

Characterizing the focused states – To quantitatively assess the degree of focus, we computed a dimension-less ratio (e_R ; see Fig.10) that characterizes the relative degree of energy brought into a narrow band spanned by $f_H - f_L$. For $[f_L, f_H]$=[1,2], e_R typically is small/negligible (i.e., energy is spread across the spectrum, not ``focused" into that narrow region). However for the Tuvan singers in the focused state, $e_R(1,2)$ is relatively larger (upwards of 0.5 and higher) and sustained across time. The situation was more complex for singer RT (Fig.10 middle), perhaps tied to the observation that he demonstrated multiple focused states (Fig.8).

Figure 10 – Quantifying focus

$$e_R(f_L, f_H) = \left(\frac{\int_{f_L}^{f_H} P(f)\, df}{\int_0^{f_{max}} P(f)\, df} \right)^2$$

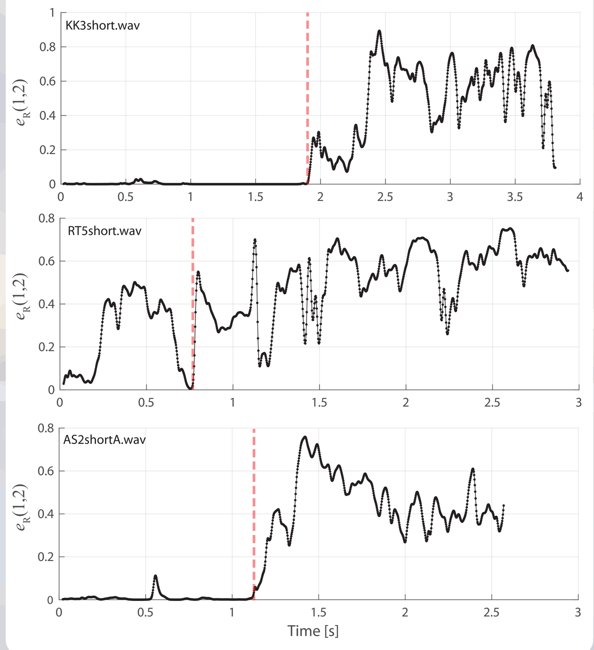


Figure 11 – Dynamic 2-D MRI

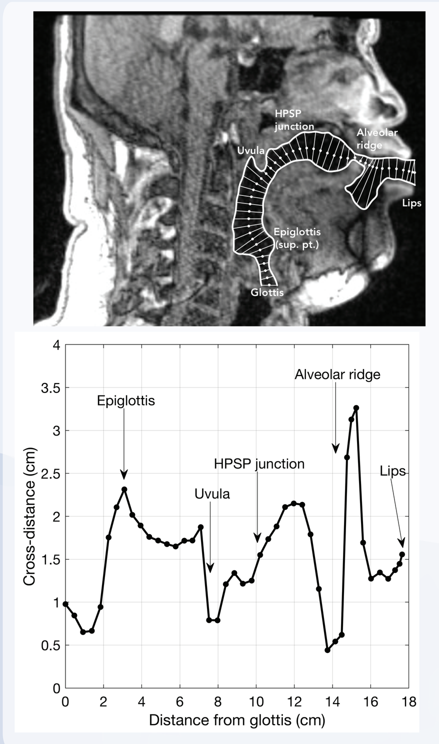
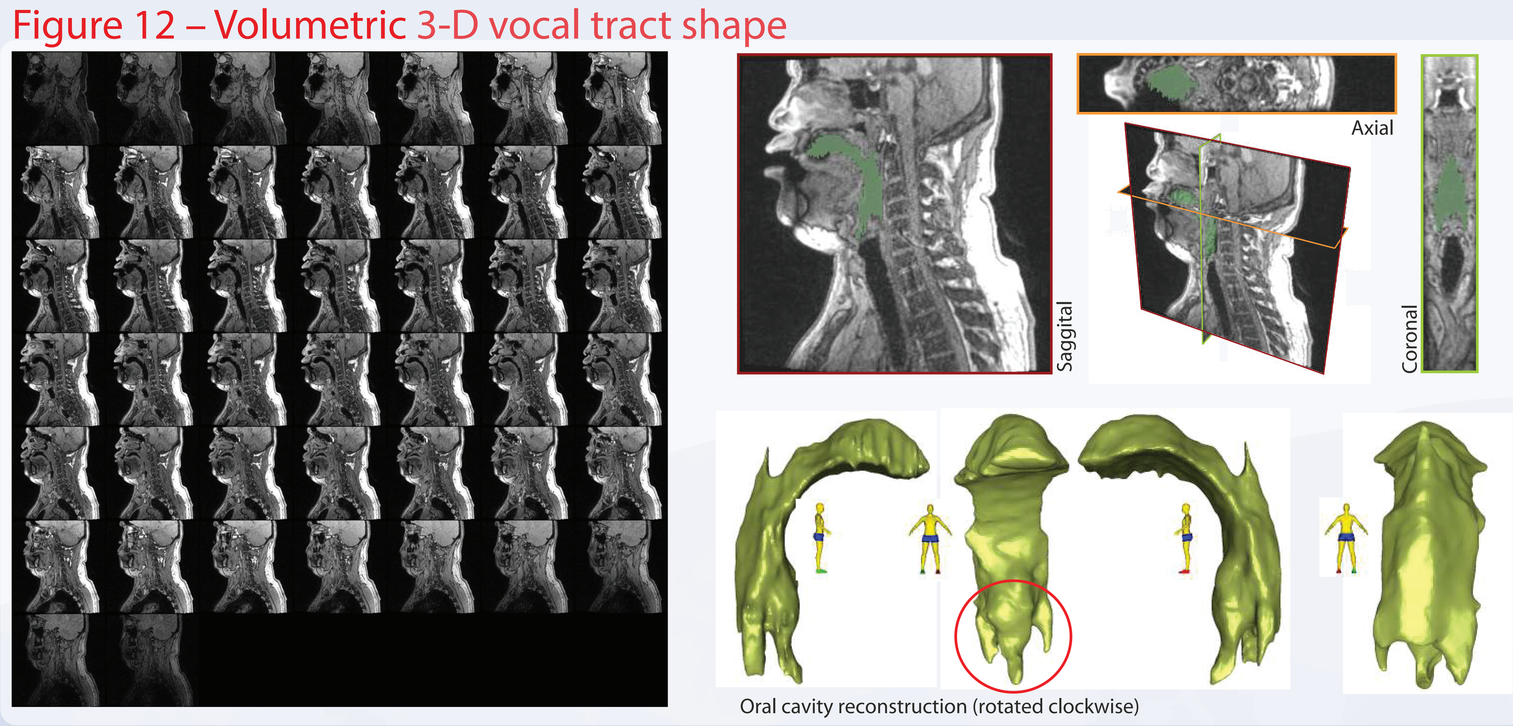


Figure 12 – Volumetric 3-D vocal tract shape



MRI

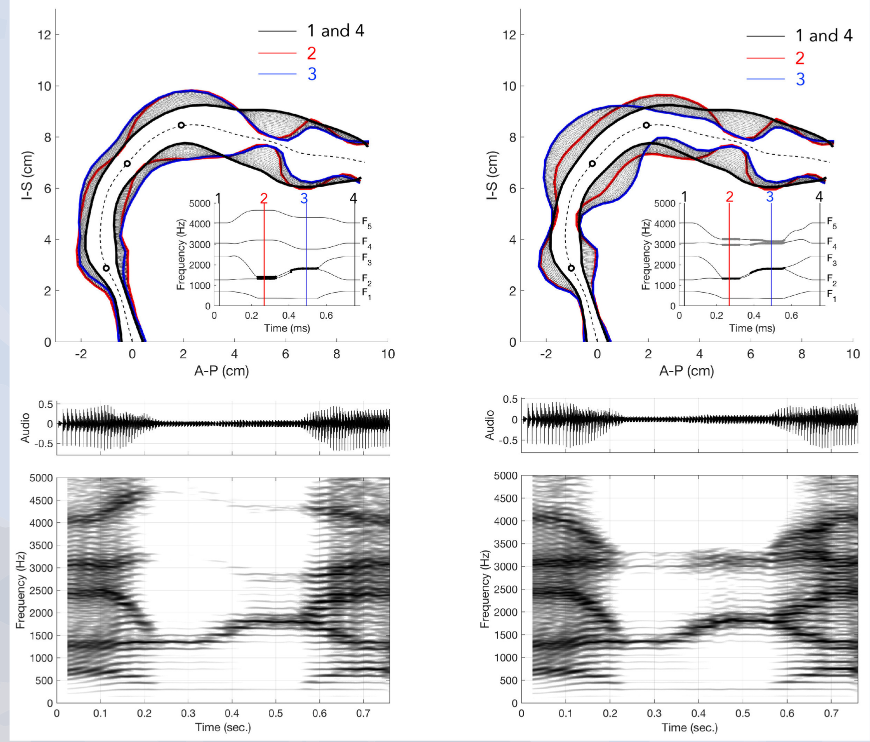
Dynamic 2-D – These data (including the associated sound spectra) allow us to examine changes heading into a focused state as well as how manipulations in song ``pitch" are achieved. There are two distinct constrictions located at 8 cm and 14 cm from glottis, respectively, and correspond roughly to the uvula and alveolar ridge (Fig.11). Additionally, the vocal tract is expanded in the region just anterior to the alveolar ridge. This occurs because the retroflex position of the tongue tip and blade that produces the constriction at 14 cm results in opening the sublingual space. It is the degree of constriction at these two locations that is hypothesized to be the primary mechanism for controlling the frequency at which an overtone is enhanced in the output signal.

Area functions – From the dynamic mid-sagittal frames (Fig.11 top), we can estimate the cross-sectional distance as a function of distance along the glottis (Fig.11 bottom). Area functions can then be determined (e.g., Fig.14 inset; see Modeling section).

Volumetric 3-D – Singer RT was able to sustain a steady-state Sygyt-style note to allow the 3-D data set to be collected (e.g., (Fig.12 left). The air-space in the oral cavity was manually segmented. Careful attention was paid to the parts of the oral cavity that were affected by shadow from the dental implant. The air cavity was manually repainted to be approximately symmetric in this affected part using the coronal and axial view (Fig.12 right).

Piriform sinuses – A potentially revealing feature is apparent in the volumetric data: the piriform sinuses (Fig.12 red circle in bottom right). These are small cavities just above the larynx and act as side-branch resonators [Dang & Honda, 1997]. It has been argued that they play a role in filtering properties of the vocal tract and thereby may be an important to facilitate overtone focusing.

Figure 13 – Modeling the dynamics



Modeling

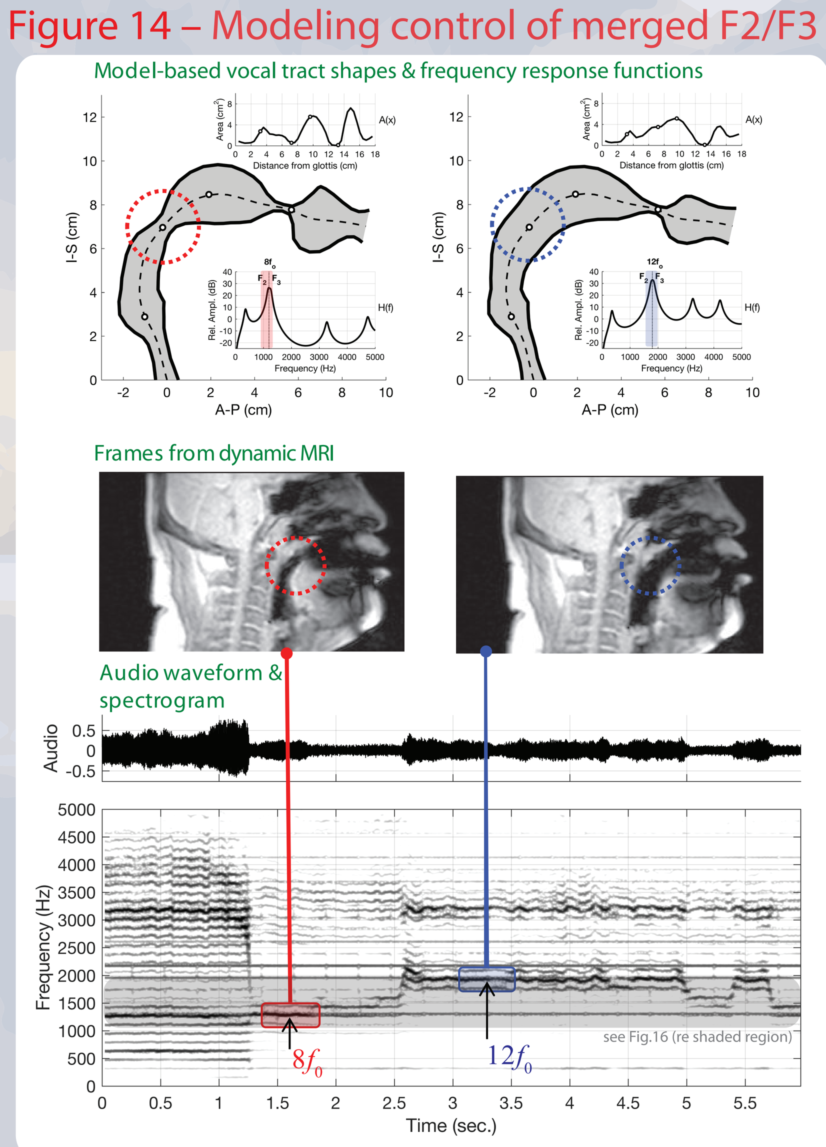
Focused state – We assume the singer configures the vocal tract to deliberately merge two formants (F2 and F3) together so that they enhance the amplitude of a selected harmonic of the voice source. Then the aim became to investigate how the vocal tract can be systematically shaped with precisely-placed constrictions and expansions to keep the formants in that focused cluster and move such along the frequency axis to allow for selection of a range of voiced (i.e., source) overtones (Fig.13).

Control – By first generating a tight constriction in the oral cavity near the alveolar ridge, and then modulating the degree of constriction in the uvular region of the upper pharynx, the model suggests how an overtone singer could potentially "play" (i.e., select) various harmonics (Fig.14).

Rapid transition – Further, the analysis suggests that small changes in the vocal tract are sufficient to make large changes in the frequency responses. Thereby, the short timescale of the transition in to/out of a focused state can plausibly be captured by the model (Fig.13).

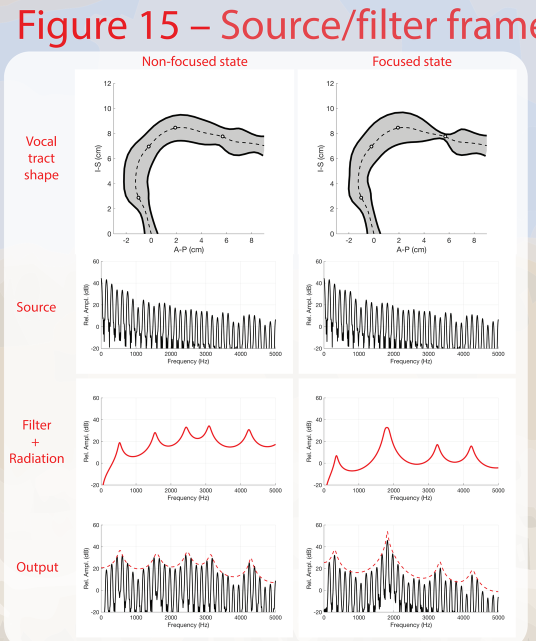
Multiple states – Additional variations in the back cavity were explored (Fig.13 right) to ascertain how a second higher-frequency focused state could be achieved (e.g., merging F4 and F5).

Figure 14 – Modeling control of merged F2/F3



Discussion

Figure 15 – Source/filter framework

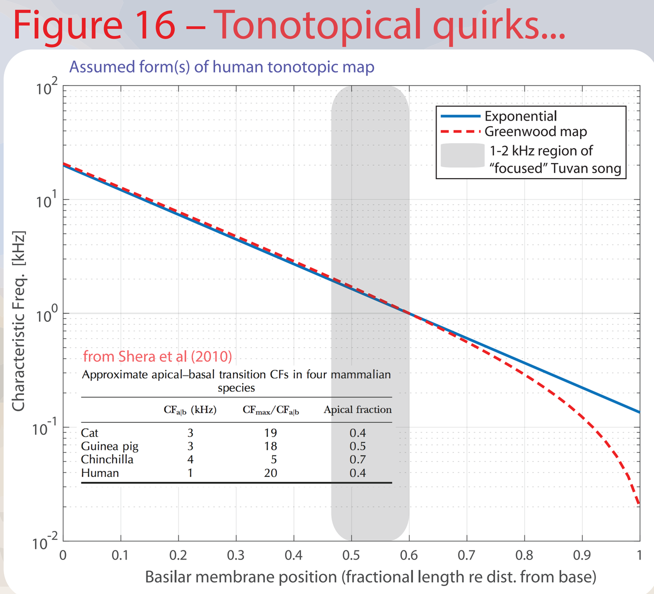


Linear vs Nonlinear?

In addition to nonlinear phonetic considerations, the sudden transition into a focused state (~50-60 ms, as evident in Fig.15) is suggestive of a nonlinear effect. However, acoustic analysis (e.g., lack of subharmonics, stable overtone structure) and ability of the (linear) model to well capture salient features of the data (Fig.13) argue for focused song to be an inherently linear process, consistent with the source/filter model (Fig.15).

However, we did observe instabilities that showed hints of nonlinearity (Fig.9), suggesting a secondary role that warrants further study/consideration (e.g., nonlinear source–filter coupling, Titze 2008)

Figure 16 – Tonotopical quirks...



Approximate apical–basal transition CFs in four mammalian species

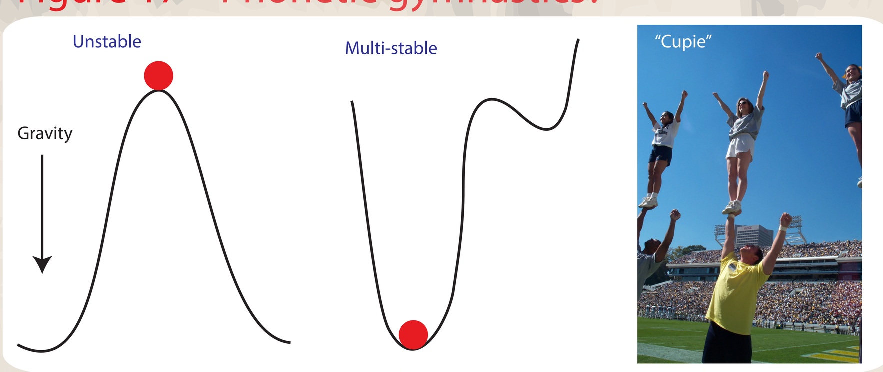
	$CF_{a|b}$ (kHz)	$CF_{max}/CF_{a|b}$	Apical fraction
Cat	3	19	0.4
Guinea pig	3	18	0.5
Chinchilla	4	5	0.7
Human	1	20	0.4

Exploiting a cochlear transition?

Ample evidence exists suggesting that there are differences in how the "basal" and "apical" portions of the cochlea process sound, as evident in the "Greenowood" map (Fig.16). Tuvan "Sygyt" song chiefly focuses energy into an octave band spanning 1-2 kHz, which sits right about that transition point.

We hypothesize that Tuvan song creates such a salient perceptual effect because it straddles these two regions. Thereby, Sygyt-style song indirectly capitalizes off a functional quirk of the cochlea.

Figure 17 – Phonetic gymnastics?



Biomechanical considerations

We argued that the focused state and sudden transition into such can be described within a linear framework. But those coupled with the brief instabilities hint the underlying motor control strategies a singer would need to learn/employ. We suggest two possibilities that a focused state represents:

- Unstable equilibrium that must be actively maintained through biomechanical feedback (left panel of Fig.17), akin to balancing a ruler vertically on your upturned palm
- Stable equilibrium in a less favorable energy configuration and more subject to perturbations (middle panel of Fig.17), a possible analogy being a "cupie" (a two-person move in cheerleading; right panel of Fig.17)

Such hypotheses could also help explain the importance of learning/practice in throat singing, and how some singers can achieve a more salient focused state than others.

We note that our study found inconsistencies with the mechanisms proposed by Levin & Edgerton. However, such may be a result of a multitude of strategies that singers might take to achieve a focused state.

Summary

Novel acoustic and MRI data underlying Tuvan throat singing were measured

"Sygyt-style" song is achieved by precise manipulation of vocal tract to *focus* overtones into narrow 1-2 kHz range

Analysis reveals that overtone focusing can be understood by linear source/filter model and is another means to achieve *biphonation* (i.e., "two notes produced simultaneously by a single singer")

Note: Digitized acoustic recordings can be accessed/listended to via: **www.yorku.ca/cberge/tuvan/**

References

• Adachi S, Yamada M (1999) J. Acoust. Soc. Am. 105(5):2920–2932.
• Aksenov AN (1973) Asian Music 4(2):7–18.
• Bloothooft G et al (1991) J. Acoust. Soc. Am. 92(4), 1827-1836.
• Bunton K, Story BH, & Titze I. (2013) Proc. Meet. Acoust. ICA2013. Vol. 19. No. 1.
• Dang J & Honda K (1997) J. Acoust. Soc. Am. 101(1), 456-465.
• Grawunder S (2009) *On the Physiology of Voice Production in South-Siberian Throat Singing: Analysis of Acoustic and Electrophysiological Evidences*. Frank & Timme GmbH
• Leighton R (2000) *Tuva or bust!* (WW Norton & Company)
• Levin & Edgerton (1999) Sci. Am. 281(3):80–87
• Shera CA, Guinan JJ, Oxenham AJ (2010) J. Assoc. Res. Otolaryngol. 11:343-355
• Stevens KN (2000) *Acoustic phonetics*. MIT press.
• Story BH, Titze IR, Hoffman EA (1996) J. Acoust. Soc. Am. 100(1):537–554.
• Titze IR (2008) J. Acoust. Soc. Am. 123(4), 1902-1915.
• Wilden I et al (1998) Bioacoustics, 9(3), 171-196.

Acknowledgements
Support provided by the Canadian National Science and Engineering Research Council (NSERC), the Fields Institute for Research in Mathematical Sciences, and the Kavli Institute for Theoretical Physics.

Biophysics

YORK UNIVERSITÉ UNIVERSITY