The mechanics and acoustics of the singing voice: registers, resonances and the source-filter interaction

Joe Wolfe, School of Physics, UNSW Sydney
Maëva Garnier, GIPSA-lab, Univ. Grenoble Alpes, CNRS, Grenoble, France
Nathalie Henrich Bernardoni, GIPSA-lab, Univ. Grenoble Alpes, CNRS, Grenoble, France
John Smith, School of Physics, UNSW Sydney

This is a pre-publication version

Abstract.

Singing combines speech and music—two activities common to most human cultures. It involves precise, coordinated control of breath, vocal folds, vocal tract, and their interactions. This chapter first introduces the interaction between laryngeal airflow and vocal-fold vibration that produces voiced sounds. It then discusses laryngeal mechanisms or registers, and how nonlinearities in the larynx produce the harmonically rich voice spectrum. Acoustical properties of the vocal tract and its resonances filter the voice spectrum, but also affect vocal fold vibration. Finally, it discusses how singers adjust their tracts using resonance tuning to produce louder, more stable voice with less effort.

Keywords [up to5]:

singing, laryngeal mechanism, register, vocal tract resonance, tuning

Singers of different sex and voice category cover a range of several octaves by using different laryngeal mechanisms to produce musical sound in the larynx. For expressive purposes, and also to be heard over accompaniment in large venues, they can also produce a large range of sound power – see Figure 8.1.

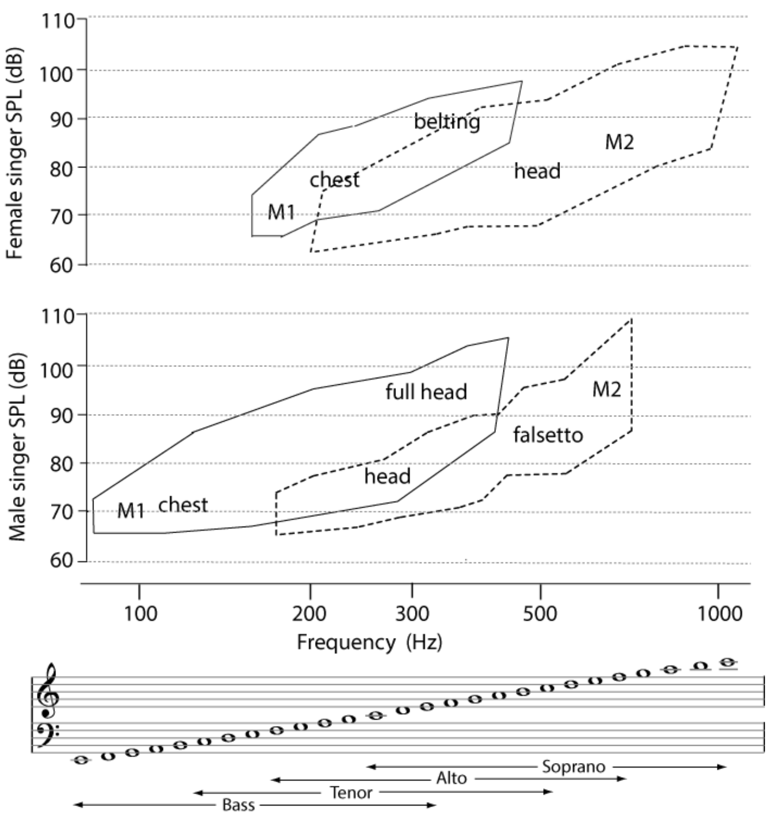

Figure 8.1. Pitch (horizontal) and sound level ranges (vertical) are exhibited in Voice Range Profiles. On each pitch, a singer exhibits his/her range of loudness (vertical scale) in each of the laryngeal mechanisms. Average data are sketched here for male and female classical singers (after Roubeau *et al.*, 2004). Laryngeal mechanisms (M1 and M2) and their transitions are explained later.

This chapter concentrates on the mechanics and acoustics of singing. We start by describing briefly how the high-pressure air from the lungs passes through the larynx, where it is modulated by the regularly vibrating vocal folds (sometimes misleadingly called 'vocal cords'). The pulses of air passing through the larynx produce sound waves in the vocal tract (and sometimes the nasal tract as well). These sound waves are modified by the acoustical response of the tract and some of their power is emitted from the lips as song.

These processes are often discussed separately, calling them respectively 'source' and 'filter'. The source at the larynx produces sound waves that comprise many different frequencies. Then the vocal tract, with its time-varying shape, filters this signal so that some bands of frequencies (called formants) are radiated from the lips with relatively high efficiency, while others are more attenuated. In practice, the source and the filter affect each other. We discuss in turn: source, filter and their interaction.

Basic science of voice production

Vocal fold vibration

The vibration of the vocal folds is not due to periodic muscular excitation at the vibration frequency. Rather, the flow of breath between them produces self-oscillation of the folds: vibration of the vocal folds modulates the airflow from the lungs, but this airflow is also the source of power for their vibratory motion.

Different physical effects are involved. These have been examined in detail in several physical models of voice production (*e.g*. Helmholtz, 1877; van den Berg, 1958; Flanagan and Landgraf, 1968; Awrejcewicz, 1990; Fletcher, 1993; Adachi and Yu, 2005) and good reviews are available (*e.g*. Titze, 2000; Zhang, 2016). Here, we introduce the basic principles.

During breathing with no phonation, the vocal folds are wide open, so the air flows readily from lips to lungs during inhalation and *vice versa* in exhalation. In producing a voiced sound, muscles attached to cartilages at the back of the vocal folds contract to pull them close together into the adducted or closed state. In this state, a pressure excess below the folds and flow between them can produce repeated cycles of opening and closing, which we call phonation. To a limited extent, the vocal fold vibration can be compared to the vibration of a mass on a spring, with a supply of energy necessary to maintain the vibration, compensating for energy losses in each cycle. Oscillations of the folds lose proportionally more mechanical energy per cycle than, say, a metal spring, so their vibration requires proportionally more energy input. How is energy from the steady, high-pressure airflow from the lungs transferred into the vibratory motion of the vocal folds?

One way uses the fluctuating pressure in the airflow between the folds. To accelerate air from low to high speed requires a gradient from high to low pressure. It follows that, in smoothly flowing air and neglecting other effects, the pressure is low where the speed is high and *vice versa*. Thus high-speed airflow between the folds creates a suction that acts to draw them together. If the phase of the variation in speed is appropriate, this variation in pressure can sustain oscillation (Flanagan and Landgraf, 1968; Zhang *et al*., 2006).

Another source of power involves vertical motion of the vocal folds. The turbulence of air downstream from the glottis has the result that the pressure is on average lower above (downstream from) the vocal folds than below. This pressure difference does positive work on the vocal folds when they move upwards and negative work when they descend. If the motion of the vocal folds were symmetrical in time, the total work around a cycle with steady pressure difference would be zero. However, observations (Baer, 1981; George *et al.*, 2008) indicate that the glottis is more nearly closed during ascent than during descent, which means that the average pressure difference can do significant positive work on the folds round each cycle (Boutin *et al*., 2015).

Pitch, loudness and timbre

All else equal, the natural frequency of a vibrating mass decreases with increasing mass. Hence it is not surprising that male voices, produced by larger larynges with longer and heavier vocal folds, have lower pitch than female. Similarly, an infection of the throat, which increases the fold vibrating mass, lowers the pitch. With given mass, their pitch can be varied by altering their length, tension and geometryThe vocal fold tension can be increased by stretching them and/or contracting the muscle inside, which also changes the mass distribution. All else equal, increasing subglottal pressure increases the power of the sound generated at the glottis. It also increases the frequency of vibration. Consequently, for a singer to control pitch and loudness independently, compensatory control is required, and a *messa di voce* (a slow *crescendo* and *decrescendo* at constant pitch) is a regular exercise for many singers.

The spectrum of the voice is one component of its timbre or voice quality. The spectrum depends on the subglottal pressure, the degree and nature of adduction of the vocal folds and their vibratory pattern. Both the power output and the spectrum depend strongly on the vocal tract acoustic response, as we shall see later.

Broadband sounds

Some speech sounds, called unvoiced, do not involve vibration of the vocal folds. In whispered speech, turbulence is produced by rapid airflow through a narrow glottis. In unvoiced fricatives (*e.g.* /sh/, /f/), the turbulence comes from a constriction formed using the tongue and hard palate or lip and teeth, respectively. In unvoiced stop consonants (/p/, /t/, /k/), the rapidly varying flow as well as turbulence is involved. Unvoiced sounds are used extensively in some styles, such as beatbox. However, because vowels are often prolonged in song, we shall concentrate chiefly on them.

The voice source spectrum

The vibration of a mass on a simple spring produces a wave with the shape of a sine function: its spectrum comprises a single frequency. (The frequency in hertz (Hz) is the number of vibrations per second: see Fig. 1.) In singing, vocal fold vibration is usually periodic, *i.e.* repeats itself after a defined period, but its motion and the waves so produced can have a more complicated shape than simple sine functions. A periodic wave with frequency f_o can always be represented as the sum of a set of harmonics, *i.e.* sine waves with frequencies f_o, $2f_o$, $3f_o$, $4f_o$ *etc*., plus a constant. The amplitudes and phases of these different harmonics are the spectrum of the wave. A sung note comprises several to dozens of audible harmonics. So a bass singing at f_o = 100 Hz (near G2, the G at the bottom of the bass clef) produces harmonics at 100 Hz, 200 Hz, 300 Hz *etc*.

The components with frequencies between several hundred and several thousand Hz are particularly important because human hearing is most sensitive in this range. This has an important implication for basses: much of the vocal power they generate falls between 80 and 300 Hz, and thus falls in a range where hearing is insensitive (and also where accompanying instruments are often powerful). At pitches low in the voice range, only the higher harmonics fall in the ear's sensitive range, so these harmonics are especially important.

Filtering by and radiation from the vocal tract: resonances and formants

The harmonic-rich signal generated at the larynx is very different from that radiated from the lips, largely because of resonances in the vocal tract, which acts as an acoustic duct. (Resonance refers to the capacity of a vibrating system to vibrate with larger amplitudes at a particular band of frequencies.) The tract is open at the lips and nearly closed at the larynx, with tract length ~17 cm for a typical man. As an approximation, we should expect it to behave very roughly like a 17 cm pipe, closed at one end and open at the other, with a baffle approximating the effect of the face on radiation from the lips. In fact, for the neutral vowel / /, detailed measurements of the acoustical properties of the tract are surprisingly similar to those of a simple cylinder (Hanna *et al*., 2016).

A uniform cylinder 17 cm long, closed at one end and open at the other, has resonant frequencies at about 500, 1500, 2500, 3500 Hz and beyond. These frequencies fall above the range of most men's voices, so a man might be surprised to observe that his trombone-range voice has a resonator of roughly piccolo length. Indeed for both sexes, the resonances usually lie above the fundamental frequency of the speaking voice, in which case the resonances chiefly affect the higher harmonics (see Figure 8.2).

At or near the resonances, a relatively small air vibration (at high pressure) at the closed end produces a much larger flow of air (at low pressure) out from the lips and into the acoustic environment: frequencies near these resonances are radiated very efficiently. In a consensus notation, the resonances of the tract are named $R1$, $R2$ *etc*., while the powerful bands of frequency in the output sound are formants $F1$, $F2$ *etc.* (ANSI, 2004; Titze *et al*., 2015).

The first several acoustic resonances of the tract usually fall within a few hundred Hz of those of the simple pipe with the same length, but their values depend strongly on the variable shape of the vocal tract. Oversimplifying, enlarging the mouth aperture (*e.g.* from 'oo' to 'ah') increases $R1$ and increases $R2$ proportionately less. Moving the position of the tongue constriction from the back to the front of the mouth ('oo' to 'ee') increases *R*2, with less effect on *R*1. *R*3 is influenced by lip spreading/rounding (Lindblom & Sundberg 1971). In any given language and accent, each vowel corresponds to a small range of values of (*R*1,*R*2) and sometimes *R*3. *R*4 and *R*5 have effects on the quality of the voice, which we discuss later. The role of resonances and formants in speech is described in texts on acoustical phonetics (*e.g*. Stevens, 2000; Clark *et al*., 2007; Ladefoged and Johnson, 2014). This topic benefits from sound recordings and multimedia, which we provide online (Music Acoustics, 2017a).

The linear source-filter model is shown in Figure 8.2. The glottal flow is taken from a published curve, which makes the assumption that there is no vertical fold motion. This flow is synthesized and input to a hardware model. The transfer function (in this case the transpedance) and the output pressure are then measured experimentally (Wolfe *et al*., 2016).

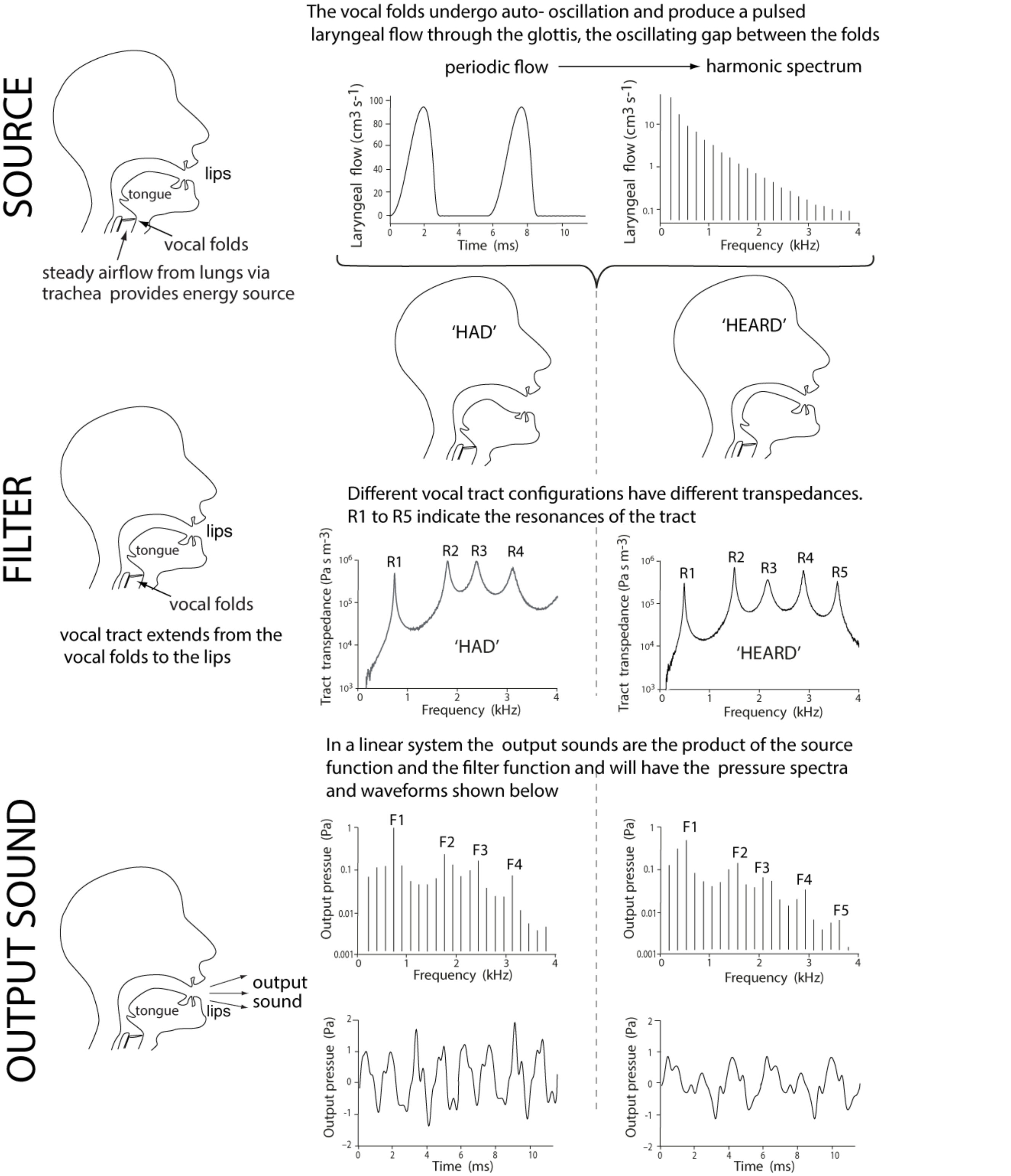

Figure 8.2. The source, filter response and output pressure, measured on hardware vocal tract models corresponding to the vowels in 'had' and 'heard'.

Some common methods of studying voice source, vocal-tract filter and their interactions

Laryngeal imaging techniques

The laryngoscope was developed by Babington to look through mouth and nose to the inside of the throat and used by anatomist and singing teacher Manuel Garcia (Castellengo, 2005). Since then, this mirror on a long handle has evolved into complex rigid or flexible endoscopes to examine the pharynx and larynx. This often provides a view of the vibrating vocal folds, though sometimes epiglottic tilt, aryepiglottic or vestibular constrictions obscure the view. Combined with stroboscopy or, more recently, high-speed digital imaging, this can monitor the lateral motion of the vocal folds and mucosal-wave displacement (Deliyski *et al*., 2008).

Electroglottography

Electroglottography is used to measure vocal-fold contact area non-invasively for clinical and scientific study and for singing-voice assessment, and biofeedback for singing training (Childers and Krishnamurthy, 1985; Colton and Conture, 1990; Henrich *et al*., 2004). The EGG (or laryngograph), uses electrodes mounted on the skin of the neck, either side of the larynx (Rothenberg and Mahshie, 1988). A very weak electric current at MHz frequencies, passed through the neck, is modulated by the vocal folds: more current flows when the folds are in contact.

Acoustic and aerodynamic measurements

The acoustic pressure signal can be measured with a microphone placed outside or at the lips. It can be used to estimate acoustical glottal flow by inverse filtering (Alku, 2011). The expired airflow (at lips and nostrils) can be measured using a mask placed on the face (Rothenberg, 1973). Aerodynamic pressure inside the mouth can be measured with a pressure sensor. The subglottal pressure can be directly measured by tracheal puncture in, or below, the crico-thyroid joint, or estimated from the mouth pressure on plosive-vowel syllabic trains. As subglottal pressure is a key parameter for voice dynamics, it is often used for singing-voice assessment.

Measuring and estimating resonances and formants

Formants – defined as bands of enhanced power in the sound (ANSI, 2004; Titze *et al*., 2015) – are present in recordings of the voice. Such spectral-energy enhancements are produced by resonances in the vocal tract. Within the framework of linear source-filter theory, linear predictive coding (LPC) is commonly used to estimate the vocal tract transfer function that underlies the resulting spectral envelope of the radiated voice (Atal and Hanauer, 1971). Strong assumptions are made about the glottal waveform, the acoustic radiation at lips and the number of poles of the vocal-tract filter (sometimes referred to as 'formants' in the speech processing community), in order to estimate resonance frequencies and bandwidths. The accuracy usually depends on the spacing between harmonics, *i*.e. f_o This may be sufficient for low pitch voices (~100 Hz) but not for high pitch.

Vocal tract resonances can be measured with greater precision by exciting the tract with a signal containing many frequencies. One method uses a shaker to excite the larynx mechanically (Sundberg, 1975; Djeradi *et al*., 1991), with the complication that the acoustic response of the neck tissues influences the results. The present authors inject broad band sound into the mouth using a small, flexible pipe at the lips, then measuring simultaneously the spectrum of the voice and the response of the tract to the injected sound (Epps *et al.*, 1997; Joliveau *et al*., 2004) – see Figure 8.3. Its chief application so far in singing research has been in studying resonance tuning, which is discussed below.

Alternatively, properties of the 'filter' (strictly the vocal-tract transfer function) can be derived from the tract geometry by solving the wave equation for sound in the tract, subject to the boundary conditions expected at its interface with tissues (Fant, 1980; Story *et al*., 1996; Perrier *et al*., 1992). This is either done in three dimensions by finite elements, or else by approximating the volume with an area function: a plot of cross-sectional area as a function of length. In the latter case, the duct is then approximated as a one-dimensional waveguide with varying cross-section. X-ray tomography and Magnetic Resonance Imaging can produce two- or three-dimensional images of the vocal apparatus.

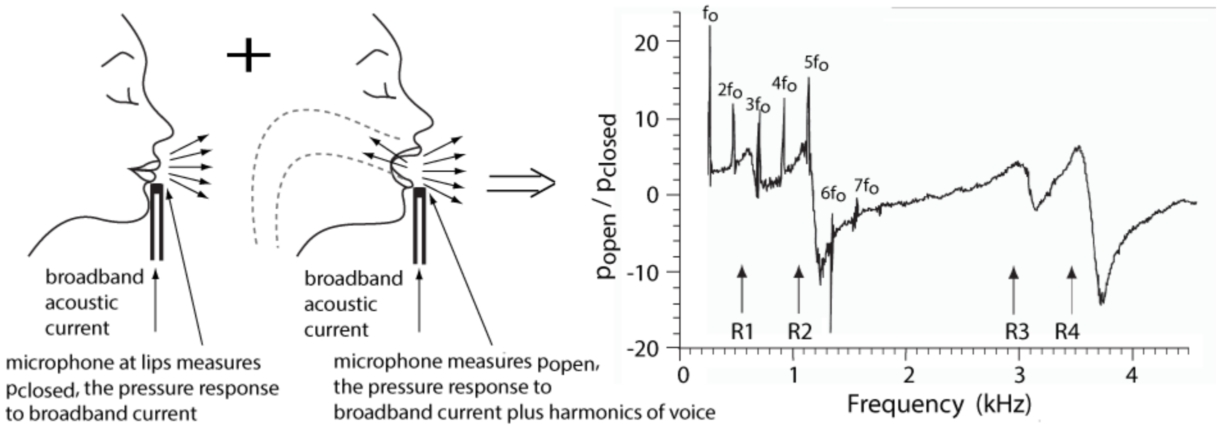

Figure 8.3. Measuring vocal tract resonances acoustically. The resonances appear as peaks in the quasi-continuous curve – the response of the tract to the injected current. Some of the harmonics appear above the curve. (Note how the fifth harmonic benefits from its proximity to the second resonance.)

Laryngeal mechanisms, singing-voice registers and transitions (*passagios*)

The large pitch range of singers is not achieved by continuous variation of the control variables: rather, there are qualitatively different laryngeal mechanisms or modes of operation of the vocal folds (Henrich, 2006; Roubeau *et al.,* 2009).

Mechanism 0 (M0)

This laryngeal mechanism enables very slow vibration, from a few Hz to ~80 Hz (Hollien *et al.*, 1977; Hirano, 1982; Blomgren *et al.*, 1998). The laryngeal intrinsic muscles are relaxed, leading to short and very thick vocal folds. In M0, the contact phase is long and the glottis opens very briefly. This voice production does not require high subglottal pressure, and thus it is commonly used in vocal warmup exercises prior to singing. Pulse, vocal fry or strohbass singing are produced using M0.

Mechanism 1 (M1)

In speech, M1 is the most commonly used mechanism, by both men and women. It is the main laryngeal mechanism used by classical bass, baritone, tenor and female alto voices. It enables production of the low to middle voice range. In M1, all the vocal fold layers (from superficial part of *lamina propria* to *vocalis* muscle) participate to the vibrating mass, inducing thick vocal folds and a vibration along the whole length. Both vibrating mass and vibration amplitude are larger than in M2. Often, the glottis is closed for more than half of its cycle. M1 is used by male and female singers singing in their chest-voice register, and by male singers for singing in head-voice register.

Mechanism 2 (M2)

This laryngeal mechanism is used by classical sopranos, mezzo-sopranos and male altos (counter-tenors), and by non-classical male and female singers, as in the female 'legit' quality of contemporary styles. It differs from M1 by a decoupling in the layered structure of vocal folds, resulting in stretched and thinner vocal folds with smaller vibratory amplitude and reduced contact phase (Hirano, 1982). The glottis is open for a larger fraction of each vibratory cycle than in M1, at least half of its duration. It is used to produce falsetto voices for male singers, and the female head voice.

Mechanism 3 (M3)

In M3, the vocal folds are even thinner and more stretched than in M2. The stretching effect forces them to vibrate with a much reduced amplitude and shorter glottal contact phase. Allowing production of very high pitches, it is often observed in children but is used only by a subset of adults, such as coloratura sopranos, jazz or pop female singers. It is equivalent to what is called bell, flute, whistle, or flageolet registers in singing.

Transition and overlap

There is considerable overlap in both pitch and loudness between M1 and M2 (see Figure 8.1), and between M2 and M3. Because of the pitch overlap, in which either mechanism may be used, singers can choose where in a musical phrase to make the change, or to make the change between phrases. For altos and tenors who use M2 as well as M1, 'managing the break' (or *primo passagio*) is a necessary art. It is interesting that, despite the differences in overall range, the *primo passagio* covers roughly the same range for both sexes (Fig. 1).

Sopranos who use M3 have the further complication that the range over which they make the transition between M2 and M3 (the *secondo passagio*) has some overlap with the range over which they may also change between two different resonance tuning strategies. This is discussed below.

Recent and current research on vocal tract adjustments in singing

While singers in some genres rely on microphones and amplifiers to produce voices loud enough to compete with amplified instruments, opera singers train so as to produce solo voices that can sometimes compete with a whole orchestra. Similar observations can be made of singers in other styles, and also of actors and orators. How is this possible?

One way to increase the power of the radiated voice, without increasing vocal effort at the glottal level, is to improve the acoustic radiation at the lips. This can often be achieved by increasing lip aperture and therefore the radiation area (Flanagan, 1960).

Another way to produce increased sound power for a given effort is to increase the amplitude of one (or more) of the voice harmonics by tuning the first or second vocal tract resonance frequency to that of the harmonic(s). The bandwidth of vocal tract resonances is typically 50 to 100 Hz (Hanna *et al.*, 2016). Consequently, low voices usually have harmonics falling close to each of the resonances and thus benefitting from the enhanced output power. For high voices, however, the relatively high f_o means large spacing between harmonics, and the possibility of missing out on the benefit of one or more tract resonances. Consider a soprano singing E5 at f_o = 650 Hz, with harmonics at 650, 1300 and 1950 Hz. If she articulates the vowel /ɜ/ with $R1$ and $R2$ at 500 and 1500 Hz, she misses the resonances completely. She could produce the same sound level with less effort if she were to adjust the vocal tract shape so as to tune $R1$ near 650 Hz (and perhaps also $R2$ to 1300 Hz). Some possibilities for resonance tuning are shown in Figure 8.4 and explained in the caption.

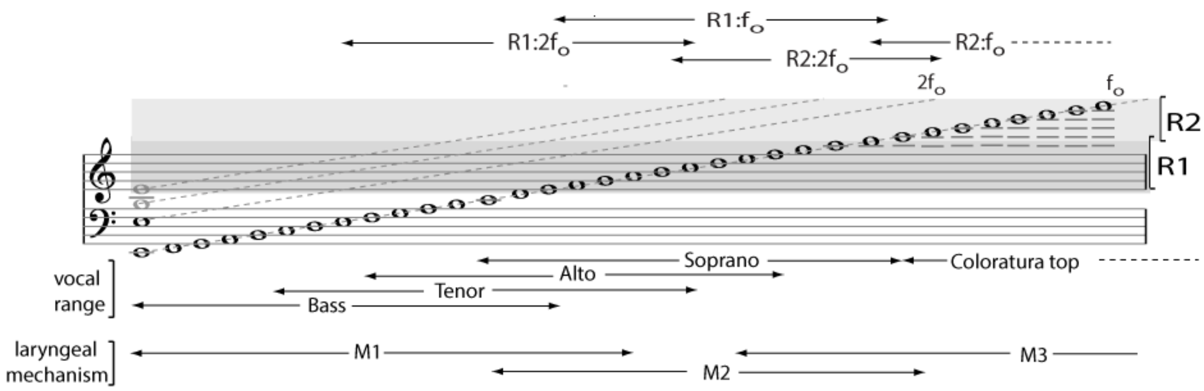

Figure 8.4. The two shaded bands represent typical frequency ranges of *R*1 and *R*2. The horizontal arrows at the top show some of the different strategies for tuning a resonance to a harmonic. For example, R2:2f_o means that the second resonance (*R*2) is tuned to the second harmonic (2f_o). The sloping dashed lines above the notes represent the pitch of voice harmonics (not equally spaced in the diagram because sung pitch goes as log *f*). Their overlap with the ranges of *R*1 and *R*2 determine the limits of the various tuning strategies.

Resonance tuning by sopranos

Sundberg (1975) measured the resonances of a professional soprano miming singing in the high range and observed $R1$ to increase to follow f_o. (Measurements were made during miming because the response to the measurement signals were much smaller than the voice signal.) This adjustment is employed whenever the pitch ascends above the normal range for $R1$ in speech: as pitch increases, singers increase the lip aperture and make other articulatory adjustments to keep $R1$ near the fundamental, f_o, now called $R1$:f_0 tuning. Joliveau *et al*. (2004) confirmed this by directly measuring resonances during singing. $R1$:f_0 tuning is used not only by professional singers but also by non-professional singers (Garnier *et al.*, 2010). It is used as high as E6-F#6 by some singers, but most stop at roughly C6, because it is difficult to open the mouth wide enough, or to make other adjustments that raise $R1:f_0$ above about 1 kHz. In this range, $R1:f_0$ tuning is sometimes accompanied by a complementary $R2:2f_0$ tuning (Henrich *et al.*, 2011).

Resonance tuning in the coloratura range and the M2/M3 transition

Sopranos who sang well above C6 used a different tuning strategy, tuning the second resonance to the fundamental ($R2:f_o$); they also used a voice quality and glottal behavior consistent with mechanism M3, the whistle register. Four of these singers produced an overlap range over which they could sing with either a 'full head' or 'fluty resonant' quality, with glottal behaviors suggesting M2 and M3 respectively. The possible mechanism and resonance tuning strategies available to such singers are shown in Figure 8.5.

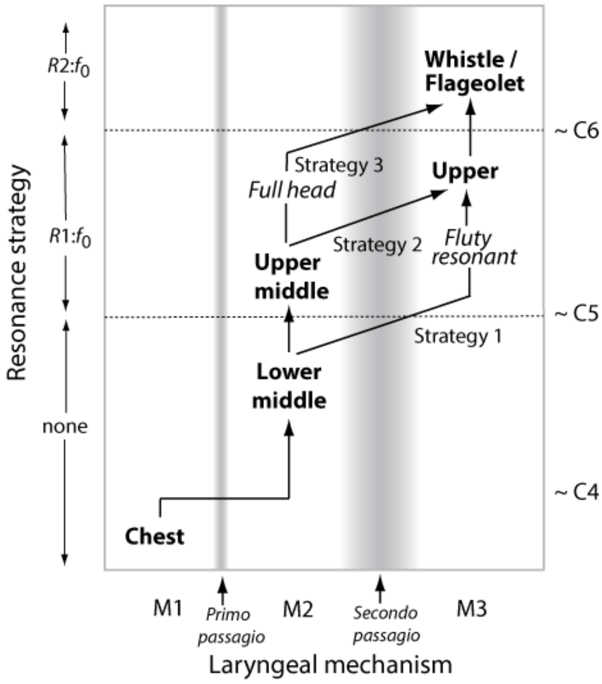

Figure 8.5. Some combinations of the M2-M3 mechanism change (secondo *passagio*) and resonance tuning (from $R1:f_o$ to $R2:f_o$) are available to sopranos in the high range.

Resonance tuning by other voice categories

Altos and sometimes tenors use $R1:f_0$ tuning in their high range, but for men it is observed only for the closed vowels (*e.g.* 'oo') that have low values of *R1* and that therefore fall in their singing range (Henrich *et al.*, 2011).

$R1:2f_o$ (tuning $R1$ to the second harmonic) is used by some altos. It is also used for 'belting' in music theatre (Bourne and Garnier, 2012), in Bulgarian folk singing (Henrich *et al.*, 2007) and shouting (Garnier *et al.*, 2008). Tenors and baritones generally use $R1:2f_o$ and $R1:3f_o$ tunings for sections of their range and various different tunings to higher harmonics at lower pitch (Henrich *et al.*, 2011). A further example of (possibly inadvertent) resonance tuning is shown in Figure 8.3, where the fifth harmonic and the second resonance coincide: $R2:5f_o$.

Consequences for vowel intelligibility in singing

As noted above, $R1$ and $R2$ determine $F1$ and $F2$ and thus the vowel. Resonance tuning shifts these values, sometimes far from the values implied by the vowel written in the score. Does this reduce comprehension? Audio examples in Music Acoustics (2017b) demonstrate the loss in a context with little information from consonants. Further, in the most consistent resonance tuning (that practiced in the high soprano range), the harmonics are so widely separated that there is little formant information to lose.

The singer's formant

Male singers trained in the Western classical tradition usually produce a strong formant in a range near 3 kHz, which is attributed to a clustering of the resonances $R3$, $R4$ and $R5$ (Sundberg, 1987, 2001). Such a resonance adjustment is associated with a lowering of the larynx to create a supraglottal constriction (Sundberg, 1974). This functions as an acoustic impedance matcher: it enhances power transmission from the glottis into the main tract. An important advantage of such a spectral enhancement is that it falls near the maximum sensitivity of the ear (~3 kHz) and above most of the main power band of an orchestra. Consequently, it increases perceived loudness and improves the identification of voice above the accompaniment. (High voices have less need of a singer's formant, because the fundamental and/or first few harmonics already fall in a sensitive range of the ear.)

Source-filter interaction

How can the filter affect the source?

A simple analogy cautions against the neglect of source-filter interaction. The vocal apparatus consists of two flaps of vibrating tissue separating resonant ducts upstream (the trachea) and downstream (the vocal tract). This invites analogies with say a trombone, in which two flaps of vibrating tissue (the player's lips) separate resonant ducts upstream (the vocal tract) and downstream (the instrument). The resonances of a trombone have a huge effect on the source: they largely determine the pitch. This is not the case for the voice: even when $R1$:f_o resonance tuning is practiced, the tuning is not usually as consistent as would be expected if the vocal folds were being driven at resonance. Further, for most of the vocal range of basses, tenors and altos, f_o lies well below $R1$.

Nevertheless, it is interesting to ask whether the large variation in acoustic load in the vicinity of resonances makes vocal fold vibration less stable or less controllable (Titze, 2008). The soprano voice is the obvious place to start, using pitch glides on a given vowel, so that f_o crosses $R1$, or slow diphthongs at constant pitch, so that $R1$ crosses f_o (Titze *et al*., 2008). Under conditions that minimize the natural tendency for some resonance tuning in pitch glides, vocal instabilities tend to occur above $R1$ (Wade *et al*., 2017). Further, when $R1$ was removed entirely (by singing into an acoustically infinite waveguide), instabilities were reduced and concentrated near the lower limit of M2, rather than at the typical frequencies where f_o crosses $R1$. So it appears that the varying acoustic load near a resonance can destabilize vocal fold vibration, though some singers are able to manage this perturbation.

The trachea is also a resonant duct, so one might also expect vocal instabilities when f_o crosses its lowest resonance at about 630 Hz (about D#5). No clustering of instabilities was found there, suggesting either that the upstream resonance does not disturb the voice, or that (because its frequency hardly varies) sopranos are accustomed to dealing with it.

How can the source affect the filter?

The vocal-tract resonances depend on the boundary conditions at tube entrance (glottis) and end (lips) (Barney *et al*., 2007). A small effective glottal area approximates a duct closed at the glottis, and the frequencies of the first few resonances increase with increasing effective glottal area. For different laryngeal mechanisms, the maximum glottal aperture varies, as does the fraction of a cycle for which it is open (Gauffin and Sundberg, 1989; Klatt and Klatt, 1990). Measurements in speech show that the first resonances occur at lower frequency in normal speech than in whispering (Kallail and Emanuel, 1984; Swerdlin *et al*., 2010), because whispering uses a larger glottal aperture (Solomon *et al*.,1989). Recently, systematic shifts in resonances were also measured for male operatic singers singing the same pitch in their chest or falsetto voices (Henrich *et al*., 2014).

Conclusion

Singing requires a skillful coordination of breath, vibrating laryngeal structures and articulation. Its complexity is reflected in the many years needed for training a professional voice. Even just controlling pitch and loudness independently requires subtle compensation. Managing the laryngeal mechanisms and the 'breaks' between them may be troublesome. For many singers, the various articulatory adjustments required for different modes of resonance tuning are also challenging. All this without discussing (for want of space) breath control and the artistic and phonetic constraints. As well as admiring singers for their artistry, we have several reasons to admire them for their vocal technique.

References

Adachi, S., & Yu, J. (2005). Two-dimensional model of vocal fold vibration for sound synthesis of

voice and soprano singing. Journal of the Acoustical Society of America. 117, 3213–3224.
Alku, P. (2011). Glottal inverse filtering analysis of human voice production—a review of estimation and parameterization methods of the glottal excitation and their applications. Sadhana, 36(5), 623-650.
ANSI (2004). ANSI S1.1-1994, *American National Standard Acoustical Terminology* (Acoustical Society of America, Melville, NY)
Atal, B. S., & Hanauer, S. L. (1971). Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. 50, 637-655
Awrejcewicz, J. (1990). Bifurcation portrait of the human vocal cord oscillations. J. Sound Vib. 136, 151–156
Baer, T. (1981). Observation of vocal fold vibration: Measurement of excised larynges," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (Univ. Tokyo Press, Tokyo, 1981), pp. 119–133
Barney, A., De Stefano, A., & Henrich, N. (2007). The effect of glottal opening on the acoustic response of the vocal tract. Acta Acustica united with Acustica, 93(6), 1046-1056
Blomgren, M., Chen, Y., Ng, M. L., & Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. J. Acoust. Soc. Am. 103, 2649-2658
Bourne, T. & Garnier, M. (2012) Physiological and acoustic characteristics of the female music theater voice J. Acoust. Soc. America, 131, 1586-1594
Boutin, H., Smith J. & Wolfe, J. (2015). Laryngeal flow due to longitudinal sweeping motion of the vocal folds and its contribution to auto-oscillation. J. Acoust. Soc. America, 138, 146-149
Castellengo, M. (2005). Manuel Garcia Jr: A clear-sighted observer of human voice production. *Logopedics Phoniatrics Vocology*, *30*(3-4), 163-170
Childers, D. G., & Krishnamurthy, A. K. (1985). A critical review of electroglottography. *Critical reviews in biomedical engineering*, *12*(2), 131-161
Clark, J., Yallop, C. & Fletcher, J. (2007). *An Introduction to Phonetics and Phonology*, Blackwell, Oxford
Colton, R. H., & Conture, E. G. (1990). Problems and pitfalls of electroglottography. *Journal of Voice*, *4*(1), 10-24
Deliyski, D. D., Petrushev, P. P., Bonilha, H. S., Gerlach, T. T., Martin-Harris, B., & Hillman, R. E. (2008). Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. *Folia Phoniatrica et Logopaedica*, *60*(1), 33-44
Djeradi, A., Guérin B., Badin, P. & Perrier, P. (1991). Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method. J. Phonetics, 19**,** 387–95
Doval B., d'Alessandro C. & Henrich N. (2006) The spectrum of glottal flow models, Acta Acustica united with Acustica, vol. 92, pp. 1026-1046
Epps, J., Smith, J.R. & Wolfe, J. (1997)A novel instrument to measure acoustic resonances of the vocal tract during speech. Measurement Science and Technology, 8, 1112-1121
Fant, G. (1980). The relations between area functions and the acoustic signal. *Phonetica*, *37*, 55-86.
Flanagan, J. L. (1960). Analog measurements of sound radiation from the mouth. J. Acoust. Soc. Am. 32, 1613-1620
Flanagan, J. L., & Landgraf, L. (1968). Self-oscillating source for vocal tract synthesizers. IEEE Trans. Audio Electroacoust. AU-16, 57–64
Fletcher, N.H. (1993). Autonomous vibration of simple pressure-controlled valves in gas flows. J. Acoust. Soc. Am. **93**, 2172-2180
Garnier, M., Henrich, N., Smith, J. and Wolfe, J. (2010) "Vocal tract adjustments in the high soprano range" J. Acoust. Soc. America. 127, 3771-3780
Garnier, M., Henrich, N., Crevier-Buchman, L., Vincent, C., Smith, J. & Wolfe, J. (2012). Glottal behavior in the high soprano range and the transition to the whistle register. *J. Acoust. Soc. America*. **131**, 951-962
Gauffin, J, & Sundberg, J (1989). Spectral correlates of glottal voice source waveform characteristics. J. Speech Hear. Res. 32, 556–565
George, N. A., de Mul, F. F. M., Qiu, Q., Rakhorst, G., & Schutte, H. K. (2008). Depth-kymography: high-speed calibrated 3D imaging of human vocal fold vibration dynamics. Phys. Med. Biol. 53, 2667–2675

Hanna, N., Smith J. & Wolfe, J. (2016). Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation. J. Acoust. Soc. Am. 139, 2924–2936

Helmholtz, H,L.F. (1877) *On the Sensations of Tone*, (trans A.J. Ellis, reprinted Dover, NY, 1954, 95-98

Henrich, N., d'Alessandro, C., Doval, B., Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. J. Acoust. Soc. Am. 115, 1321-1332

Henrich, D. N. (2006). Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. Logopedics Phoniatrics Vocology, 31, 3-14

Henrich, N., Kiek, M., Smith,. J. & Wolfe, J. (2007). Resonance strategies in Bulgarian women's singing: a pilot study. Logopedics Phoniatrics Vocology, 32, 171-177

Henrich, N., Smith, J., & Wolfe, J. (2011) "Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones" J. Acoust. Soc. America, 129, 1024-1035

Henrich Bernardoni, N., Smith, J., Wolfe, J. (2014). Vocal tract resonances in singing: Variation with laryngeal mechanism for male operatic singers in chest and falsetto registers. J. Acoust. Soc. Am. 135, 491-501

Hirano M. (1982) *The role of the layer structure of the vocal fold in register control*. Vox Humana: University of Jyvaskyla; pp. 50–62

Hollien H, Girard GT, Coleman RF. (1977) Vocal fold vibratory patterns of pulse register phonation. Folia Phoniatr. 29, 200-205

Joliveau, E., Smith, J. & Wolfe, J. (2004) "Vocal tract resonances in singing: the soprano voice", J. Acoust. Soc. Am. 116, 2434-2439

Kallail, KJ, & Emanuel, FW (1984). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. J. Speech Hear. Res. 27, 245–251

Klatt, DH, & Klatt, LC (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. 87(2), 820–857

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Nelson Education

Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. J. Acoust. Soc. Am., 50, 1166-1179

Music Acoustics (2017a) *Voice Acoustics*. http://newt.phys.unsw.edu.au/jw/speech.html Accessed 19/4/2017

Music Acoustics (2017b) *Sopranos: resonance tuning and vowel changes* http://newt.phys.unsw.edu.au/jw/soprane.html Accessed 19/4/2017

Perrier, P., Boë, L. J., & Sock, R. (1992) "Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients" *J. Speech, Language, and Hearing Research*, 35, 53-67

Rothenberg, M. & Mahshie, J.J. (1988). Monitoring vocal fold abduction through vocal fold contact area" J Speech Hear Res. 31: 338–51

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal airflow waveform during voicing. J. Acoust. Soc. Am., 53, 1632-1645

Roubeau, B., Castellengo, M. E. L., Bodin, P., & Ragot, M. (2004). Phonétogramme par registre laryngé. [Phonetograms for laryngeal registers] Folia Phoniatrica et Logopaedica, 56, 321-333

Roubeau, B., Henrich, N., & Castellengo, M. (2009). Laryngeal vibratory mechanisms: The notion of vocal register revisited. Journal of voice, *23*, 425-438

Solomon, NP, McCall, GN, Trosset, MW & Gray, WC (1989). Laryngeal configuration and constriction during two types of whispering. J. Speech Hear. Res. 32, 161–174

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press

Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. J. Acoust. Soc. Am. 100, 537-554

Sundberg, J. (1974). Articulatory interpretation of the 'singing formant'. J. Acoust. Soc. Am. 55, 838-844

Sundberg, J. (1975). Formant technique in a professional female singer. Acustica, 32, 89-96

Sundberg, J. (1987). *The science of the singing voice*. Northern Illinois University Press, DeKalb, Ill.

Sundberg, J. (2001). Level and centre frequency of the singer's formant. J. Voice 15, 176-186.

Swerdlin, Y., Smith, J. and Wolfe, J. (2010). The effect of whisper and creak vocal mechanisms on vocal tract resonances J. Acoust. Soc. America. 127, 2590-2598

Titze, I.R. (2000) *Principles of Voice Production*. National Center for Voice and Speech, Iowa

Titze, I. (2008) Nonlinear source-filter coupling in phonation: theory. J. Acoust. Soc. Am. 123, 2733–2749

Titze, I. R., Riede, T, and Popolo, P. (2008). Nonlinear source-filter coupling in phonation: vocal exercises J. Acoust. Soc. Am. 123, 1902

Titze, I.R., Baken, R.J. Bozeman, K.W., Granqvist, S. Henrich, N., Herbst, C.T., … Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. J. Acoust. Soc. America. 137, 3005-3007

van den Berg, J. W. (1958). Myoelastic-aerodynamic theory of voice production. J. Speech and Hearing Research , 1, 227–244

Wade, L. Hanna, N., Smith J. and Wolfe, J. (2017). The role of vocal tract and subglottal resonances in producing vocal instabilities. J. Acoust. Soc. America, 141, 1546-1559

Wolfe, J., Chu, D., Chen, J.-M. & Smith J. (2016). An experimentally measured Source-Filter model: glottal flow, vocal tract gain and radiated sound from a physical model. Acoust. Australia, 44, 187-191

Zhang, Z., Neubauer, J., & Berry, D. A. (2006). The influence of sub-glottal acoustics on laboratory models of phonation. J. Acoust. Soc. Am. 120(3), 1558–1569

Zhang, Z. (2016). Mechanics of human voice production and control. J. Acoust. Soc. Am., 140, 2614–2635

Bios

Joe Wolfe is a professor in the School of Physics, UNSW Sydney, leading a laboratory devoted to the fundamental physics of the voice and of musical instruments.

Maëva Garnier is a speech scientist at the French National Center for Scientific Research (CNRS), conducting research on the notion of vocal effort at the Speech and Cognition Department of GIPSA-lab at Grenoble, France.

Nathalie Henrich Bernardoni is a voice scientist at the French National Center for Scientific Research (CNRS), having also a personal practice as a singer and choir conductor. She heads the Speech and Cognition Department of GIPSA-lab at Grenoble, France.

John Smith is an Honorary Associate Professor in the School of Physics at UNSW, where he currently researches the acoustics of musical instruments, the human voice and how they interact.

Glossary

Adduction:	The bringing together of the vocal folds to close the aperture between them (the glottis)
Chest voice:	A singing-voice register commonly used by male and female singers to sing in the lower part of their vocal range
Cricothyroid joint:	joint connecting cricoid and thyroid cartilages
Electroglottograph:	A measuring device that passes a small, high frequency electric current between two electrodes attached to either side of the neck at the level of the larynx. Variations in the electrical conduction give information about the varying degree of vocal fold contact
Endoscope:	An optical device that allows direct observation and filming of the larynx
Filter:	A system that responds with different amplitudes to different frequencies, attenuating some frequencies more than others
Formant:	A prominent band of frequencies in a sound spectrum
Glottis:	The aperture between the two vocal folds, though which air can pass
Head voice:	A singing-voice register mainly used by female singers and few male singers to sing in the upper part of their vocal range
Hertz, Hz:	The standard unit of frequency. One hertz is one vibration per second
Kilohertz, kHz:	One kHz = 1000 Hz. See 'hertz'
Laryngeal mechanism:	Biomechanical configuration of the laryngeal vibrator, related to specific adjustments in the microstructure of the vocal folds, that allows the production of the human voice over different

frequencies. Four laryngeal mechanisms have been identified in the male and female adult voice from the bottom to the top of their vocal range:

- Laryngeal mechanism M0 uses very thick and relaxed vocal folds, with minimal muscular activity of vocalis and crico-thyroid muscles. This mechanism is used to produce the lowest frequencies from a few Hz to less than about 100 Hz
- Laryngeal mechanism M1 uses thick vocal folds with increased muscular activity of vocalis. All layers within each fold are coupled, implying a relatively large vibrating mass over the fold's thickness
- Laryngeal mechanism M2 uses thin vocal folds with increased muscular activity of the crico-thyroid muscles. The deep layer (vocalis muscle and ligament) of each vocal fold is decoupled from its superficial layer, making it possible to sing at high pitches with reduced vibrating mass.
- Laryngeal mechanism M3 uses thin and vocal folds under high tension. This mechanism, if available, allows production of the highest frequencies of the a vocal range.

Laryngoscope: An optical device that allows direct observation and filming of the larynx (see also Endoscope)

Modal voice: The most commonly used speaking-voice register (in contrast to whispering or creaky voice)

Phonation: The process whereby a flow of air between the vocal folds causes them and nearby tissues to vibrate, which in turn modulates the flow of air and creates an audible sound

Primo passaggio: In ascending pitch, the first transition that occurs in a vocal range, related to a coupling/decoupling of vocal fold layers, and corresponding to the transition between M1 and M2 laryngeal mechanisms (typically around D4-E4 in both men and women)

Register: A voice quality observed over a range of consecutive pitches

Resonance: Some vibratory systems have a large vibration response at a particular frequency. These are called resonances. In the vocal tract, the resonances are due to standing waves: constructive interference due to sound waves travelling in opposite directions.

Resonance tuning: The act of adjusting one or more of the vocal tract resonances so that it is close to one of the harmonics of the voice

Secundo passaggio: In ascending pitch, a second transition that occurs in the (female) vocal range, typically around E5.

Singer's formant: A band of relatively high power in the voice spectrum around 3 kHz.

Sound spectrum: A measurement of sound pressure or sound power as a function of frequency

Source-filter model: A description of voice production as being due to a voice source (the production of a harmonically rich sound at the larynx) whose spectrum is then filtered by the acoustical response of the vocal tract and radiated to the open space at lips (and nostrils)

Trachea: The respiratory duct below the larynx that links it to the lungs

Transpedance: The ratio of the sound pressure at the lips to the input acoustic flow.

Vocal folds: Multi-layered fibrous tissue within the larynx, anchored on the thyroid cartilage at the front of the neck and the arytenoid cartilages at the back

Vocal tract: All cavities located between the glottal plane and the lips/nostrils

Wave equation for sound: An equation that relates the variation of sound pressure in time with its variation in space. Its solution usually describes a sound wave.