PROCEEDINGS of ISMA 2019
International Symposium on Music Acoustics

13 - 17 September 2019 in Detmold, Germany

A real-time feedback system for vocal tract tuning

Marie JEANNETEAU[1], Noel HANNA[1], John SMITH[1], Joe WOLFE[1]

[1] School of Physics, UNSW Sydney, Australia

Abstract

The first resonance of the vocal tract (R1, an impedance minimum when measured at the lips) is tuned by some singers: all of the sopranos we have studied, whether trained or not, tune R1 to f_o over the range roughly C5 to C6; altos and tenors sometimes tune it to one of the lower harmonics in their upper range. Sopranos who sing in the range substantially above about C6 tune the second resonance R2 to f_o. The first purely acoustic impedance maximum is tuned by saxophonists and clarinettists under some conditions. This paper introduces a public version of software that can be used in a system to teach vocal tract tuning using a nearly real-time display and a graphical user interface. Our first use of it was to teach R2: f_o tuning to sopranos. In a one-hour session, most subjects learned tuning over a limited range, but only while the interface provided visual feedback. One of the authors (MJ) spent several weeks with the system and learned to extend her range of R2 variation and could use R2: f_o tuning over from F5 to E6, with and without the visual feedback. Training resonance tuning on saxophone is also proposed.

Keywords: Resonance tuning, Vocal tract, Feedback system

1 INTRODUCTION

In speech, the fundamental frequency f_o of the vibrating vocal folds, and hence that of the voice, is typically ~100-300 Hz and usually falls below the range of the resonances of the vocal tract, which typically lie above about 300 Hz. Further, the frequency response of the vocal tract is controlled independently of f_o and is used for a largely unrelated phonetic purpose: the first two resonances, R1 and R2, produce formants F1 and F2, which are broad peaks in the spectral envelope. Regions on the (F2,F1) plane identify vowels, and the trajectories F1(t) and F2(t) contribute to identifying consonants [*e.g.* 1].

In performance on wind musical instruments, it is possible to achieve a degree of competence without adjusting vocal tract resonances as a function of f_o, (which here means the fundamental frequency of the note being played). This is not surprising: With regard to the pressure difference acting across a reed or a brass player's lips, the acoustic impedances of the bore and the vocal tract are in series. Usually, the impedance of the bore of reed or wind instruments is rather greater than that of the vocal tract, so that the latter has only a modest effect.

Nevertheless, there are important circumstances in which both singers and some wind instrumentalists tune the vocal tract, as we explain below. In some cases, the tuning is counter-intuitive. In others, there may be no simple auditory feedback. For these reasons, visual feedback may be helpful in learning the art of resonance tuning. For that purpose, high precision in the feedback is not necessarily required, and so techniques rather simpler than those used in an acoustics research lab may suffice. For that reason, we decided to make a public version of software that, with relatively simple and easily obtained hardware, provides feedback in nearly real time of the frequencies of vocal tract resonances. This paper first introduces resonance tuning, then describes the new, simple system, then summarises the results of a preliminary study using it.

1.1 Resonance tuning in singing

In normal nomenclature, tract resonances correspond to resonant modes having a pressure node near the open lips and a pressure antinode near the larynx: such a resonance acts as an effective impedance matcher from the high impedance at the glottis to the low impedance of the radiation field at the lips. Resonance R1 has no pressure node between the two ends, R2 has one and Rn has $n-1$.

The normal range of the soprano voice, from about C4 (middle C) to nearly C6 (a soprano's 'high C'), covers approximately the range of R1. A soprano who varied R1 independently of f_o (in other words, one who allowed the librettist rather than the composer to specify the vocal tract shape) would often pay the penalty of losing the

[1] jeanneteau.marie@gmail.com, N.Hanna@unswglobal.unsw.edu.au, john.smith@unsw.edu.au, j.wolfe@unsw.edu.au

loudness boost provided by the impedance matching possibility of R1, and perhaps also risk vocal instabilities because of the phase of the acoustic load at the larynx [2].
Observations of sopranos [3,4] and direct measurements of their resonance [5] showed that, over this range, they tune R1 very near to f_o: what we call R1:f_o tuning. Untrained as well as trained sopranos use this technique: the auditory feedback of producing a louder sound with less effort is apparently enough for singers to learn it. Altos and tenors sometimes use R1:f_o tuning in their upper range and altos, tenors and baritones use R1:nf_o tuning occasionally: *i.e.* they tune R1 to one of the harmonics of the voice [6,7].
Because of the vocal tract's length and the limit to the ability to open the lip aperture, it is difficult to increase R1 much above 1 kHz, which corresponds roughly to the typical upper limit of pitch range for most sopranos, *i.e*. about C6 or 'high C'.
A minority of sopranos sing pitches well above C6. In a study of those who do use this range [8], we found that most make a transition, in the region of C6, to R2:f_o tuning: they tune the tract's second resonance to the fundamental of the voice. This may be counter-intuitive: in spite of having learned (whether taught or not) to increase the lip aperture when ascending a scale, they would need suddenly to reduce the aperture to go higher. This observation led us to wonder whether the difficulty in learning R2:f_o tuning might be a cause of the upper limit to the range of most sopranos. This in turn was one incentive for the current project.

1.2 Resonance tuning in wind instrument performance

Reed instruments usually operate at a frequency very near to a maximum in the impedance spectrum of the bore[2]. To access the second register (operated by the second resonance and second impedance peak), players usually open a small register hole, which weakens and detunes the first resonance, making it unplayable. For higher registers, either more register holes are used, or some of the tone holes serve that purpose.
Saxophones have a larger cone angle than other reed instruments. This has the intended consequence of making them louder, but makes the third impedance peak too weak to drive the reed [9]. For that reason, many saxophonists are limited to the first two registers, giving a pitch range of about two octaves and a fifth.
To access the third and higher registers (known to saxophonists as *altissimo*), players learn to tune an impedance peak[3] of the vocal tract to the frequency of the desired note [10,11]. The series combination of bore and tract impedance then drives the reed at that frequency. Clarinettists also use vocal tract tuning to access the *altissimo* registers [12].
Both clarinettists and saxophonists also use tract tuning for pitch glides [12] and for controlling multiphonics [13].
Providing feedback that might teach resonance tuning to reed players was a second motivation for this project and we are continuing with work on that aspect.

2 MEASUREMENTS

2.1 Measuring resonance tuning in and outside the laboratory

For both singing and reed playing, we have measured vocal tract impedance spectra using the capillary method: a broadband source with a high output impedance and a microphone are combined in an impedance head. This is first calibrated to remove the frequency response of the source. The different experimental details are described elsewhere [12] and a schematic is shown in Fig 1. Briefly, a synthetic broadband signal, synthesised on a computer, is output via an interface to an audio power amplifier and then to a horn driver. An inverted horn matches it to a flexible tube, next to which is attached a microphone. The microphone is connected via a preamplifier to the audio interface and thence to the computer.
Measuring the frequency of resonances for singers has the challenge of ecological relevance: it is possible to make precise measurements of the vocal tract (and trachea) using a multiple microphone system in a duct sealed to the lips [14,15], but this has the disadvantage of making normal singing impossible. For this reason, we have used a flexible pipe as the source and bound a microphone to it—Fig 1. This makes a simple impedance head that, held to touch the lower lip, can measure the frequency of resonances with little inhibition of normal singing. For measurements on clarinet and saxophone players, we install the source and the microphone in the mouthpiece, which then becomes an impedance head. For lip reed instruments, two thin tubes are inserted in the corner of the mouth of the player; these then become the impedance head [16,17].

[2] Reed instruments resemble the voice in that the bore resonances have a pressure antinode at the source and a node near the bell or the start of the array of open tone holes. Of course, the qualitative difference is that, in instruments, it is the resonator and not the source that largely determines the pitch—inadvertent squeaks excepted.

[3] In the notation used for the voice, an impedance peak at the voice would be an antiresonance, though the term resonance is sometimes used for both maxima and minima in acoustic impedance.

2.2 Hardware

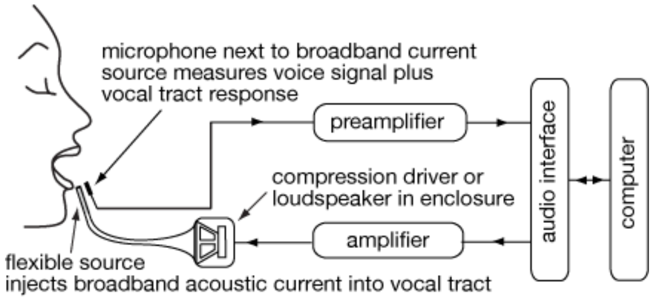

Figure 1 – A schematic (not to scale) of the hardware used for resonance feedback in singing.

In the lab, we use research-quality microphones, preamplifiers and audio interfaces. For the purposes of providing feedback, less expensive items may be used. A key element is a horn driver that delivers significant power at the lower frequencies of interest; many commercial options are available. Linearity in the microphone is important, so a low sensitivity microphone is desirable. However, we have in the past used cheap tie-clip microphones with acceptable, though not good, results. A 3D print design for the horn, along with other information, will be posted soon to our web site. As this contribution is submitted as a demonstration presentation, details will be available at the demonstration.

2.3 Calibration

For singing, we ask subjects to hold the source and microphone at the lower lip, with the mouth closed. We then inject a broad band signal comprising a sum of sine waves whose magnitudes and relative phases are adjusted to improve the dynamic range. The spectral envelope of the microphone signal then exhibits the frequency-dependent gain of the source, microphone and amplifiers, and the acoustic impedance of the radiation field at the lips, baffled by the face. To 'flatten' this curve, a new output signal is generated having Fourier components proportional to the reciprocals of those of the first measurement [18]. When this is output, the linear gain of the entire system is compensated and the next pressure signal measured (under the same conditions) is close to flat. A further iteration is often helpful to compensate for nonlinear effects in microphone or speaker. The flattened signal measured with closed lips we call p_{closed}.

For measuring the tracts of wind players, our calibration load is an acoustically infinite waveguide, into which the impedance head is inserted. Ideally these ducts should be straight to avoid reflections, so we have them installed in the ceiling space of the building. However, less precise measurements are possible using calibrations made on a long coil of pipe, provided precautions are taken to avoid crosstalk between loops of the coil [19].

2.4 Measurements

The calibration using infinite pipes, whose impedance is a pure resistance, means that the pressure measurements made with this source calibration are proportional to the acoustic impedance (provided that the measured impedance is small compared to the output impedance of the source). For the measurements in the mouths of wind players, maxima in the measured probe signal give the vocal tract impedance maxima.

For singing, the source is calibrated on the radiation impedance, as described above. When this calibrated signal is output during singing, the current source is loaded with the acoustic impedance of the vocal tract, as measured at the lower lip, in parallel with the radiation field, still baffled by the face, albeit with an altered geometry. We call this microphone signal p_{open}. We then plot the ratio $\gamma = p_{open}/p_{closed}$. Both calibration and measurement are shunted by the radiation impedance, the source is approximately a current source for these loads, so γ is approximately equal to the ratio of their impedance. At low frequencies, maxima in γ are a good approximation to the resonances of the tract, with precision usually much greater than that given by formant estimations. However, neither this technique nor formant estimation gives a good measurement of resonance amplitude or bandwidth.

Measurements of singers and players have the serious complication that the harmonics of the note played or sung usually have a much greater amplitude than the response of the system to the probe signal. Figure 2 gives an example for a measurement of p_{open}/p_{closed} on a singer. Here, the harmonics at nf_o are prominent. Peaks in the broadband signal indicate the resonances, as marked. Observe that the fifth harmonic falls close to R2 and thus receives an amplitude boost, called the formant F2. In this case, however, the R1 resonance falls between the second and third harmonic, so no harmonic is boosted by this resonance: R1 does not produce a formant F1 in this case.

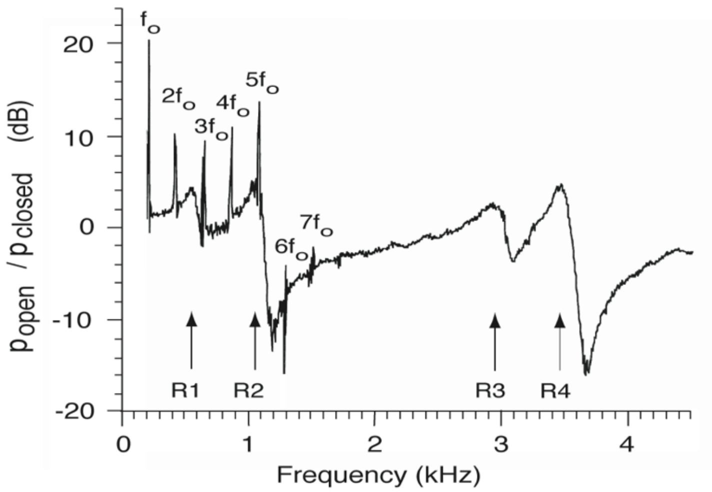

Figure 2 – A plot of the measured pressure ratio, mouth open to mouth closed.

We note that the approximations made above in interpreting p_{open}/p_{closed} are less accurate at high frequencies and that peaks in γ correspond less well with resonances. This limitation does not apply to measurements made with the lips sealed around a duct-type impedance head [*e.g.* 14,15]. However, the geometry in that case precludes ecological articulation.

3 SOFTWARE INCLUDING GRAPHICAL USER INTERFACE

3.1 Shareware software

The shareware software has a graphical user interface (GUI) with panels for measurement parameters, calibration, display, recording and editing and a number of other features. These are described in more detail elsewhere [20].

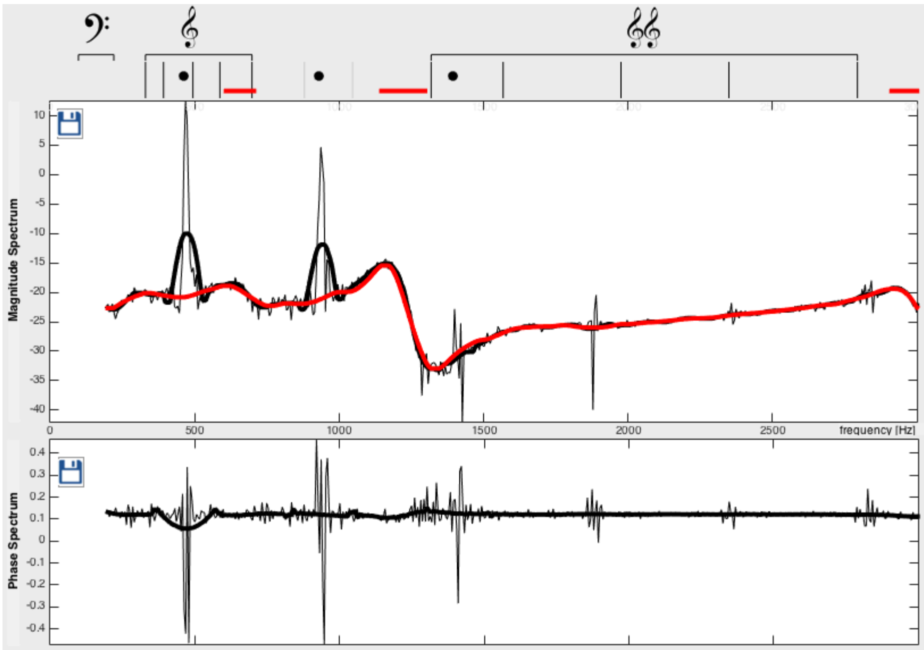

Figure 3 – A panel from a screen grab during a singing measurement.

Figure 3 shows a panel from a screen grab. The two graphs show the magnitude and phase of p_{open}/p_{closed} during a sung note (Bb4). The ordinates are dB and radians. The abscissa in the middle is frequency in Hz. Above the graph this is converted to pitch using the three clefs, bass, treble and supertreble (two octaves above treble). The fine black line is the raw p_{open}/p_{closed} signal, which is smoothed to give the thick black trace. The harmonics are then removed and interpolated to give the red line, which is thus the response of the vocal tract to the injected signal. On the music staves are shown horizontal red bars, which are estimates of the resonance frequencies, including error bars.

The singer in question is not tuning resonances—the pitch is towards the lower end of the range over which sopranos practise resonance tuning. In a tuning exercise, the singer would sustain the note while making an articulatory change and simultaneously watch the positions of the horizontal bars, adjust articulation to move one of the bars (R1 or R2) to overlap with one of the harmonics (nf_o), thus to achieve the desired Rn: nf_o tuning.

4 EXAMPLES OF USE

4.1 Teaching basic articulation skills

For some learning exercises, it is helpful to learn the articulatory gestures while miming, with glottis closed and with velum raised (*i.e.* without nasalisation). Figure 4 shows the effects of these separately. In the 'normal' condition, with glottis closed and velum raised (top left), R1, R2 and R3 are visible. Opening the glottis (top right) then adds the trachea below the glottis. One would expect this roughly to double the length of the duct, with an open end approximating the remote end where the trachea branches rapidly. This would roughly double the number of expected resonances in a given frequency range [15]. One extra resonance does indeed appear at ~1 kHz. The higher whole-duct resonances have not appeared probably because the glottis was only partly open and the inertia of the air in the glottis was large enough to isolate the two tracts at high frequencies. Finally, lowering the velum gives a single strong resonance in the range viewed. (For measurements at the lips, and with the velum completely lowered, it makes little difference whether the glottis is open or closed.)

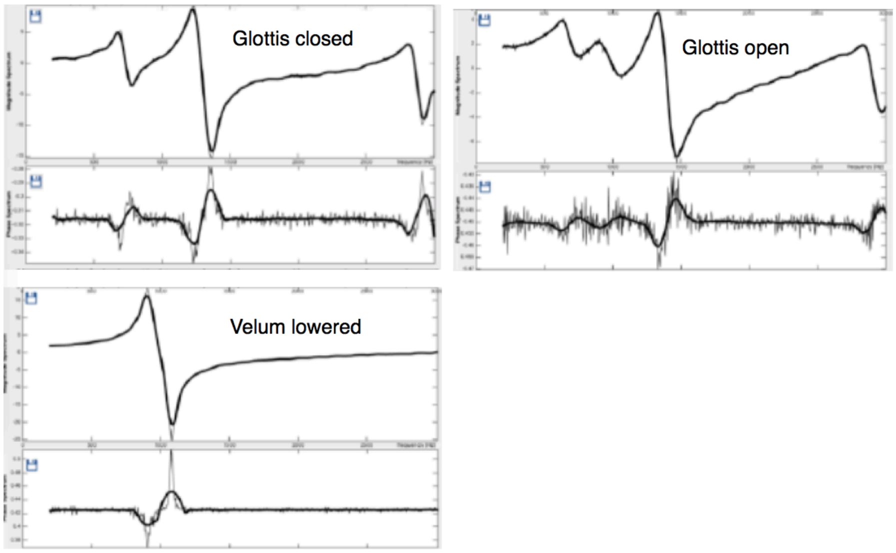

Figure 4 – Three screen grabs with the gestures indicated. Again, the frequency axis is from 0 to 3 kHz.

4.2 Training R2:f_o tuning

Can visual feedback teach resonance tuning? We conducted a preliminary experiment using this interface.
As mentioned above, sopranos do not need help to learn R1:f_o tuning. But those who sing in the range well above high C usually use R2:f_o tuning. Could we teach this skill? A formal account of this project has been submitted elsewhere, but the simple results will be described here.
Eight sopranos came to the lab for a one-hour session each. All of them had a self-described range with an upper limit near high C (C6). Using the hardware and software described above, each of them first tried to learn how to control glottis and velum while miming (see Fig 4). In the next stage, they mimed the vowel /*a*/, then gradually adjusted the vocal gesture to put R2 (the second red bar in Fig 3) at a pitch near G5. Then they heard that note on a glockenspiel, and attempted to sing it, without changing the mouth or tongue shape. Those that achieved the goal then sang up a scale, tuning R2 as they went.

4.3 Training results

Of the eight subjects, six learned to control the glottis and velum when miming—the other two had difficulty in closing the glottis. Once these two did close the glottis, they then had difficulty sealing the nasal tract with the velum.
One of these two was never able to completely seal the velum when miming. The other seven were able to use the visual feedback to lower R2 to a desired value while miming: one could lower it as far as D5.
Of the six subjects with good independent control of glottis and velum, five learned to lower R1 when singing.
However, only two learned how to perform R2:f_o using the visual feedback in the one-hour session (see Fig 5).
After the training, subjects sang an ascending scale while R1 and R2 were measured by the experimenter, without displaying visual feedback. In these measurements, all eight subjects were observed to use R1:f_o tuning over

the suitable range, but none spontaneously switched to R2:f_o near the top of their range. Evidently, one hour is not enough time to teach what is, for sopranos, a counter-intuitive technique: to reduce the lip aperture on an ascending step in a scale.
Just one soprano has used this interface for a sustained period. One of the authors (MJ), a soprano with no formal training, spent an estimated half an hour per day over a period of three months while developing the software used here. It took her about 4 weeks to have competent and independent control of the glottis, velum and vocal tract gesture.
After three months, she was able to use R1:f_o up to C#6 and R2:f_o over the range from E5 to E6, without needing to use the feedback system. Further, she extended her comfortable range by a fifth (A5 to E6) and her absolute upper limit to F#6, the range D6 to F#6 being only possible with R2:f_o tuning. The results of the study summarized here will (we hope) appear in more detail later.

5 CONCLUSIONS

Doing precise measurements of vocal tract resonances requires sophisticated hardware and has, until recently, required software that has gradually evolved for research purposes and was not widely available. New software has been written for use in feedback applications and is available as shareware. Simple versions of the hardware and software suitable for singing will be available for demonstration in the conference demonstration session. Software download and details are at http://www.phys.unsw.edu.au/jw/broadband.html

ACKNOWLEDGEMENTS

We thank the Australian Research Council for its support and our volunteer subjects.

REFERENCES

1. Clark J, Yallop C, Fletcher J. An introduction to phonetics and phonology. Blackwell, Oxford 2007.
2. Titze IR. Nonlinear source-filter coupling in phonation: Theory. J Acoust Soc Am. 2008; 123: 2733–2749.
3. Sundberg J. Formant technique in a professional female singer Acustica 1975; 32: 89–96.
4. Sundberg J, Skoog J. Dependence of jaw opening on pitch and vowel in singers. J. Voice 1997; 11:301–306.
5. Joliveau E, Smith J, and Wolfe J. "Vocal tract resonances in singing: The soprano voice," J Acoust Soc Am 2004; 116:2434–2439.
6. Henrich N, Smith J, Wolfe J. Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. J Acoust Soc Am. 2011; 129: 1024-1035.
7. Henrich N, Kiek M, Smith J, Wolfe J. Resonance strategies in Bulgarian women's singing. Logopedics Phoniatrics Vocol. 2007; 32: 171-177.
8. Garnier M, Henrich N, Smith J, Wolfe J. Vocal tract adjustments in the high soprano range. J Acoust Soc Am. 2010; 127: 3771-3780.
9. Chen JM, Smith J, Wolfe J. Saxophone acoustics: introducing a compendium of impedance and sound spectra. Acoust. Aust. 2009; 37: 18-23.
10. Chen JM, Smith J, Wolfe J. Experienced saxophonists learn to tune their vocal tracts. Science 2008; 319: 726.
11. Scavone GP, Lefebre A, Da Silva AR. Measurement of vocal-tract influence during saxophone performance. J Acoust Soc Am 2008; 123: 2391–2400.
12. Chen JM, Smith J, Wolfe J. Pitch bending and glissandi on the clarinet: roles of the vocal tract and partial tone hole closure. J Acoust Soc Am. 2009; 126: 1511-1520.
13. Chen JM, Smith J, Wolfe J. Saxophonists tune vocal tract resonances in advanced performance techniques. J Acoust Soc Am. 2011; 129: 415-426.
14. Hanna N, Smith J, Wolfe J. Frequencies, bandwidths and magnitudes of vocal tract and surrounding tissue resonances, measured through the lips during phonation. J Acoust Soc Am. 2016; 139: 2924–2936.
15. Hanna N, Smith J, Wolfe J. How the acoustic resonances of the subglottal tract affect the impedance spectrum measured through the lips. J Acoust Soc Am. 2018; 143: 2639-2650.
16. Tarnopolsky A, Fletcher N, Hollenberg L, Lange B, Smith J, Wolfe J. Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I: Experiment. J Acoust Soc Am. 2006; 119: 1194-1204.
17. Boutin H, Fletcher N, Smith J, Wolfe J. Relationships between pressure, flow, lip motion, and upstream and downstream impedances for the trombone. J Acoust Soc Am. 2015; 137: 1195-1209.
18. Epps J, Smith JR, Wolfe J. A novel instrument to measure acoustic resonances of the vocal tract during speech. Measurement Sci. Tech. 1997; 8: 1112-1121.
19. Dickens P, Smith J, Wolfe J. Improved method of measuring reflection or impedance spectra using adapted signal spectra and resonance-free calibrations. Int. Cong. Acoust., Sydney. 2010. M. Burgess, ed.
20. Music Acoustics. http://www.phys.unsw.edu.au/jw/broadband.html UNSW 2019.