Second International Workshop on the History of Speech Communication Research (HSCR 2017)
18-19 August 2017, Helsinki, Finland

ORIGINS OF SOURCE FILTER THEORY FROM MERSENNE (1636) TO MÜLLER (1839)

Daniel Aalto[1,2]
[1]*Communication Sciences and Disorders*
Faculty of Rehabilitation Medicine, University of Alberta
[2]*Institute for Reconstructive Sciences in Medicine*
Misericordia Community Hospital, Edmonton, Canada
aalto@ualberta.ca

Abstract: Source filter theory is central in the modern description of speech production. The key concepts underpinning the theory are the sustained oscillation of vocal folds and the vocal tract resonances that modulate the source signal. In this short essay the development of these concepts is superficially followed from early 17th to late 19th century. Marin Mersenne was the first to experimentally verify the relationship between the pitch of a sound (of a cord) and its fundamental frequency, published in 1636. By analogy, Mersenne concluded that vocal folds must vibrate with the same frequency as a string with the same pitch. While Mersenne could not explain why several frequencies could be simultaneously present in glottal excitation, Daniel Bernoulli suggested superposition of waves as a solution in 1753, which was further developed by Joseph Fourier. Johannes Müller's report of excised larynx experiments in 1839 can be considered as the first version of source filter theory. Finally, once X-ray imaging during speech production was made possible in 1897, the stage was set for experimental validation of source filter theory.

1 Introduction

The main events that contributed to the creation of source filter theory can be traced back to key experiments and theoretical developments that are documented in books, correspondence, and journal articles. However, limited availability and accessibility of written sources may distort the story that emerges.

In preparing this essay, the author has sought both original sources and earlier descriptions of related historical trajectories. The original sources include Mersenne's Harmonie Universelle [1], Daniel Bernoulli's Hydrodynamica and medical writings [2], Fourier's paper on heat conduction [3], Willis's theory of vowels [4], and Müller's book on physiology (English translation) and his paper on excised larynges [5, 6]. In all historical texts automated search with suitable anatomical (e.g., larynx, vocal folds, etc.) and acoustic keywords (e.g., vibration, resonance, etc.) were used to limit detailed reading to relevant content. In addition to the original historical sources, descriptions of them were used to create a balanced interpretation of the sources in their historical and scientific context. Parallel historical trajectories describing the development of pitch perception models [7], general acoustics [8, 9], speaking machines [10], and vowel production studies [11] were used to identify when and how related concepts emerged to guide the interpretation of the texts regarding speech production.

2 Marin Mersenne and fundamental frequency

Marin Mersenne (1588–1648) is usually credited for determining the relationship between frequency and pitch of a vibrating string [8]. He attended to a school run by Jesuit priests in France,

10.21437/HSCR.2017-7

the same school where René Descartes was accepted a few years later [12, p. 16]. Robert Lenoble characterizes Mersenne as the secretary of the scientists of the time. True, Mersenne had a vast correspondence (compiled later by Armand Beaulieu to fourteen thick volumes) with his contemporaries. Hence, many of the ideas presented by Mersenne reflect the collective knowledge of the scientists of the time.

This is particularly true for Harmonie Universelle, a monograph published in 1636 describing in detail the theory of music and practical instructions regarding interpretation of music (e.g. embellishments) and performance technique including instrumental and singing music. Musical theory is summarized by Mersenne under title *Abregé de la Musique speculative* (adjusted following Cohen [13, p. 100]) as

1. Sound is nothing but a percussion (literally: "beating") of the air, taken up by the sense of hearing when it [the ear] is touched by it [the sound]. Loudness is proportional to the amount of air struck; pitch is proportional to the number of percussions per unit time.
2. The relative sweetness of the consonances is determined by how often the percussions that make up their sounds coincide.
3. The relative sweetness of the consonances is calculated as follows. If the percussions of a certain interval unite, for example, every second time (the octave), it is sweeter than if they unite, for instance, every third time (the fifth).
4. Given a string of a certain length, thickness, and material, quadrupling its tension makes its sound rise one octave.
5. The velocity of straight sound is 453 meters per second independent of loudness and the direction of the wind; the echo's velocity is 319 meters per second.

The significance of Mersenne's work is in compiling the existing knowledge about sound and music to one work and in advancing the knowledge by experimentally establishing the connection between frequency and pitch. In the argument, Mersenne describes the Pythagoran law (frequency is inversely proportional to the length of a string) and goes into further detail relating the tension and the mass of the string to the pitch. This law was first described by Beeckman in 1616, communicated to Descartes in 1618 who was in constant correspondence with Mersenne [12, p. 483]. However, sometimes Galileo is attributed as the inventor of the law [8, pp. xii–xiii], potentially because Mersenne was promoting Galileo's ideas.

To experimentally test the laws of the vibrating strings, Mersenne produced a series of strings that were predicted to have a theoretical pitch difference of exactly one octave. As the strings got longer, there was some discrepancy between the observation and the formula [1, 1.III, propos. XIII, p. 187]. As a byproduct, Mersenne noted that the lowest audible frequency was twenty beats per second [1, 1.III, propos. V, p. 170]. At the same page Mersenne noted that with very long strings (seventeen foot) there is no audible sound, but on the other hand the frequency of the oscillations is visible and can be counted by eye. This remarkable experiment defined through the chain of octaves the frequency of the pitch.

The Harmonie Universelle is divided into two parts each consisting of multiple "books" (fr. *livre*). The fundamental frequency of a vibrating string was derived in detail in the third book. The reasoning was strictly deductive with an emphasis on the logical validity of the steps. However, as explained by Cohen, where Galileo and Descartes soughted certainty fitting with the deductive writing style of the time, Mersenne aimed at precision. When experimental evidence was at odds with universal laws, Mersenne acknowledged the discrepancy. As an example, Mersenne studied the difference between theoretical predictions in his vibrating string formulas and the frequencies of oscillating cords and found a general second order correction factor accounting for the deviations [13, p. 101].

Mersenne and his contemporaries Descartes and Galileo shared a mechanistic view of physiology: the body was operated by a spirit living in the brain. While Mersenne's exploration of senses, in particular vision and the function of the eye, was accurate and neutral with respect to the metaphysical considerations of the mind, he was not very well read in the biological literature of the time [12, pp. 499–502]. Instead of relating his work to that literature, Mersenne chose to describe how the human machine was operated to produce sound from a mechanistic point of view. Mersenne reserved almost one hundred pages to describe the human voice. He opened the book by briefly explaining the structural parts and their roles (adapted from [1, Traitez de la voix et des chants, I, p. 1]):

> The parts that produce voice are called the lungs to push air, windpipe to conduct the air, larynx with an opening resembling a reed of a flute, palate whose concavity makes air or sound to resonate, gargareon (uvula), tongue which forms speech (*parole*) by its movement, front teeth serving the voice production by letting tongue touch them, air without which all the other parts would be nothing, and the mouth of which the lips form most of the letters that are called consonants, in particular the ones that the Hebrews call *labials*.

Mersenne used three dimensions to describe the voice: weak-strong (*foible & forte*), clear-hoarse (*claire & rauque*), and flat-sharp, i.e. low-high (*grave & aigue*) [1, Traitez de la voix et des chants, I, prop. IV, p. 6]. Mersenne explained that

> the strong [voice] resulted from violent movement of the thoracic muscles, the clear [voice] from well regulated humidity in cartilages, membranes, and the laryngeal muscles, and the hoarse from too great humidity or dryness of the same parts.

In explaining the vocal fold relationship to the fundamental frequency, Mersenne used an analogue to the flute and pipe organ. The long descriptions of intralaryngeal muscles and their role in regulating the phonation reveal that Mersenne was able to correctly deduce the impact of various muscles on phonation. These reasonings were based by intuitive application of knowledge gained from musical instruments.

Mersenne applied the voice production theory that he had developed to various questions. For instance, in the proposition XXII he asked whether an individual can produce two voices simultaneously, in proposition XXIX he explained how to build a parabola that collects sounds from different locations in a room to one focal point, in proposition XLVIII he discussed the amount of possible and producible sound combinations given the vowels and consonants in a language, and in the proposition LI he described how to help deaf children to learn to speak.

All the topics chosen for discussion were thoroughly treated with an easy-to-follow logic. A long explanation was devoted to articulatory description of different "letters" by place of articulation [1, Traitez de la voix et des chants, I, proposition XLIII, p. 56–59]. Mersenne explained that five vowels, *a, e, o, i, u* were sufficient to derive a vowel in any language by interpolating between *a* and the other four. For these five vowels he described the characteristic positions of the tongue and the lips. Mersenne emphasized that the roles of larynx and supraglottal articulators were distinct:

> Or il faut premierement remarquer que les voyelles ne se forment pas par la seule ouuerture du larynx, & de la glotte, qui n'a nulle autre vertu que de former les sons graues & aigus, forts & foibles, clairs, & rauques, &c. car les sons ne seroient nulle voyelle si l'on n'auoit point de langue, dont le plus simple abbaissement qui se fait

> au bout forme la premier voyelle A, lors qu'elle s'estend, & qu'elle soustient le son; l'O se fait quasi par la mesme situation de la langue, car elle se retire & s'enfle fort peu vers le milieu du palais [1].

As much as Mersenne provided insight to the human voice production, he could not explain why multiple frequencies could be heard simulatenously in the human voice and in vibrating strings. However, Mersenne's description of speech production was articulatorily accurate with conceptual separation of vocal source and vocal tract. In particular, Mersenne was the first to quantify phonation in terms of the frequency of glottal pulses.

3 Daniel Bernoulli and Joseph Fourier

Daniel Bernoulli (1700–1782) is best known for his work in hydrodynamics that he published in 1738. However, before devoting his time to physics, he studied medicine graduating in 1720, and wrote a doctoral dissertation about respiration in Basel in 1721 [2, V. Zimmermann in Part 1, p. 6]. After working as a physician in Italy for two years [14], he moved to St Petersburg to serve in the Academy of Sciences where he worked with Leonhard Euler and began writing Hydrodynamica [15, G.K. Mikhailov in Chapter 9]. In 1733, he returned to Basel as a professor in anatomy and botany. Although Bernoulli's passion was clearly in mathematics and physics, his contribution to medicine was substantial, including a treatise on muscular mechanics (1726), and on mechanical work done by the heart (1737, 1748) [2]. Bernoulli did not directly address blood circulation or airflow in the respiratory system in *Hydrodynamica* although the principles presented in the book directly apply to these physiological situations and left a permanent trace into the history of medicine [14].

The relevance of Bernoulli's work to source filter theory is twofold. First, the Bernoulli principle was presented in *Hydrodynamica* and later applied to glottal flow and sustained vocal folds vibration [16]. Second, Bernoulli introduced the concept of superposition of waves (as explained by Robert Lindsay in [8, p. xv]) in 1755 that was crucial in shaping the theory of Fourier.

Joseph Fourier (1768–1830) presented his solution of the heat equation based on spectral decomposition, i.e. as an infinite sum of trigonometric functions, to the Institut de France in 1807. It was not received well by Laplace and Lagrange and the paper was published much later [17]. However, it soon became a well respected manuscript in mathematics [15, Chapter 26]. The mathematical challenges in handling infinite series of trigonometric functions inspired a considerable amount of mathematical work. Earlier, the analysis of the vibrating string by d'Alembert had already sparked a controversy about the definition of a function in 1747 and got intertwined with the problems of convergence in early 19th century. The interrelated mathematical concepts and arguments are described in detail by Gonzalez-Velasco [17].

As Fulop points out, the value of Fourier's work was in generalizing the earlier superposition concepts of Euler and Bernoulli and in creating an integral transform that could handle aperiodic signals as well [18]. However, it was not until 1966 before the convergence problems in the Fourier's theory were finally solved by Lennart Carleson [19].

To conclude, both Daniel Bernoulli and Joseph Fourier pushed their respective fields forward and created mathematical tools and concepts that became instrumental to speech production research. However, there appears to be no direct evidence to demonstrate that they would

---

[1] But first a remark has to be made that vowels are not formed just by opening larynx, and the glottis, which have no other quality than to form low and high, strong and weak, clear and rough, etc. sounds, since the sounds would be no vowels at all if one has no tongue, which forms the first vowel A if simply lowered to the bottom, as it extends and supports the sound; the O is produced at the almost same tongue position, since it retracts and expands only little towards the middle of the palate.

have applied their knowledge to the human voice.

4 Johannes Müller and excised larynges

Johannes Müller (1801–1858) was a prominent anatomist and physiologist advocating microscopic research and physiological experimentation. His comprehensive book on human physiology and his experimental paper on excised larynges may be considered the first account of source filter theory [5, 6, 11]. Similar to Mersenne, Müller synthesized knowledge of his field together with advancing it with new experimental evidence.

Müller summarized the speech production acoustics in the following [6, p. 695]:

> The circumstance of the vocal tube in front of the vocal cords being double, consisting of the oral and nasal canals, seems to have no other effect on the height of the notes than a simple tube; but it causes their character to be altered by the resonance. I attached to the short anterior tube of an artificial larynx with caoutchouc tongues a bifurcated tube; the result was no raising of the notes, but they were rendered more sonorous.

In addition to the artificial larynges, Müller used extensively excised human and animal (at least bovine, ovine, and simian) larynges coupled with acoustic tubes. Hence, Müller was able to experimentally verify that the source signal and the vocal tract resonances were, to a large extent, independent of each other. However, Müller mentioned that elongation of the tube that was coupled with the excised larynx lowered the pitch by a semitone, or very rarely, an entire tone [5]. Underlying these reasonings is the observation from musical instruments, in particular reed instruments, that a resonance (usually the lowest) of the instrument forces the frequency of the reed to lock to the resonant frequency. It appears that Müller felt it necessary to explain why such a locking does not take place in the larynx.

At the surface, the earlier quotes from Mersenne and Müller show little difference in their description of the voice production. However, Müller could provide the frequencies of the vocal tract resonances that he observed in his experiments, whereas Mersenne could only describe the frequency of the vocal source oscillations. While Müller was not the first one to determine the frequencies of the vocal tract resonances, he was apparently the first to experimentally study the interaction between vocal fold and vocal tract oscillations. However, Müller never delved into the signal processing of the speech sounds nor mentioned Fourier's work although he extensively cited earlier works in medicine and speech research including Willis's study of vowels [4]. In contrast, Hermann Helmholtz, who was listening to Müller's lectures on anatomy and physiology in Berlin and whom Müller mentored to a successful career in science, managed to bring together the theory of resonance and vowel production [20].

When John Strutt, better known as Lord Rayleigh, published a textbook about acoustics in 1877, Helmholtz asked for him to include speech acoustics as a third volume [21]. Rayleigh followed Helmholtz's advice and expanded the second edition to include a chapter about speech, published in 1894. However, it is clear that a consensus about vowel production mechanisms had not yet been formed.

In 1890's two remarkable developments can be pinpointed that fundamentally impacted the 20th century phonetics. First, anatomical imaging data during vowel production was presented by Max Scheier in the meeting of otolaryngologists less than a year after the invention of the X-ray imaging technology [22, 23]. The new technology later provided crucial data for refining and validating the models of vocal tract acoustics [24]. Second, sound recordings and Fourier analysis of these waveforms became gradually more common establishing the source filter framework through a new concept, the formant [18, pp. 52–56], [25].

5 Conclusions

Source filter theory developed gradually from qualitative, anatomical descriptions of speech production. In the early 17th century Mersenne deduced the fundamental frequency of phonation by analogy, hence drafting the first quantitative account of vowel production. Bernoulli proposed that multiple frequencies may be present in the vibration of a body which Fourier confirmed in the early 19th century theoretically. Müller's work on artificial and excised larynges in 1839 articulated the essence of the source filter theory by observing that resonances of the vocal tract and the vibrating frequency of the vocal folds are largely unrelated. It is noteworthy that the key concepts of vocal source oscillations and spectral filtering were established before sound was ever recorded.

6 Acknowledgments

The author is grateful for Louis Clerc, Gabriela Constantinescu, Fred Cummins, and anonymous reviewers for feedback that improved the manuscript.

References

[1] MERSENNE, M.: *Harmonie Universelle.* Sebastien Cramoisy, Paris (digitally accessible at imslp.org), 1636.

[2] SPEISER, D.: *Die Werke von Daniel Bernoulli.* Birkhäuser, Basel, 1996.

[3] FOURIER, J.: *Théorie analytique de la chaleur.* Firmin Didot, Paris, 1822.

[4] WILLIS, R.: *On the vowel sounds, and on reed organ-pipes. Transactions of the Cambridge pholosophical society*, pp. 231–276, 1830.

[5] MÜLLER, J.: *Über die Compensation der physischen Kräfte am menschlichen Stimmorgan.* A. Hirschwald, Berlin, 1839.

[6] MÜLLER, J.: *Elements of physiology.* Translated from German by W. M. Baly, Lea and Blanchard, Philadelphia, 1843.

[7] DE CHEVEIGNÉ, A.: *Pitch perception models.* In *Plack, C. and Oxenham A. (eds), Pitch*, pp. 169–233. Springer, 2005.

[8] STRUTT, J. W.: *The theory of sound, 2nd ed. revised and enlarged 1894 with a historical introduction by R. B. Lindsay.* Dover Publications, 1929.

[9] ROSSING, T. D.: *Springer Handbook of Acoustics.* Springer, New York, 2007.

[10] FAGYAL, Z.: *Phonetics and speaking machines: on the mechanical simulation of human speech in the 17th century. Historiographia Linguistica*, 28, pp. 289–330, 2001.

[11] HONDA, K.: *Evolution of vowel production studies and observation techniques. Acoustical Science and Technology*, 23(4), pp. 189–194, 2002.

[12] LENOBLE, R.: *Mersenne ou la naissance du mécanisme.* Paris, J. Vrin, 1943.

[13] COHEN, H.: *Quantifying music, the science of music at the first stage of the scientific revolution, 1580–1650.* D. Reidel Publishing Company, Dordrecht, 1984.

[14] FYE, W.: *Profiles in cardiology – Johann and Daniel Bernoulli. Clinical cardiology*, 24, pp. 634–645, 2001.

[15] GRATTAN-GUINNESS, I.: *Landmark writings in Western mathematics, 1640–1940*. Elsevier, 2005.

[16] COOPER, D.: *Voice: a historical perspective; Woldemar Tonndorf and the Bernoulli effect in voice production. Journal of Voice*, 3(1), pp. 1–6, 1989.

[17] GONZELZ-VELASCO, E.: *Connections in mathematical analysis: the case of Fourier series. American Mathematical Monthly*, 99, pp. 427–441, 1992.

[18] FULOP, S. A.: *Speech spectrum analysis*. Springer, 2011.

[19] CARLESON, L.: *On the convergence and growth of parital sums of Fourier series. Acta mathematica*, 116(1), pp. 135–157, 1966.

[20] FINGER, S. and N. J. WADE: *The neuroscience of Helmholtz and the theories of Johannes Müller: Nerve cell structure, vitalism, and the nerve impulse. Journal of the history of the neurosciences*, 11, pp. 136–155, 2002.

[21] HELMHOLTZ, H.: *Book review: The theory of sound. Nature*, 19, pp. 117–118, 1878.

[22] SCHEIER, M.: *Anwendung der Röntgenstrahlen für die Physiologie der Stimme und Sprache. Deutsche Medicinische Wochenschrift*, 1897.

[23] DOMMANN, M.: *Durchsicht, Einsicht, Vorsicht – eine Geschichte der Röntgenstrahlen 1896–1963*. Chronos-Verlag, Zurich, 2003.

[24] FANT, G.: *Acoustic theory of speech production*. Mouton, The Hague, 1960.

[25] TITZE, I. R., R. J. BAKEN, K. W. BOZEMAN, S. GRANQVIST, H. HENRICH, C. T. HERBST, D. M. HOWARD, E. J. HUNTER, D. KAELIN, R. D. KENT, and J. KREIMAN: *Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. The Journal of the Acoustical Society of America*, 137(5), pp. 3005–3007, 2015.