Mechanical Speech Synthesis in Early Talking Automata

Gordon J. Ramsay

Address:

Spoken Communication Laboratory
Marcus Autism Center
1920 Briarcliff Road NE
Atlanta, Georgia 30329
USA

Email:

gordon.ramsay@emory.edu

Early attempts at synthesizing speech using mechanical models of the vocal tract prefigure modern embodied theories of speech production.

Introduction

Three centuries of scientific research on speech production have seen significant progress in understanding the relationship between articulation and acoustics in the human vocal tract. Over this period, there has been a marked shift in approaches to experimentation, driven by the emergence of new technologies and the novel ideas these have stimulated. The greatest advances during the last hundred years have arisen from the use of electronic or computer simulations of vocal tract acoustics for the analysis, synthesis, and recognition of speech. Before this was possible, the focus necessarily lay in detailed observation and direct experimental manipulation of the physical mechanisms underlying speech using mechanical models of the vocal tract, which were the new technology of their time. Understanding the history of the problems encountered and solutions proposed in these largely forgotten attempts to develop speaking machines that mimic the actual physical processes governing voice production can help to highlight fundamental issues that are still outstanding in this field. Many recent embodied theories of speech production and perception actually directly recapitulate proposals that arose from early talking automata.

The Voice as a Musical Instrument

By the beginning of the seventeenth century, the anatomy of the head and neck was already well understood, as witnessed by the extraordinarily detailed illustrations found in many books of the period (e.g., Casserius, 1600). An example of a midsagittal cross section of the vocal tract from the first anatomy textbook published in color (Gautier d'Agoty and Duverney, 1745), correctly reproducing all of the major anatomical structures, is shown in **Figure 1**. However, the exact function of the many different structures within the vocal tract and the origin of the human voice were still an active topic of discussion. From the earliest definition of the science of acoustics in the landmark article by Sauveur (1700) and even before, analogies were drawn between speech and music that drove much of the debate.

The first clear understanding that the geometry of the vocal tract directly shapes the timbre of speech was published by Marin Mersenne in his book *Harmonie Universelle* (Mersenne, 1636). In the sixth volume of that remarkable tome, Proposition XXXVI "explains how to construct a set of organ pipes, to pronounce vowels, consonants, syllables, and utterances," correctly inferring that appropriately manipulated tube shapes excited by a reed would produce corresponding speech sounds. Later, the focus shifted to the function of the larynx, with much heated argument about how the vocal folds were able to create sound. Dodart (1700) proposed that the glottis acts as a wind instrument, blown by air flowing over the edges of the hole between the vocal folds, whereas Ferrein (1741) claimed instead that the vocal cords vibrate like a string instrument, bowed by the air from the lungs. Reviewing the evidence from

https://doi.org/10.1121/AT.2019.15.2.11

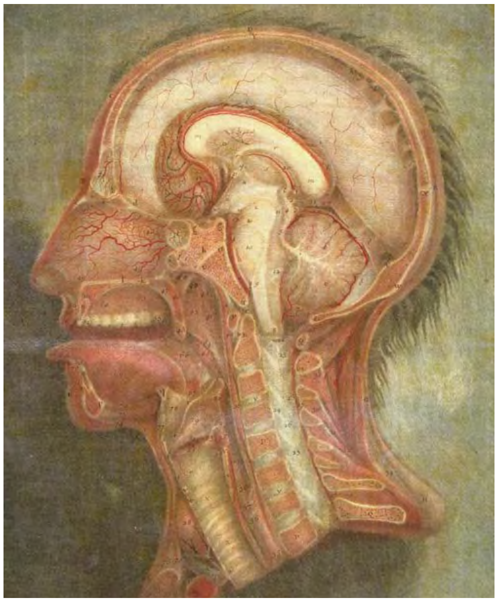

Figure 1. Midsagittal section of the vocal tract from the first color anatomy book of the head and neck. All of the structures of the respiratory, oral, and nasal tracts are accurately labeled in exquisite detail, including the trachea (x), vocal folds (85), jaw (p), tongue (65), palate (54), velum (45), and lips (L, M). The flow of air from the trachea through the vocal folds into the oral and nasal cavities was clearly understood as was the vibration of the vocal folds; the shaping of the oral cavity by the jaw, tongue, and lips; and the action of the velum in closing off the nasal cavity during speech. Reproduced from Gautier d'Agoty and Duverney (1745).

these apparently contradictory viewpoints, Ferrein himself, and later Vicq d'Azyr (1779), concluded that the vibration of the vocal folds, the shape of the glottis, and the glottal airflow could not be meaningfully separated and were all responsible for sound generation, in many respects predicting the modern myoelastic-aerodynamic theory of vocal fold vibration (van den Berg, 1958). By the middle of the eighteenth century, the analogy between the vocal tract and a very special kind of musical instrument was no longer in doubt. The open issue was how to "play" the vocal instrument to produce speech.

Mechanical Reproduction of the Voice

It took until the late eighteenth century for all of these early ideas by Mersenne, Dodart, and Ferrein to be fully explored and implemented. The basic component mechanisms underlying speech production were by now understood: a pair of lungs to create an aerodynamic flow, a pair of vocal folds vibrating under tension and blown by the glottal air flow to create sound, and a tube shaped like the vocal tract to form sound into speech.

Mechanical analogs were proposed, drawing again on comparisons with musical instruments: a pair of bellows for the lungs, a vibrating reed or membrane for the vocal folds, and organ pipes for the mouth and nose. Only the control mechanism and the confidence that a mechanical speaking machine could actually be built were lacking. These were provided by Vaucanson (1738), who constructed an automaton flute player that played tunes by blowing into a real flute. Drawing on a long history of mechanisms used in musical clocks and chamber organs (cf. Kircher, 1650), dating back to before the middle ages, Vaucanson ingeniously employed a revolving cylinder studded with pins to coordinate the timing and activation of a set of levers moving the articulators of his automaton, leaving physics to do the rest. Generalizing the same idea, Engramelle (1775) later published a monograph detailing how individual musical performances could be systematically transcribed onto pinned cylinders, as in a modern music box, and used to drive a mechanical organ for playback. These were the first examples of programmable musical instruments and also the first examples of musical automata designed to reproduce the actions of human musicians. It did not escape the imagination of contemporaries of both Vaucanson and Engramelle that the same mechanism could also be used to synthesize human speech (Doyon and Liaigre, 1966; Séris, 1995).

Kratzenstein's Vowel Tubes and Kempelen's Speaking Machine

The first instantiation of Mersenne's original proposal appeared in 1780, when Christian Gottlieb Kratzenstein, a professor in Copenhagen, won first prize for a competition proposed by Leonhard Euler at the Imperial Academy of St. Petersburg in 1777. Euler asked whether it might be possible to construct a set of organ pipes similar to the traditional *vox humana* stops, which would perfectly imitate the vowels *a, e, i, o,* and *u*. Kratzenstein (1780) responded by making five tubes of metal and wood (**Figure 2**) that he shaped by trial and error to produce approximations of the different vowel sounds when blown with a free reed. Notably, none of these bore any recognizable resemblance to the shape of an actual vocal tract.

At around the same time, Wolfgang von Kempelen spent 20 years making several attempts to create a mechanical speaking

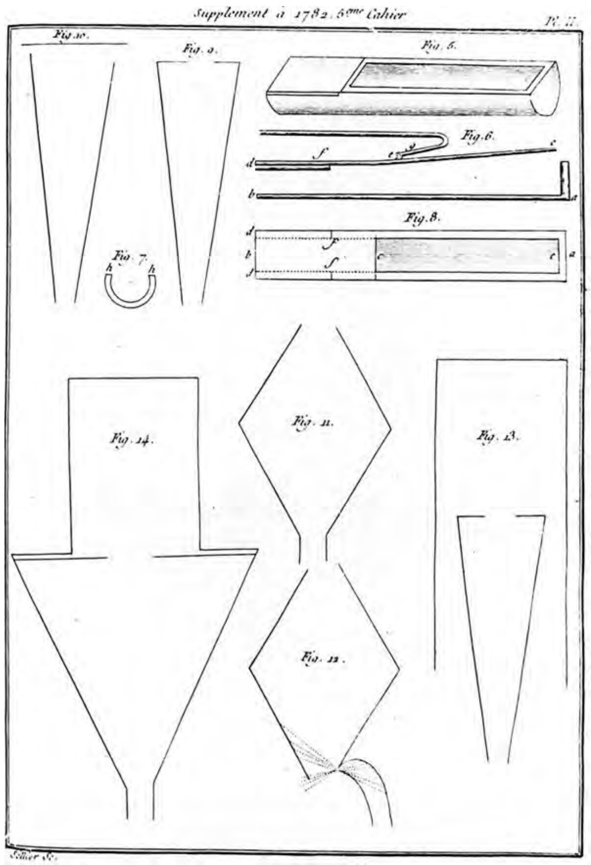

***Figure 2.** Kratzenstein's five vowel tubes for the vowel sounds a (two cross sections in Figs. 9, 10), e (Fig. 11), i (Fig. 12), o (Fig. 13), and u (Fig. 14), excited by a free reed (Figs. 5-8). Kratzenstein remarks that the passage from o (Fig. 13) to u (Fig. 14) is achieved by the stricture of the upper cavity. Reproduced from Kratzenstein (1780).*

machine, as summarized in his famous book (Kempelen, 1791). After working on various pipe models like Kratzenstein, Kempelen eventually returned to analogies between music and speech. Inspired by a rustic bagpipe, he adopted the familiar model of wooden bellows, counterbalanced by a weight, to blow air through a vibrating ivory-and-leather reed mounted inside a box and venting through a flared gutta-percha tube, respectively simulating lungs, larynx, and mouth (**Figure 3**). In the process of developing his model, he thought carefully about the relationship between the detailed articulations of all the different parts of the vocal tract and the phonetic contrasts that he found to be important in different languages, realizing that many speech sounds can be compared and discriminated based on voicing, aspiration, frication, or nasality as well as the shape of the vocal tract. Accordingly, to the basic model he added nostril tubes that could be opened or closed; a means of damping the vibration of the reed to cut off voicing; an extra smaller side bellows that would inflate and then rapidly deflate as the mouth opening was closed off and released to create air puffs for plosives; two side tubes of different lengths bypassing the reed for fricatives; and, finally, a metal wire that could be pushed onto the reed to create a rattle resembling a trill. The bellows were pressed with the right elbow, and the levers and openings on the top of the main box were operated with the right hand to control the secondary modifications, while the left hand was partially inserted into the mouth tube and manipulated empirically to create the primary articulation of the sound. Modern reconstructions have shown that, in the hands of a skilled operator, unrestricted whole sentences can be produced intelligibly on demand, but the perceptual quality is far from natural, as everyone remarked at the time (Liénard, 1967; Brackhane, 2011).

These first two vocal tract models have been described in detail elsewhere many times (e.g., Chapuis and Gélis, 1928; Dudley and Tarnóczy, 1950) and are the most well-known but perhaps the least interesting of all the mechanical speaking machines. Kratzenstein's tubes reproduce isolated vowel sounds acoustically but fail to accurately simulate even the geometry of the vocal tract, let alone the underlying physics. Kempelen's speaking machine captures many of the physical mechanisms responsible for sound production in the vocal tract, albeit crudely, but totally sidesteps many of the crucial issues of articulatory timing and control by reducing the model to a passive musical instrument that needs to be harnessed back to the actions of a human musician. The shaping of the mouth tube, for example, is largely provided by the operator's hand and can only be learned through a great deal of practice and listening. The next two models ultimately resolved these problems.

The "Talking Heads" of the Abbé Mical

An almost exact contemporary of Kempelen, the Abbé Mical was an impecunious cleric, the younger son of a wealthy family from the Dauphiné in France who ran away from the church to pursue mechanics. From 1776 to 1785, he exhibited across Paris first one, then two, carved wooden "talking heads" that produced not only single speech sounds and syllables but also whole sentences and even an entire dialogue (**Figure 4**). On his request, his invention was examined by the Académie des Sciences, which appointed a committee of notable scientists, including Vicq d'Azyr, Laplace, and Lavoisier, to produce a report (de Milli et al., 1784; Chapuis and Gélis, 1928; Hémardinquer, 1961).

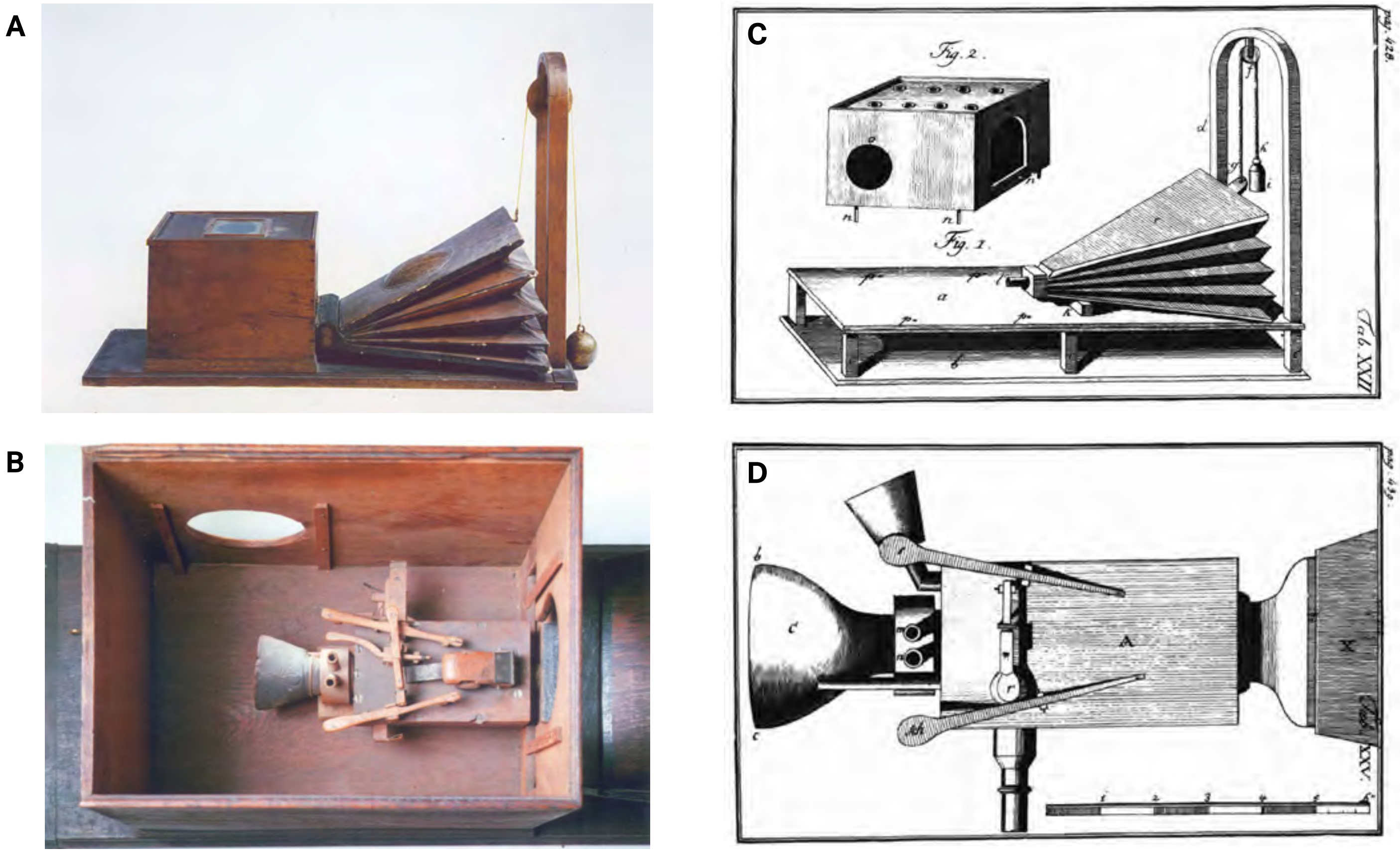

Figure 3. *Kempelen's speaking machine.* ***A*** *and* ***B****: photographs of the mechanism exhibited in the Deutsches Museum in Munich, Germany, thought to be a reconstruction of the original.* ***C*** *and* ***D****: corresponding illustrations from Kempelen (1791).* ***D****: details shown (see text for discussion) include the bellows (X), the box containing the vibrating reed (A), and the mouth tube (C); the other pipes (m and n) and levers (r and sch) are extra modifications needed for nasals, fricatives, and trills. Photographs reproduced from the Deutsches Museum, Munich, Germany, with permission. Copyright Deutsches Museum, München, Archiv, BN37402 and BN37404.*

The same elements persisted from previous attempts. A pair of bellows and a tube mimicked the lungs and trachea and a set of valves directed airflow into a collection of boxes. The entry of each box was covered with a leather diaphragm over an elliptical hole representing the glottis, while a parchment reed covering the hole vibrated like the vocal folds; by moving a metal tongue over the reed, the tone could be adjusted. The different actuators shaping the inside of each box simulated movements of the vocal tract with levers and shutters pulled by cords. Vowels and diphthongs were produced by connecting particular boxes in sequence. Stops were produced by rapidly opening and closing shutters over the ends of boxes. Fricatives were synthesized by silencing the reed and blowing air into the boxes. Trills were produced by a special vibrating reed. Syllables were produced by sequences of movements built into the actuators for each box. Unlike previous mechanisms, the boxes seem to have modeled dynamic articulations approximating specific consonant-vowel combinations.

Most importantly, following Engramelle and Vaucanson, a pinned cylinder was turned to activate the different articulations in sequence, and the position of the pins could be altered to program whole sentences. Examples of utterances given in newspaper reports from the period include single vowels, *a, e,* and *o*; diphthongs, *oa*; syllables, *pe, la, le, fe, fai, ra,* and *ro*; and the following extended series of sentences mimicking a conversation between the two heads that was presented before Louis XVI at Versailles in September 1783 (see **Figure 4** for translation). The 1st Head begins, "*Le Roi donne la paix à l'Europe*"; the 2nd Head replies, "*La paix couronne le Roi de gloire*"; the 1st Head responds to the 2nd head, "*Et la paix fait le bonheur des peuples*"; and then addresses the King, "*O Roi adorable, père de vos peuples, leur bonheur fait voir à l'Europe la gloire de votre trône.*" Public reaction was divided. Many thought the display was sensational as proof that speech could be synthesized mechanically, whereas others complained that they could barely understand what was being said.

Figure 4. *The "talking heads" of the Abbé Mical, exhibited in Paris in 1783. The dialogue spoken by the two heads is written on the curtain hiding the mechanism (see text for actual French): "1st Head: 'The King gives peace to Europe'; 2nd Head: 'Peace crowns the King with glory'; 1st Head: 'And peace makes the people happy'; and 1st Head: 'O adorable King, father of your people, their happiness shows Europe the glory of your throne.'" At the bottom of the display is the caption: "Talking Heads: A problem solved in mechanics that up to this day had been considered unsolvable, or at least very difficult. The Academy of Sciences has said in its report that these talking heads can throw the greatest light upon the mechanism of the vocal organ and on the mystery of speech. It added that this work was worthy of its approval by its novelty, by its importance, and by its execution." Illustration reproduced from the Bibliothèque Nationale, Paris, France, with permission. Copyright Bibliothèque Nationale, Paris, Gallica, ark:/12148/btv1b8410437r.*

This mechanism is remarkable in the history of speech synthesis for several original contributions. The Abbé Mical's "talking heads" constituted the first automatic programmable speech synthesizer. They were the first synthesizer based on concatenation of phonetic units and also the first synthesizer based on replicating the serial ordering of articulations seen in the human vocal tract, using a form of demisyllable synthesis. Finally, as the examples above show, they synthesized the first dialogue between two machines. Against all of these advances is the undeniable deficiency that they seem to have relied on coordinating and switching between multiple separate mechanisms to produce the illusion of speech instead of controlling a single unified model of the vocal tract, which partly accounts for the poor and unnatural sound quality.

Faber's "Euphonia"

In the early nineteenth century, an Austrian mathematician and astronomer, Joseph Faber, came across Kempelen's book and built a replica of the mechanical speaking machine for himself. Quickly realizing the problems involved in playing it by hand, he determined to improve on the original design by turning it into a keyboard instrument to automate the means of control. Like Kempelen's machine, he used bellows and a vibrating reed as lungs and larynx to provide an airflow and sound source. However, he replaced the simple half-open mouth tube that Kempelen shaped by contortions of the hand with a fully configurable model of the whole vocal tract. By mounting six adjustable metal blocks back-to-back in a square box behind a pivoting tongue resting on a rotating jaw that terminated in a pair of moveable rubber lips, he was able to create tube shapes with front and back cavities that could be easily and directly related back to actual vocal tract configurations. As in a real vocal tract, the tongue, jaw, and lips moved to create simple opening and closing movements in the front cavity of the model, incorporating natural articulatory constraints, whereas the blocks behind resemble a six-section area function that could be used to create a more detailed shaping of the back cavity. Following Kempelen, Faber also added a more realistic nasal cavity, inserted a rotating vane to interrupt the airflow for trills, and used a moveable lever to alter the effective length of the reed, raising and lowering the pitch or cutting off the reed vibration completely.

The real ingenuity of Faber's talking machine, however, lay in the control mechanism. The jaw, tongue, lips, six blocks, nasal opening, vane, and reed were all controlled by a set of vertical rods such that the continuously adjustable heights of the rods entirely determined the configuration of the vocal tract model at any point in time. Faber connected the rods to a keyboard with 14 keys and 3 pedals, using a grid of cross-patched levers to transform each single key or pedal depression into a predetermined setting for multiple rods. He

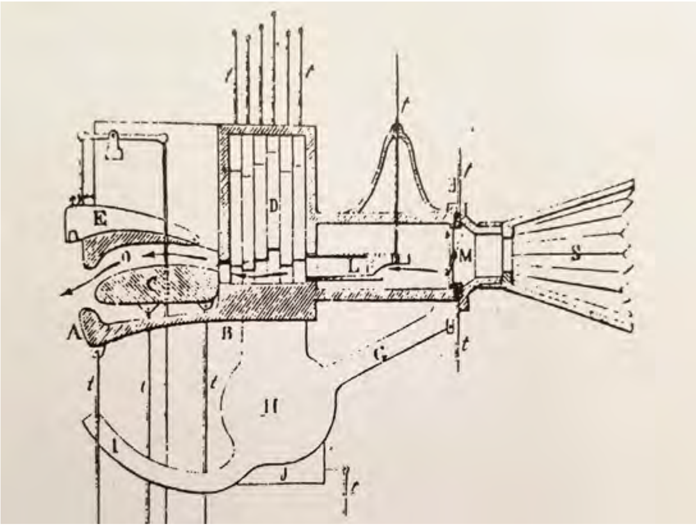

Figure 5. *The mechanism of Faber's "Euphonia," illustrated by du Moncel (1882), showing the bellows (S), the vibrating reed (L), the trill mechanism (M), the moving blocks (D), the jaw (A and B), the tongue (C), the palate (E), the nasal cavity (G, H, and I), and the levers (t) connected to the keyboard;* ***arrows*** *indicate the direction of airflow. Reproduced from du Moncel (1882).*

set up the keyboard mapping so that the keys would produce the following basic sounds: *a, o, u, i, e, l, r, v, f, s, ch, b, d,* and *g,* which we now recognize phonetically as subsets of vowels, liquids, fricatives, and plosives. One of the pedals controlled nasality and another controlled voicing and pitch, with the last pedal manipulating the bellows. By pressing different keys and pedals either together or in sequence while pushing air out of the bellows, any utterance that could be transcribed in terms of these basic articulations could be played on the keyboard and intonation could be added by continuously varying the pitch pedal. Spring loading in the key and rod mechanisms contributed inertia and damping, resulting in a smooth interpolation between adjacent sounds, approximating coarticulation. When demonstrating his invention, Faber would ask for sentences from the audience in any language, which he would then play back on the machine after transcribing them phonetically in his head, reportedly with a German accent. In London in 1846, he even managed to get it to sing "God Save the Queen."

A full description of the machine and its operation was published by du Moncel (1882), and the only known diagram of the mechanism is reproduced from that account in **Figure 5**. An illustration of the entire machine is shown in **Figure 6**, taken from a contemporary newspaper report in Paris in 1877, exactly one hundred years after Mical.

Faber called his talking automaton the "Euphonia" and first presented it in 1841, giving public performances almost continuously for many years until his death in 1864. Afterward, his invention was further improved and exhibited as the "Amazing Talking Machine" by his niece, Marie Trunka, and nephew-in-law, Samuel Husserl, until 1887. Fascinating accounts of the travels and travails of Faber's mechanical speaking machine can be found in Altick (1978) and Lindsay (1997a).

Faber's speaking machine is notable for many significant advances in the synthesis of human speech. For the first time, all sound production occurs by modulating the passage of airflow through a single tube shape, just like in a real vocal tract, whereas previous models had to coordinate multiple separate mechanisms to create different classes of sound, breaking the analogy between model and physics. Also for the first time, synthesis is controlled by direct manipulation of the vocal tract area function, under physiological constraints inspired by the articulators. Completing the gamut of musical analogies, this was the first keyboard instrument to successfully play speech, and this is scientifically significant because the keyboard mechanism encapsulated the need for coordination of multiple

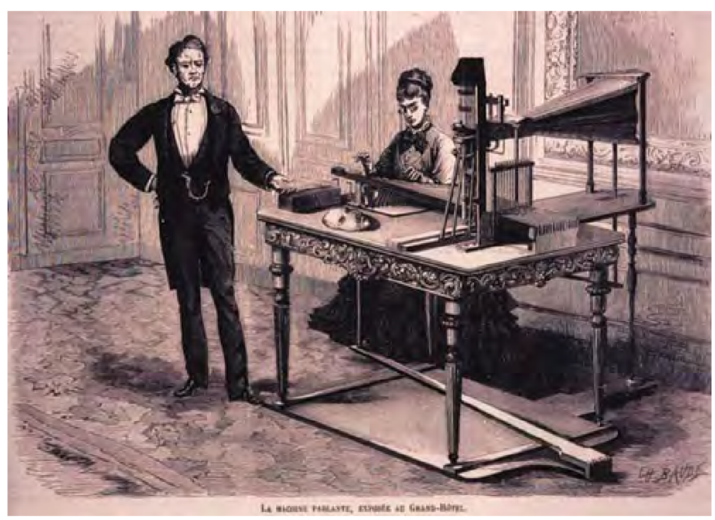

Figure 6. *Faber's "Euphonia," exhibited at the Grand Hôtel in Paris in 1877 by his niece and nephew-in-law. This is one of the few illustrations of the whole speaking machine, showing the bellows driven by a foot pedal, the box containing the mechanism shown in* ***Figure 5****, and the system of rods and levers linking the different articulators to the keyboard. Reproduced from L'Illustration (January 1877).*

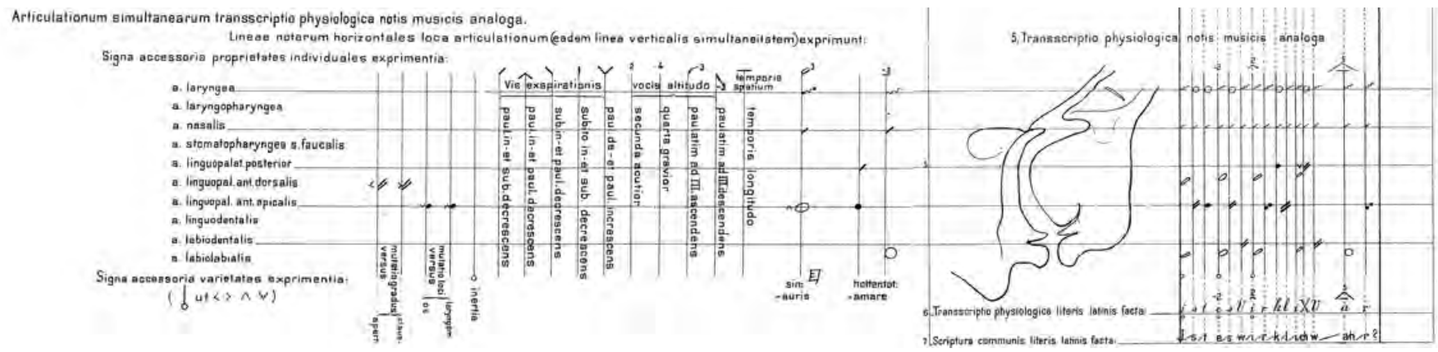

Figure 7. *Techmer's "articulatory score," inspired by early speaking machines. The caption reads "A physiological transcription of simultaneous articulations by analogy with musical notes. The* ***horizontal lines*** *of notes express the place of articulation, whereas the* ***vertical lines*** *express the simultaneity of articulations, with additional signs denoting individual properties of the articulations and their variants." Definitions of different signs are shown on the main chart. Two examples transcribing the Chinese word for "ear" (with a tone) and the Khoisan word for "to love" (with a click) are given to the left of the schematic vocal tract, while a full articulatory score for the German phrase "Ist es wirklich wahr?" is given on the right. Reproduced from Techmer (1880).*

vocal tract articulations, simultaneously and sequentially, to produce each single coherent unit of speech. This was also the first system for unrestricted real-time speech synthesis, allowing the performer to play any desired sequence of speech sounds based on concatenation of discrete phonetic units defined by the keys. It was the first system to systematically incorporate the synthesis of intonation and performed the first song ever sung by a machine. All of these achievements are quite remarkable for their time and represent the zenith of what was accomplished with mechanical synthesis until very recently.

Techmer's "Articulatory Scores"

Faber's "Euphonia" was exhibited across the whole of Europe, Russia, and America for more than 40 years, with hundreds of newspaper accounts of public demonstrations that often also refer back to Mical, Kempelen, and Kratzenstein. Although these speaking machines are now almost forgotten, they were well-known to the major scientists of the day, not only as public spectacles for entertainment but also as genuine objects of scientific inquiry into the nature of human speech (Lindsay, 1997b).

One of those scientists, Friedrich Techmer, developed an entire theory of articulatory phonetics based on his fascination with talking automata. In his "Habilitationschrift" from the University of Leipzig, published in 1880, he reviewed the last century of progress in speech physiology and phonetics and attempted to synthesize many of the principles embodied in the speaking machines, with which he was intimately familiar, into a practical method for phonetic transcription. Techmer (1880) proposed a symbolic alphabet of vocal tract constriction types as the fundamental units of speech. Revisiting the analogy between music and speech, he proposed that utterances could be represented in terms of such constrictions, organized according to an "articulatory score" by analogy with a traditional score in music. Different lines of the score represent independently controllable articulators; single notes represent the activation of specific articulators at particular points in time, whereas chords represent the simultaneous coordination of multiple articulator combinations to achieve particular constrictions. Critically, for the first time, the temporal as well as the spatial coordination of articulatory movements is explicitly represented, and timing becomes an intrinsic part of the representation of speech. The example he gave to illustrate his system, an analysis of the German phrase, "*Ist es wirklich wahr?*" is partly reproduced in **Figure 7**.

Techmer's system can be viewed as the natural extension to speech of Engramelle's earlier proposal for transcribing individual musical performances. It could potentially have been employed to combine Mical's programmable cylinder mechanism with Faber's vocal tract keyboard to yield a fully automated, programmable, mechanical speech synthesizer, realistically simulating the articulations and acoustics of a physical vocal tract to reproduce natural speech.

Unfortunately, popular and scientific interest in mechanical speech synthesis largely vanished with the appearance of

Edison's phonograph in 1877, and attention shifted to acoustic reproduction of the speech signal rather than simulation of the physical system that produced it, using the next new technology.

Modern Recapitulations? Links to the Prehistory of Speech Synthesis

When reviewing the history of mechanical speaking machines, it is remarkable how many of the problems and solutions that preoccupied whole generations of early speech scientists continue to reappear today, with striking parallels in modern theories of speech production and perception.

How is sound produced in the vocal tract? The source-filter model of speech production (Fant, 1960) was never explicitly articulated before the twentieth century, yet in all of the speaking machines there was clear recognition early on of the need for an aerodynamic flow from the lungs, a vibratory or turbulent source from the larynx, and the shaping of sound by a tube. Debates about the generation of sound by the motion of the vocal folds and glottal airflow that began with Dodart and Ferrein continue to be central to current research on aeroacoustics and fluid-structure interaction in speech (McGowan, 1992). Kempelen, Mical, and Faber all experimented extensively with different glottal geometries, making meticulous empirical observations about the influence of the glottal shape and vocal fold tension on airflow, vibration, turbulent noise, and the quality of the resulting sound. They realized the importance of damping the reed with leather and leaving a gap to bias the airflow to avoid irregular vibration and harshness. Experiments on excised larynges and mechanical analogs of the vocal folds continue this vein of inquiry to the present day (e.g., Birk et al., 2017), albeit with the added novelty of computer simulations. A further constant thread has been the realization that the mechanisms of human speech have parallels in vocal production across other species. Casserius (1600) includes comparative anatomies of the larynx in a variety of creatures (cf. Negus, 1929), whereas Vicq d'Azyr (1779) and Kempelen (1791) both consider in detail analogies between sounds and sound production in humans and other animals (cf. Fletcher, 1992).

Which vocal tract shape produces what sound? Understanding the complex many-to-one relationship between vocal tract geometry and acoustics is a perennial theme. Puzzling to all of these investigators was the difficulty in deriving appropriate tube shapes corresponding to particular speech sounds, perhaps because they lacked the ability to see inside the vocal tract but also because of the laws of physics. The low-frequency eigenmodes of the vocal tract are only sensitive to long-wavelength perturbations of vocal tract geometry, so any tube that replicates the macroscopic shape of the quasi-1-D area function regardless of microscopic 3-D details will approximate the same resonances (e.g., Ungeheuer, 1962), which is the origin of the inverse mapping problem. Following the modern tendency toward increasingly overdetailed 3-D vocal tract models, Faber chose to bring his machine closer and closer to an actual vocal tract to be able to exploit physical constraints, whereas Kratzenstein, Kempelen, and Mical perhaps intuitively understood that extreme spatial accuracy or exact reproduction is not always needed, as long as a functionally equivalent tube shape is somehow achieved by hand or box. Kratzenstein's tubes, which sound like vowels but look nothing like vocal tracts, are the classic example.

How is the vocal tract controlled, and what are the underlying goals and units of speech production? For all of the speaking machines, progress toward intelligible synthesis was only made when the temporal dynamics of speech began to be accurately captured, either mechanically, as in Mical's programmable cylinder, or by harnessing human action systems to bootstrap the sequencing of vocal tract movements, as in Kempelen and Faber's wind and keyboard instruments. Mical's attempts at concatenative synthesis using fixed demisyllabic units were less successful than the flexible manual coarticulation that Kempelen's speaking machine allowed. The solution afforded by Faber's keyboard, which succeeded in yoking multiple articulators together in sequence as composite actions to realize the discrete sounds played by each key, is directly analogous to the modern concept of "coordinative structures" in task dynamics by which multiple end effectors are flexibly co-opted to realize a sequence of goals (Turvey, 1990). In the same light, Techmer's "articulatory score," which relates those goals to an alphabet of vocal tract constrictions that function as embodied phonological symbols, has striking parallels with the influential theory of gestural phonology proposed by Browman and Goldstein (1992). Articulatory control and timing have always been central theoretical and practical issues in speech, then as now.

All of these examples demonstrate that mechanical speaking machines were not simply idle amusements but rather can be considered as early attempts at fully embodied theories of speech production, successfully tackling problems that continue to preoccupy research in speech communication into the present and proposing solutions that continue to be echoed in current research. Now that embodiment has returned full cycle as a theme, it can be hoped that the historical undercurrent of productive research on mechanical speaking machines will be more fully appreciated and once again merge with the mainstream.

Acknowledgments

I gratefully acknowledge many years of helpful discussions with Louis-Jean Boë.

References

Altick, R. D. (1978). *The Shows of London.* Harvard University Press, Cambridge, MA.

Birk, V., Döllinger, M., Sutor, A., Berry, D. A., Gedeon, D., Traxdorf, M., Wendler, O., Bohr, C., and Kniesburges, S. (2017). Automated setup for *ex vivo* larynx experiments. *The Journal of the Acoustical Society of America* 141(3), 1349-1359.

Brackhane, F. (2011). Die Sprechmaschine Wolfgang von Kempelens-Von den Originalen bis zu den Nachbauten. *PHONUS* 16, 49-148.

Browman, C. P., and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica* 49(3-4), 155-180.

Casserius, G. (1600). *De vocis auditusque organis historia anatomica*. Baldinus, Ferrara, Italy.

Chapuis, A., and Gélis, E. (1928). *Le Monde des automates*. Griffon, Neuchâtel, Switzerland.

de Milli, N.-C., le Roy, J.-B., Lavoisier, A.-L., Laplace, P.-S, Périer, J.-C., and Vicq d'Azyr, F. (1783). Compte de deux têtes parlantes de M. l'Abbé Mical. *Comptes Rendus de l'académie royale des sciences, 3 septembre 1783*.

Dodart, D. (1700). Mémoire sur la cause de la voix de l'homme, et de ses differens tons. *Suite des mémoires de mathématique et de physique tirés des registres de l'académie royale des sciences de l'année MDCC*, 343-410.

Doyon, A., and Liaigre, L. (1966). *Jacques Vaucanson: Mécanicien de Génie*. Presses Universitaires de France, Paris.

Dudley, H., and Tarnóczy, T. H. (1950). The speaking machine of Wolfgang von Kempelen. *The Journal of the Acoustical Society of America* 22(2), 151-166.

du Moncel (1882). *Le Microphone, le radiophone et le phonographe.* Librairie Hachette, Paris.

Engramelle, M.-D.-J. (1775). *La Tonotechnie ou l'art de noter les cylindres: et tout ce qui est susceptible de notage dans les instruments de concerts mécaniques*. P. M. Delaguette, Paris.

Fant, G. (1960). *The Acoustic Theory of Speech Production*. Mouton, The Hague, The Netherlands.

Ferrein, A. (1741). De la formation de la voix de l'homme. *Suite des mémoires de mathématique et de physique tirés des registres de l'académie royale des sciences de l'année MDCCXLI*, 545-579.

Fletcher, N. H. (1992). *Acoustic Systems in Biology*. Oxford University Press, Oxford, UK.

Gautier d'Agoty, J., and Duverney, J. G. (1745). *Essai d'anatomie en tableaux imprimés qui représentent au naturel tous les muscles de la face, du col, de la tête, de la langue et du larinx, d'après les parties disséquées et préparées, comprenant huit grandes planches en couleur avec tables qui expliquent les planches*. Chez Gautier, Paris.

Hémardinquer, J.-J. (1961). Machines parlantes peu connues du XVIIIème siècle. *Revue du son* 103, 381-385.

Kempelen, W. (1791). *Mechanismus der menschlichen Sprache, nebst der Beschreibung seiner sprechenden Maschine*. van Degen, Vienna.

Kircher, A. (1650). *Musurgia universalis sive ars magna consoni et dissoni in X libros digesta*. F. Corbelletti, Rome.

Kratzenstein, C. G. (1780). *Tentamen resolvendi problema ab Academia scientiarum imperiali petropolitana ad annum 1780 publice propositum: 1. Qualis sit natura et character sonorum litterarum vocalium a, e, i, o, u, tam insigniter inter se diversorum; 2. Annon construi queant instrumenta ordini tuborum organicorum, sub termino vocis humanae noto, similia, quae litterarum vocalium a, e, i, o, u, sonos exprimant, in publico Academiae conventu, die XIX septembris 1780, praemio coronatum*. Imperial Academy of Sciences, St. Petersburg, Russia.

Liénard, J.-S. (1967). Reconstruction de la machine parlante de Kempelen. In *Proceedings of the 4th Conference of Acoustics*, Budapest 1967.

Lindsay, D. (1997a). Talking head. *Invention & Technology* 13(1), 56-63.

Lindsay, D. (1997b). *Madness in the Making*. Kodansha America, New York.

McGowan, R. S. (1992). An aeroacoustic approach to phonation. *The Journal of the Acoustical Society of America 83*, 696-704.

Mersenne, M. (1636). *Harmonie Universelle*. S. Cramoisy, Paris.

Negus, V. E. (1929). *The Mechanism of the Larynx*. The C. V. Mosby Company, St. Louis, MO.

Sauveur, J. (1700). Acoustique. *Histoire de l'académie royale des sciences de l'année MDCC*, 131-140.

Séris, J.-P. (1995). *Langages et machines à l'âge classique*. Hachette Supérieur, Paris.

Techmer, F. (1880). *Zur vergleichenden Physiologie der Stimme und Sprache. Phonetik*. Engelmann Verlag, Leipzig, Germany.

Turvey, M. T. (1990). Coordination. *American Psychologist* 45(8), 938-953.

Ungeheuer, G. (1962). *Elemente einer akustischen Theorie der Vokalartikulation*. Springer Verlag, Berlin.

van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *The Journal of Speech and Hearing Research* 3(1), 227-244.

Vaucanson, J. (1738). *Le Mécanisme du flûteur automate, présenté à messieurs de l'académie royale des sciences*. J. Guérin, Paris.

Vicq d'Azyr, F. (1779). Mémoire sur la voix. De la structure des organes qui servent à la formation de la voix considérée dans l'homme et les différentes classes d'animaux, et comparés entre eux. *Suite des mémoires de mathématique et de physique tirés des registres de l'académie royale des sciences de l'année MDCCIX*, 178-206.

BioSketch

Gordon Ramsay is an assistant professor in the Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, and also director of the Spoken Communication Laboratory at the Marcus Autism Center. He received his PhD in electronics and electrical engineering from the University of Southampton, UK, and was a Marie Curie Research Fellow at the Institut de la Communication Parlée, Grenoble, France, for two years. His research focuses on mathematical models of speech production and perception and early vocal development in infancy.