Péter Nikléczy & Gábor Olaszy

KEMPELEN'S SPEAKING MACHINE FROM 1791: POSSIBILITIES AND LIMITATIONS. (RECOVERING A 200 YEAR-OLD TECHNOLOGY)

INTRODUCTION

The Hungarian polymath Farkas Kempelen constructed the first speaking machine of the world between 1769–1791. This was the first articulatory synthesiser of the world. A reconstruction – based on his book – has been made in 2001 in Budapest, Hungary at the Kempelen Farkas Speech Research Laboratory of the Research Institute for Linguistcs. Kempelen's book was issued in 1791 in Vienna entitled "Mechanismus der Menschlichen Sprache". The reconstructed machine is the exact replica of the device, written and demonstrated by drawings by Kempelen in his book. By this reconstruction it came available to try the same technology for speech generation what Kempelen used more than 200 years ago. Using the reconstructed machine and getting experienced in controlling it, the present day researcher can have an opportunity to check what could Kempelen do with his machine and what kind of limitations he had to face.

1. THE SPEAKING MACHINE

Kempelen studied human articulation processes with a scientist's thoroughness. He made a number of claims that were subsequently demonstrated to be true by scientific methods. He started building his first speaking machine in 1769 and he worked relentlessly on the implementation of his grand plan: "I continued my experiments with undescribable patience and I am amazed to this very day at how I was able to work on my plan for months on end without ever making a step forward. However, the certainty that speech must be imitable strengthened my perseverance …". After twenty-two years of research and development of several machines, he completed the final product in that he then described in his book in 1791. During his initial experiments, he thought that in order to imitate the individual vowels he should operate separate devices like the pipes of an organ (*Figure 1* based on Kempelen's original engravings). In 1773, he was already able to produce four sound types resembling a vowel each with this method, but with respect to [i], he failed. It was then that he hit upon the principle "one voice – one mouth", that is, the idea that the same basic mechanisms (of voice production and vocal tract) are to be used for the production of the various speech sounds. The scientist let four years' results be lost without hesitation to try producing speech with this new approach. He wrote about this in his book as follows: "… I had to start everything anew, but I did not mind the effort or the costs because I thought that the six letters that I gained by them and that were to shed light on my new dim path in what followed, amply rewarded me. But the matter did not remain at that; with the tiresome work of my own hands, I constructed a lot more and threw them away again. … If I wrote down all my failures in as much detail as the above, my book would easily run into another extra volume … Let it suffice to say that, all in all, I threw away as much machinery with ease as would be hard to carry off even with the help of a strong horse". After many years of experimentation, the idea of a serial machine fitted with a bellow (see *Figure 2*, based on Kempelen's original engravings) and following the principle of human sound production finally took shape.

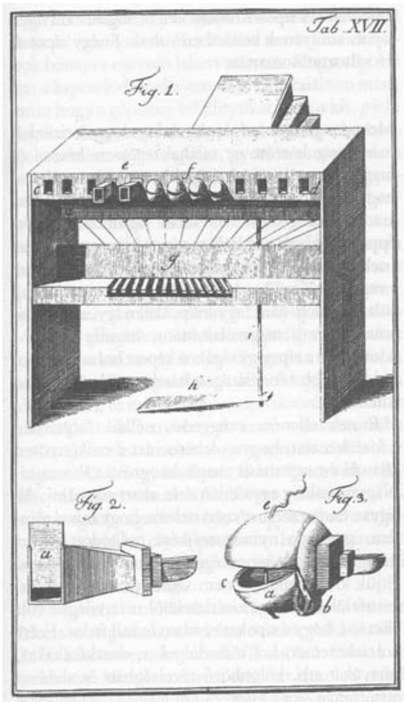

Figure 1: Kempelen's organ-like machine for the generation of vowels

From the bellow, air flowed into the "voice box", an air-tight wooden element (11 x 7 x 6 cm). Within the box were to be found a construction of a vibrating reed for voice production, as well as two valves (in fact 2 small wooden boxes with a cover) on the right side one and on the left side one, the cover of which could be opened by the help of keys on top of the "voice box" (the valves were shut in their position of rest). As a continuation of the vibrating reed, Kempelen installed a tube corresponding to the pharynx, from which he opened two nostrils (these protruded upwards as two additional little tubes). That tube disembogued into a rubber funnel (articulation channel) that stood for the oral cavity and through which voice left the machine. It was by altering the shape of that funnel by hand that sounds resembling the various vowel qualities could be produced.

2. The reconstruction

The aim of this reconstruction was threefold. Firstly we are speech researchers and also deal with speech generation, secondly our laboratory is named by Kempelen, thirdly we wanted to recover a 200 year old technology to see what Kempelen coud do with his machine at that time. The reconstruction was based mainly on his 210 years old book (both the original version and also the translated into Hungarian one). Preliminary studies have been carried out also at the *Deutsches Museum,* Munich, and at the *Hochschule für angewandte Kunst*, Vienna. Kempelen had built a number of machines during his experiments. One such earlier machine of his is kept in Munich, in the *Deutsches Museum*.

This replica was designed and built by the authors of the present paper. Our aim was to build a copy of Kempelen's last and final version of the machine, the one he described in his book in detail. What we wanted to make was not just an exhibition piece but a

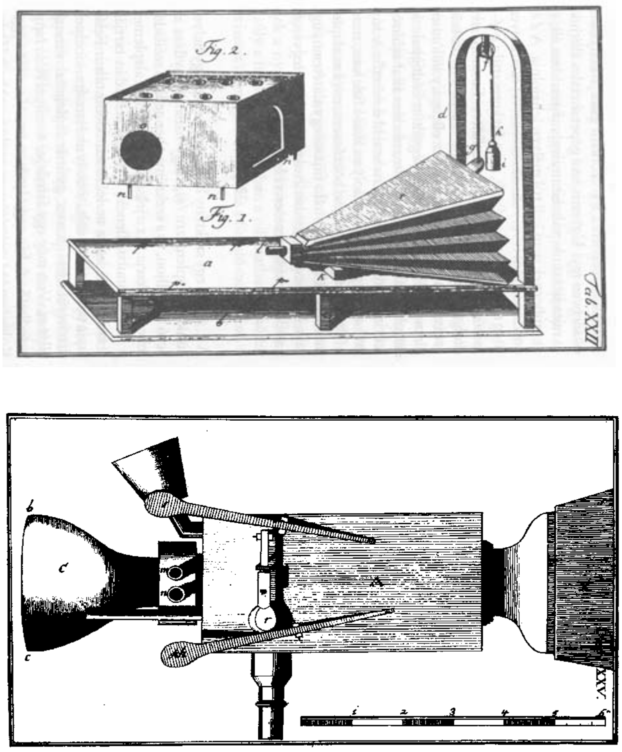

Figure 2: The final serial-order machine imitating the mechanism of human

machine that actually worked (Nikléczy/Olaszy 2003). This was important for us for a number of reasons. If the machine works, we can produce sounds (or even sound sequences) with it. We can try the ways of producing sounds that Kempelen wrote in his book. Thus we can go back 210 years and study the working of one of the most precious instruments of the

Baroque period. The acoustic patterns of the machine's speech can be studied by today's sophisticated signal processing methods and prove or disprove Kempelen's claims by measurement data. It was another important aspect of our decision that we were to present researchers of speech and of Kempelen with an instrument that had been unavailable so far. We took it to be an important task in terms of the history of science to contribute to our knowledge of the beginnings of phonetic research. And finally, we wanted to erect a monument, as it were, to the memory of Farkas Kempelen and his work in phonetics.

The machine was reconstructed (*Figure 3*) in its original dimensions (100 x 40 x 40 cm). During the reconstruction work, we had to pass through dead ends similar to the ones Kempelen had to experience. This was because what we had to replicate was a 210-year-old mechanical instrument, which did not physically exist, and even though we knew its basic features from the contemporary description, certain details of how to build it were missing. Here are a few examples. The air-tight attachment of the bellow or the design of its valve was not described by Kempelen.

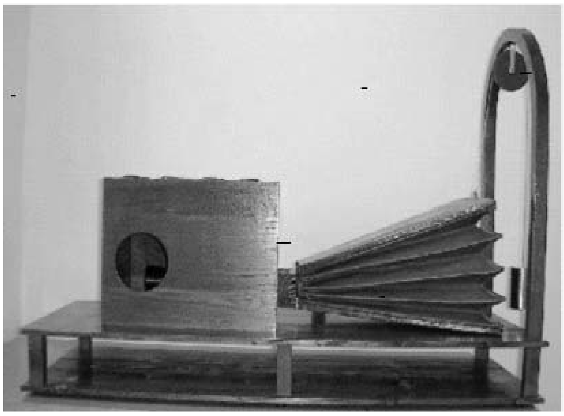

Figure 3: The working replica of the speaking machine built in 2001.
The side hole serves for reaching into the inner part by the left hand

The connection between the keys that operate the valves and the valve-caps had to be established such that the voice-box remains as close to air-tight as possible. The hermetic sealing of the voice-box is important because the air flowing in from the bellow has to result in an increase of the pressure of the air that either makes the reed vibrate or vigorously passes through one of the side valves (if they are opened). If the voice-box leaks and some of the air escapes, the machine does not produce the appropriate sounds. Another problem was caused by the determination of the size of the supplementary bellow that serves to amplify the release of voiceless stops and the force of the backlash spring. This is because, with unsuitable size and spring pressure, the apparatus does not get filled with air or does not give appropriate supplementary pressure for stop release. The reconstructed "voice-box" is shown in *Figure 4.* A similarly grave problem was to find the appropriate solution to how to shape the vibrating reed that produces voice. We have, again, followed Kempelen's detailed description and drawings, but we also used today's technology (for having voice spectrograms displayed, determining frequency components, etc.). We made and tried several reeds. (We had to build each into the machine separately, itself not a simple task given that the reed is located in the most central interior part of the machine.) The size, thickness, and way of fixing of the reed all bear upon the quality of the voice produced. Kempelen admitted that he himself had had difficulties with making a reed producing the appropriate vibrations. He could only guess that, in order to produce speech sounds, he needed a fundamental vibration that resembled the air vibrations produced by the movement of vocal cords. To control the machine both hands are needed. The right hand (five fingers) are used to press the three keys and to close or open the two nostrils, also the right arm is used to press the bellow down for producing air pressure for the wooden (5) box. The left hand is used to change the shape of the rubber "mouth". One way of it is changing the form of the rubber (by pressing sideways), other possibility is to cover the direct output in different angel scales, it is also possible that we put inside the "mouth" some fingers and by this change the figure of the "articulation channel".

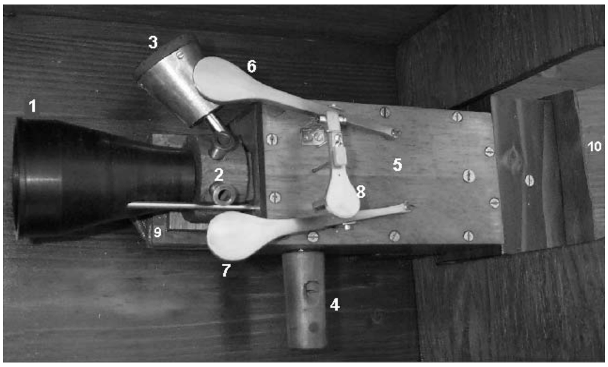

Figure 4: The sound production part in Kempelen's machine. 1- robber funnel ("mouth"), 2- nostrils, 3- resonator of [f]-like sounds, 4- resonator of [Σ]-like sounds, 5- box with the vibrating reed and with the valves for the two side resonators, 6,7- keys to open the valves inside, 8- key of [r], 9- additional bellow for increasing the pressure for stops, 10- connection of the bellow and hole for the right hand

3. Possibilities and limitations

The construction of this machine follows in its main parts the human sound production mechanism. The articulation channel is realised, the jaw and mouth movements can be imitated by changing the figure of the rubber "mouth". Tongue movements cannot be realized.

The two resonators for the spirants theoretically give the possibility to produce spirants, but in practice there are many limitations. All these mean that sounds having simple articulation can be produced quite well. Those sounds the production of which needs complex articulation movements can not be produced. Let us see the possibilities in detail. By single vowels mainly the left hand work is important. The right hand gives only the air pressure and covers the nostrils. Single vowels were produced relatively easily with the machine according to Kempelen's instructions. The easiest is the [a:]. It succeeds always. For producing this sound the nostrils have to be covered and the bellow has to be pushed down. The left palm is far away from the output (there is no obstacle against the voice coming out from the "mouth”). Similarly the [ɔ, o, u] could be produced in high probability by covering the output of the "mouth" with the palm step by step. The most problematic vowel was the [i]. For producing this the output have to be strongly covered by the fingers, only the first finger must be kept slightly away, so a very narrow aperture is formed on the output of the "mouth". Parallel with this much greater pressure have to be given by the bellow as it was needed by the other vowels. As we see production of [i] is much more complicated as that of the other vowels. Therefore it is impossible to produce the [i] sound in sound sequences. The formant structures of the vowels we produced are shown in Figure 5. How to generate consonants? Of the fricatives, the easiest to produce are [ʃ], [f], since these have separate keys ("mouth" must be closed by the left palm). Their intensity compared to that of the vowels, however, falls short of the required value due to the structure of the machine, a fact that makes their inclusion in sound sequences rather awkward. Hence, in sequences containing fricatives, the intensity of the latter is a lot lower than would be desirable, i. e. the vowel covers the

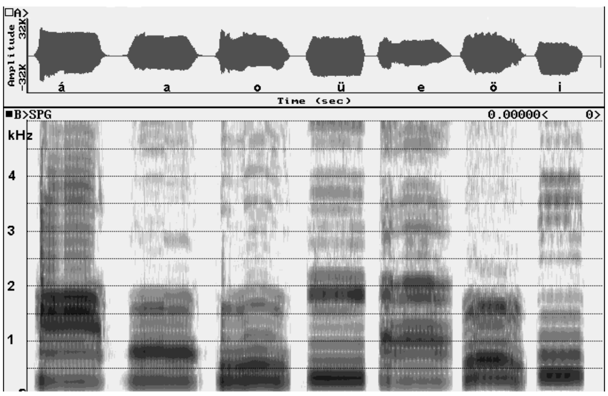

Figure 5: Oscillograms and spectrograms of vowels produced by the reconstructed machine (á=[a], a=[ɔ], o=[o], ü=[y], e=[ɛ], ö=[ø], i=[i])

consonant, and it cannot be heard in many times. One possible solution is to give higher pressure by the bellow in this case, but it is easy to see that the pressure cannot be controlled by a bellow so quick as it would be desirable by a consonant-vowel conjunction. Theoretically the voiced pairs of [ʃ] and [f], the [ʒ], and [v] can be produced by pressing the appropriate key (to produce the noise part of the sound) and let the "mouth" slightly opened (to cover it by the palm and let a narrow aperture between the two middle fingers). Such a configuration can imitate for example the labiodental articulation position in the case of producing the [v]. So, the voiced sound part will be mixed with the noise of the side resonator. In practice the [ʒ] can be produced as a single sound but not in sound sequences. The different sound intensities between the voiced part and the unvoiced make the sound not very recognizable. Sibilant [s] can be produced by the machine in a low frequency form only. One solution is to open both valves parallel. The result is an [s] like weak sound. High frequency voiceless [s] can not be produced. The third key serves for the production of [r]. Of the trills, it is only the uvular type that can be produced (this is quite adavantageous for German), given that the metal pin has to touch the vibrating reed and, while it takes over its vibration, it also hampers the movement of the reed to some extent, due to its own weight. Apical [r] cannot be produced at all.

Nasal sounds [m] and [n] can be produced in good accuracy. The "mouth" have to be closed totally and the nostrils have to be opened, while a voice sound is produced. If only one of the nostrils is opened, an [n] like sound, if both, an [m]-like sound will be heard. Nasalised vowels can be produced if the nostrils are kept open, while a vowel is generated. Nasal stops, like [ɲ] cannot be produced at all.

What is the situation by the oral stops and affricates? Of the stops, the most adequate imitations can be achieved for [b], [p]. This is not a surprise, because the bilabial closure can be imitated by total closing the "mouth" (by the left palm), – this is the first phase of the stop – and after that releasing the close quickly – this represents the burst. A relative higher pressure must be given by the bellow, and attention has to be paid to the supplementary small bellow too. If this small bellow is filled in with air the closure must be released quickly. To sum up, practice is needed to produce these sounds. The other stops, [d],[t],[g], [k], [ɟ], [c] can not be produced in good quality by the machine. This is because by these stops the closure is produced by the tongue and there is no imitation of the tongue in the machine.

The sound [l] can be produced is good quality. The quick movement of the tongue can be imitated by the quick movement of the fingers inside the rubber funnel.

4. Producing sound sequences

After some practice, we produced short sound sequences with the machine, in Hungarian (*mama* 'Mam', *papa* 'Dad', *sás* 'sedge', etc.). The oscillograms and spectrograms of Hungarian *mama* and *papa* are shown in *Figure 6*.

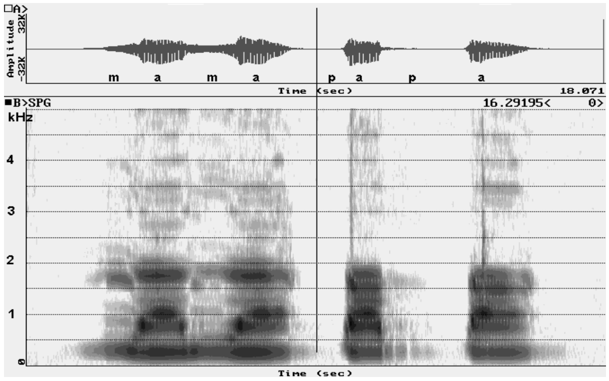

Figure 6: The structure of the sequences mama *'mam' and* papa *'dad' produced by the reconstructed machine (oscillograms and spectrograms)*

Short phrases in German (*es war* 'it was'), French (*je t'aime* 'I love you'), and English (*I go*) were also generated (see the spectrograms in *Figure 7, 8, 9*).

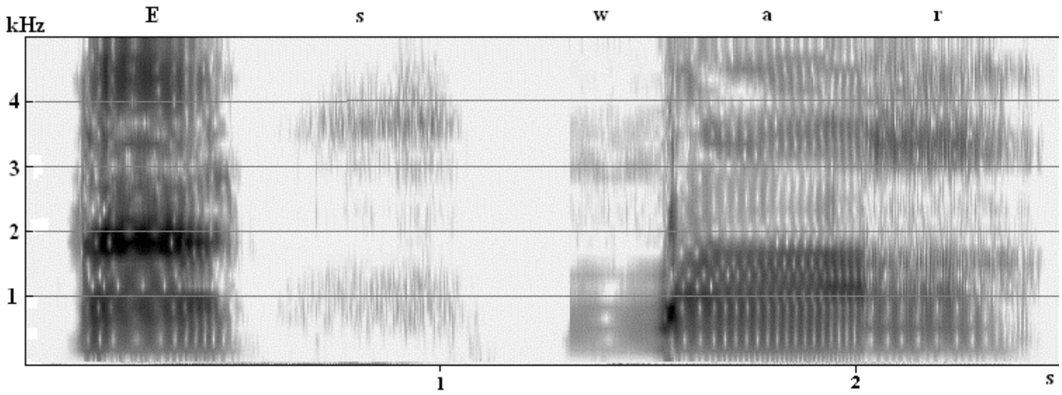

Figure 7: The spectrogram of the German sound sequence [ɛs vaːʀ] produced by the reconstructed machine

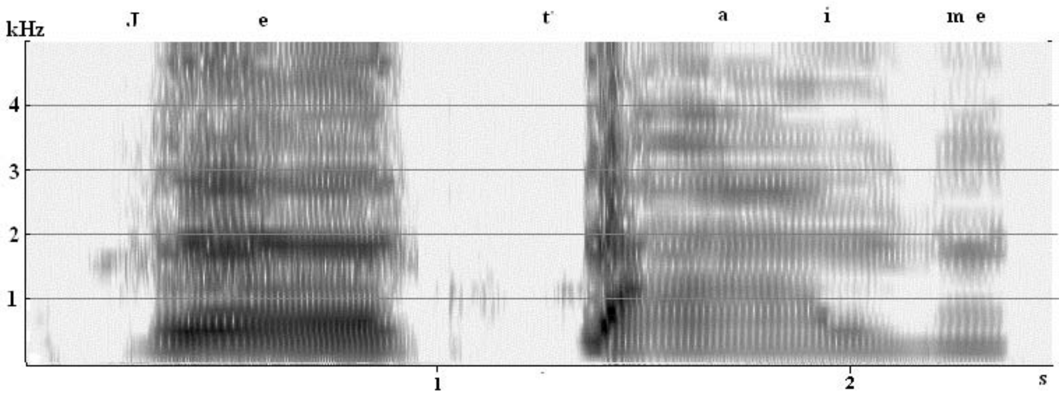

Figure 8: The spectrogram of the French sound sequence [ʒø tɛm] (I love you) produced by the reconstructed machine

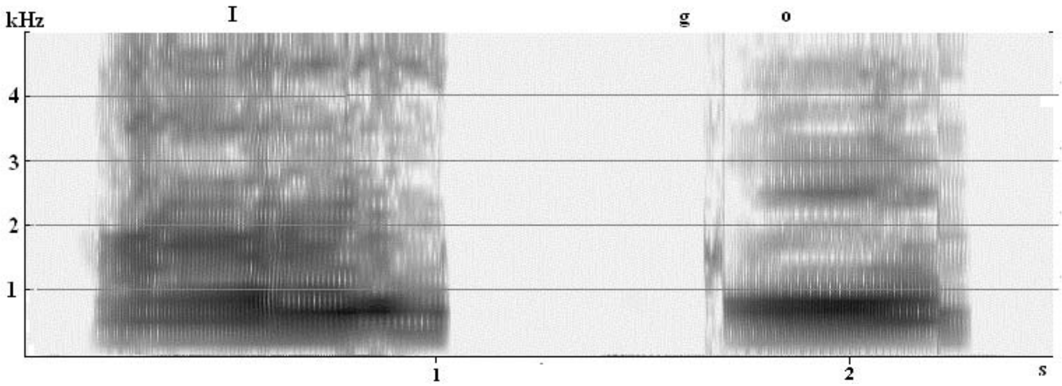

Figure 9: The spectrogram of the English sound sequence [aj go:] produced by the reconstructed machine

5. CONCLUSIONS

Our results have confirmed that Kempelen must have been able to produce sounds and sound sequences with his mechanical machine on the basis of the guidance he gave in his book. The "synthetic speech" we were able to produce was ragged and slow since to produce each sound we had to perform the appropriate movements with both hands in a split second (today's average speech rate is 13–16 sounds/s for Hungarian). Kempelen himself mentioned this physical limitation of the method. Hence, the machine is only suitable for producing short utterances. Sound quality was furthermore influenced by the actual speech sounds that the sequence to be produced contains.

ACKNOWLEDGEMENTS

The authors thank all the help that has been given to the reconstruction by the Deutsches Museum, Munich, and the Hochschule für angewandte Kunst, Vienna. This reconstruction was sponsored by Millenáris Kht., and the reconstructed machine was subsequently exhibited in Budapest for more than a year as part of the exhibition "Dreamers of Dreams; World Famous Hungarians" in 2001–2002.

REFERENCES

Gósy, M. 2000 On the early history of Hungarian speech research, *International Journal of Speech Technology* 3: 154–164.

Kempelen, F. 1791 *Mechanismus der menschlichen Sprache*, Wien.

Kempelen, F. 1989 *Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépének leírása (XXVII metszeti rajzzal)*, Budapest. [The mechanism of human speech and a description of the author's speaking machine (with 27 copperplates)].

Nikléczy, P./Olaszy, G. 2003 A reconstruction of Farkas Kempelen's speaking machine, *Proceedings of Eurospeech 2003, Vol 2*: 894–897.

Péter Nikléczy, Gábor Olaszy

Kempelen Farkas Speech Research Laboratory, Research Institute for Linguistics, Hungary

e-mail: nikleczy@nytud.hu, olaszy@nytud.hu