Perception tests with a replica of von Kempelen's 'speaking machine'

F. Brackhane

Saarland University Saarbrücken, Germany, Email: brackhan@coli.uni-saarland.de

Introduction

In the second half of the 18th century the austro-hungarian scholar Wolfgang von Kempelen (1734-1804) invented the famous 'chess turk' and other impressive mechanical apparatuses but also a 'speaking machine' with which he wanted to model the human articulatory processes and generate hardware-based articulatory speech synthesis (see Figure 1; all figures show the Saarbrücken replica of the 'speaking machine' or parts of it).

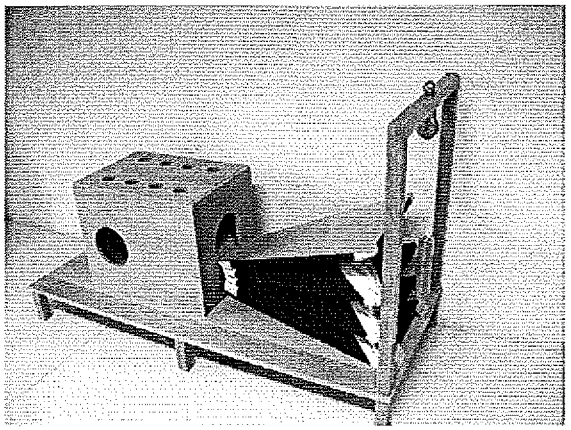

Figure 1: Kempelen's 'speaking machine'. The part on the right consists of the bellows, representing the lungs, on the left the resonator chest that covers the windchest.

The 'speaking machine'

This machine was based upon a reed pipe (as used in pipe organs, similar to the embouchure of a clarinet, see Figure 2) simulating the human vocal folds. Instead of a regular resonator von Kempelen used a rubber funnel which represented the human mouth (see (1) in Figure 3).

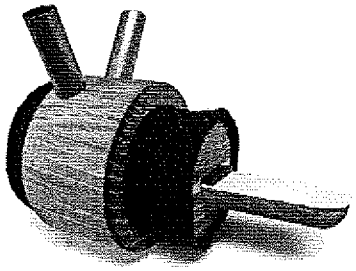

Figure 2: The 'historic' reed pipe (combined with the 'nasal tract'). The reed is made from ivory as described by von Kempelen himself.

This 'mouth' usually is completely open and can only be closed by covering it with the palm of the hand. Furthermore the 'oral cavity' does not have any analogies for teeth, tongue, soft palate etc. In its basic position the machine produces an [a]-like sound. Some other vowels like [o], [u] and [ɛ] can be approximated by modifying the closure of the 'mouth'.

To produce the bilabial plosives [p] and [t] full closure has to me made with the palm of the player's left hand and than suddenly released. Other consonants can not be synthesized in articulatory correct manner because of the lack of all places of articulation. The plosives [t, d, k, g] can be imitated by 'articulating' a bilabial plosive, which can be perceived as the intended plosive by listeners. Nasal consonants can be produced by a 'nasal tract' which von Kempelen installed between the reed pipe and the rubber funnel. By covering one or both 'nostrils' of this part, some nasal quality like [m] or [n] can be effected when the 'mouth' is closed.

The aggregate of reed pipe, nasal tract and mouth funnel is fixed to the side of a windchest (a small box made from wood, see Figure 3). On the other side the bellows are situated which represent the lungs (see Figure 1). To be able to generate fricative sounds von Kempelen invented two separate small generators which synthesizes the fricatives [ʃ] and [s]. To produce an [ʃ] he used the embouchure of a recorder which was cut off and adapted in such a way that no note was to be heard any longer but only a sort of fizzling which is reminiscent of the human [ʃ] sound. This generator is fixed on the left side of the windchest and played via a small valve inside the windchest which is activated by a lever on the top of the windchest. For [s] von Kempelen build a similar generator which conducts the air to a broad but very cramped aperture. This second generator is fixed on the right side of the windchest and operated similarly to the first one.

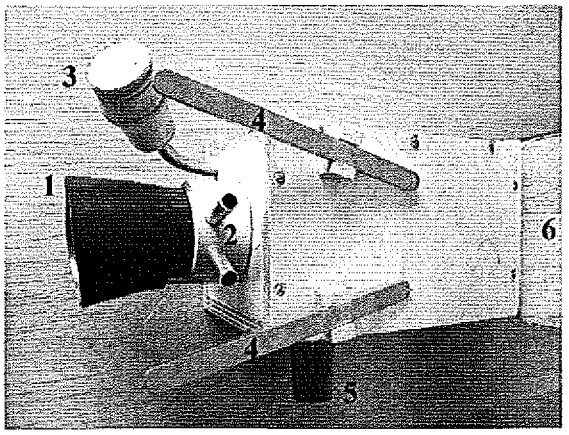

Figure 3: The windchest No. 1. From left to right: Rubber funnel, representing the mouth (1), 'nasal tract' (2) with two brass tubes as nostrils, above the generator for [s] (3). On the top of the windchest the leavers for the two fricative genarators (4), below the [ʃ] generator (5). Right the bellows (6).

Speech quality and replicas

In his book 'Mechanismus der menschlichen Sprache' [1] von Kempelen described the construction of the 'speaking machine' intricately. But in several details more clarity would have been preferable. Several contemporaries published descriptions too, which reported as von Kempelen did too, that the machine sounded very realistic, like the voice of a child aged three to six [1, 3, 4, 5].

The original 'speaking machine' is lost so that it is not possible to verify these reports, but an apparently very old replica which sometimes is attributed to von Kempelen himself is hosted in the Deutsches Museum in Munich (Germany). However this machine differs from von Kempelen's descriptions in various aspects.

Despite the aforementioned authenticity of the auditory impression it is impossible to generate most speech sounds in a human-like manner and quality because of constructional details of the machine which do not represent the human vocal tract in a realistic manner. Notable exceptions are bilabial stops, bilabial nasals as well as open vowels which allow to synthesize words like 'mama' and 'papa'. Since the publication of the 'Mechanismus' many were engaged in exploring the machine's possibilities and how to improve them [6, 7]. Various replicas of the 'speaking machine' were built too, many since the 1960s [8, 9, 10].

From 2007-09 we constructed a replica ourselves. This replica first was built with a windchest made from plywood which had bigger dimensions (85 x 140 x 80 mm) than described by von Kempelen (see Figure 3) because of constructional misapprehensions. The construction of a reed pipe like von Kempelen described it was not possible by the time so a modern reed pipe made from lead and a narrow brass reed with a fundamental frequency (F0) of 212 Hz was used.

In 2008 we were able to built a 'historic' reed pipe, made from oak wood and with a broad ivory reed like the one used by von Kempelen and as a close copy of the reed pipe of the replica in the Deutsches Museum with a F0 of 209 Hz (see Figure 2).

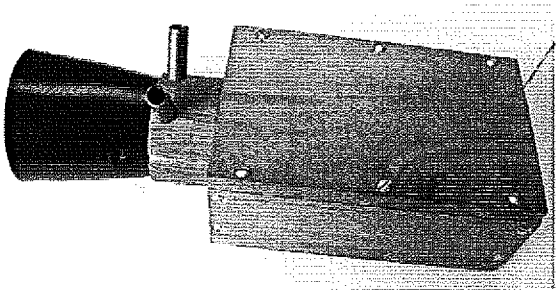

Figure 4: The 'historic' windchest No. 2, (still without fricative generators and levers)

In a third step we constructed in 2009 a second windchest in historic dimensions (40 x 92 x 66 mm) made from nut wood (Figure 4). Additionally, we produced a narrow ivory reed for the modern reed pipe with a F0 of 195 Hz and a broad brass reed for the historic one with a F0 of 258 Hz (see Figure 5). F0 frequencies were not chosen arbitrarily but are heavily dependent on thickness and condition of the materials.

Preliminary results

First tests and recordings with this replica showed that the 'sound quality' was very unlike a realistic child voice, in contrast to the historic descriptions. Other modern replicas did not sound very child-like, either. Thus, the aim was to find out why the original synthesis by von Kempelen had been judged being very realistic. Despite many vague descriptions and measures of single parts of the machine by von Kempelen himself the fundamental difference between historic and modern references could not be explained with technical problems. Reflecting some of the reports of that time a psychological reason seems to be plausible: Several of von Kempelen's contemporaries reported, that they knew the machine's intended utterances before it pronounced them. Even von Kempelen himself reported that it would be easier to understand the machine if people knew what it intended to 'say' ([1, 2, 3, 4]). Furthermore von Kempelen synthesized his utterances imitating very precisely the rhythm of the pronounced words and sentences [2].

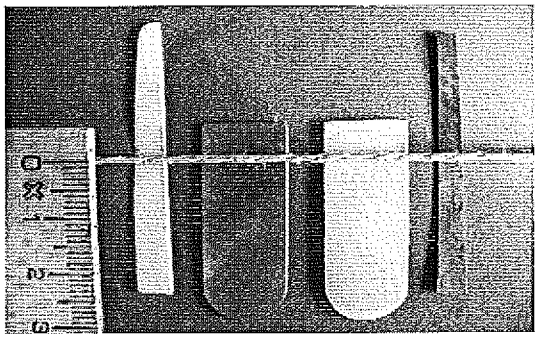

Figure 5: The 4 reeds used, from left: narrow ivory, broad brass, broad ivory, narrow brass. The line indicates the point of fixation in the pipe. Measures in cm.

Perception tests

First Experiment

To explore the machine's possibilities and limitations we performed three perception experiments. 22 subjects were asked in a online experiment to verbally describe various sounds (including giggle, breathing, scratching etc.).

Each subject had to listen to 24 stimuli in randomised order from which only one stimulus (either 'mama' or 'papa') was gene-rated by the 'speaking machine'. 12 of them heard a 'mama' stimulus synthesised by the speaking machine, 10 subjects heard a 'papa' stimulus. The remaining 23 sounds only were used as fillers. The subjects did not know anything about the used stimuli or the work of the author.

17 of the subjects described the 'speaking machine' stimulus they had heard as a human utterance or as human-like (see Table 1). Only 5 subjects described that stimulus as a sound

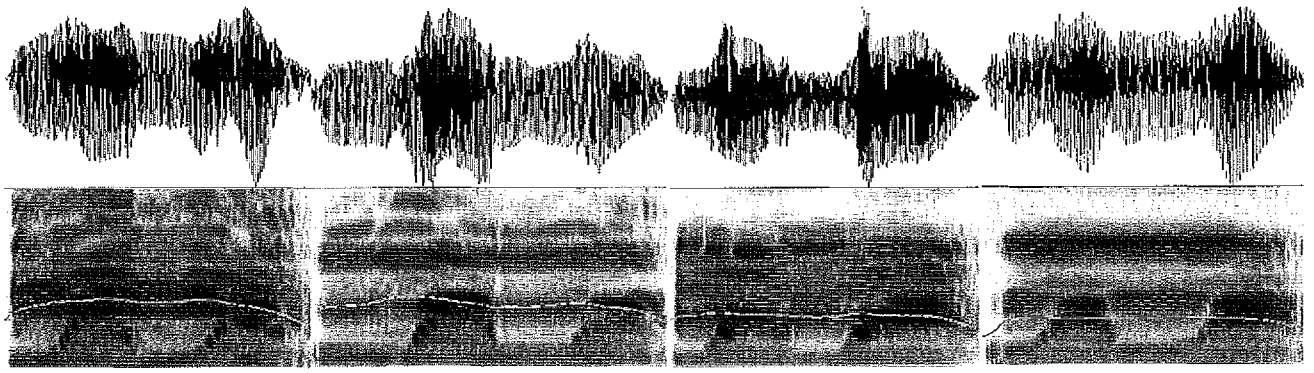

Figure 6: Waveforms, spectrograms (0-6 kHz) and F0 contours of the four settings ('mama', with adjusted durations); from the left to right: ivory reed and 'modern' windchest, ivory reed and 'historic' windchest, brass reed with 'modern' windchest, brass reed with 'historic' windchest.

of a mechanical instrument but speech-like. One subject described the 'mama' stimulus as a noisemaker.

Stimulus	Number of subjects	Judged as 'human'	Judged as 'mechanic'	others
'mama'	12	6	5	1
'papa'	10	8	2	0
	22	14	7	1

Table 1: Descriptions of the 'mama' and 'papa' stimuli in the first experiment.

Second Experiment

In a second experiment the autosuggestive factor in the perception of the speaking machine's utterances was tested. Different recordings of 'mama' and 'papa' were taken with the first windchest and three different reed pipe settings of the Saarbrücken replica ('historic' reed pipe from wood with a broad ivory reed, broad brass reed for the same pipe and a 'modern' reed pipe with a narrow brass reed) and two settings (2 different reed pipes with a F0 of 255 Hz and 328 Hz respectively) from the Kempelen replica in Budapest resulting in ten different sounds.

These stimuli were presented to 12 subjects six times in randomised order via headphones. The subjects were asked to assess each stimulus on a six point scale from 1 (very similar) to 6 (not similar), how akin it is to the voice of a child aged three to six. Results showed no clear indications for one of the reed pipe settings. But the 'papa' stimuli were judged to be more realistic than the 'mama' stimuli, those from Budapest were judged to be better than the stimuli from Saarbrücken (see [11] for a more detailed description).

Third Experiment

In the outcome of the second experiment there were some indications that varying prosodic features on the one hand and changing constructional details on the other hand would allow to generate a more 'realistic' sound with the Saarbrücken replica. We built a new windchest now designed very close to von Kempelen's description (see Figure 4) and made similarly new recordings of the word 'mama' with different prosodic patterns (rising, falling and monotone intonation contour, syllabe durations long-long, short-long and short-short). Both windchests were used with both reed pipes resulting in 36 stimuli ('historic' reed pipe with the ivory reed, modern one with the narrow brass reed, see Figure 6. The two additional reeds both did not work with the 'historic' windchest).

These were presented to 8 subjects via headphones, each stimulus five times in randomised order (in total 180 stimuli). Again the subjects were asked to assess the naturalness of each stimulus on a six point scale. The subjects were told again what they were going to hear. Interestingly, four subjects reported after having finished the experiment that they were convinced that some stimuli were with a voice of a child. This phenomenon had been observed in individual cases in the second experiment, too.

The analysis of the ratings showed that a combination of 'historic' windchest and reed pipe was rated best in most cases (see Figure 7). The seven best rated stimuli (from 2.38 to 3.18) were all based on these settings. All seven worst rated stimuli (4.47 to 4.95) were produced with the modern reed pipe (but with both windchests). The changes in timing and intonation did not affect the naturalness.

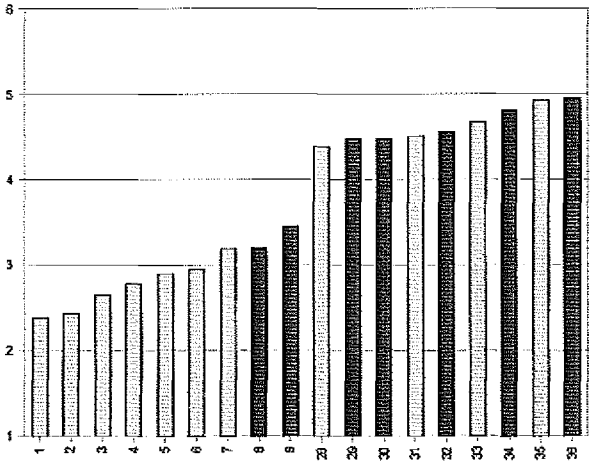

Figure 7: Diagram of the mean ratings of the 9 best and 9 worst rated stimuli of the third experiment. The bright bars indicates stimuli which used the 'historic' widchest.

Discussion

Experiments showed that naïve and non-naïve listeners are able to recognise machine-generated sounds as an authentic child voice. It has been von Kempelens aim to model a true simulation of the human vocal tract. This succeeded only in parts convincingly but however his 'speaking machine' is able to synthesize some short utterances adequately. With regard to the date of origin of this speech synthesis and concerning modern day quality of articulatory speech synthesis this can be referred to a great accomplishment.

There is evidence for using exactly von Kempelen's guidelines while constructing a replica of his 'speaking machine' although nothing can be said about the reasons why just these measures are the ideal ones. In the 'Mechanismus der menschlichen Sprache' Wolfgang von Kempelen described his 'speaking machine' very detailed in parts but cursory in other (not less important) parts without giving any reasons for the accurate measures he used. So reconstructing this fascinating machine will always mean to test out various constructional details which first seem to be less important.

Regarding our third perception experiment particularly the role of prosodic patterns is noteworthy. Monotone stimuli or those which were produced with non-natural high syllabe durations were not rated to be less natural than others. Maybe this can be explained by the condition that listeners of a child's voice did not expect fully correct speech – as von Kempelen already mentioned. So the machine's high-pitched 'child-like' voice enhances the acceptance of the synthesis.

Acknowledgements

Many thanks to Dominik Bauer and the Orgelbaumanufaktur Hugo Mayer in Heusweiler (Saarland) for supporting the reconstruction project in various ways. Thanks too to Jürgen Trouvain from the Institute of Phonetics of Saarland University for helping and advising. We wish also to thank Maria Goszy and Gabor Olaszy from Budapest University for giving useful advices and measures.

References

[1] Wolfgang von Kempelen: Wolfgangs von Kempelen k. k. Wirklichen Hofraths Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine. Degen Wien, 1791, Reprint Friedrich Fromman, Stuttgart, 1970

[2] Jens Peter Köster, Historische Entwicklung von Syntheseapparaten zur Erzeugung statischer und vokalartiger Signale nebst Untersuchungen zur Synthese deutscher Vokale. Hamburger Phonetische Beiträge – Untersuchungen zur Phonetik und Linguistik Bd. 4, Helmut Buske Verlag, 1972

[3] Anonymus, Ueber Herrn von Kempelens Schach= Spieler und Sprach=Maschine. Zweeter Brief. in: Teutscher Merkur, November 1784

[4] Johann Jacon Ebert, Nachricht von den berühmten Schachspieler und der Sprachmaschine des K. K. Hofkammerraths Herrn von Kempelen, Leipzig, Müllersche Buchhandlung, 1785

[5] Karl Friedrich Hindenburg, Ueber den Schachspieler des Herrn von Kempelen – nebst einer Abbildung und Beschreibung seiner Sprachmaschine, Müllersche Buchhandlung, Leipzig, 1784

[6] Robert Willis, Ueber Vocaltöne und Zungenpfeifen, in: Annalen der Physik und Chemie, 1832, Vol. 3, pp. 397 - 437

[7] Charles Wheatstone, Reed organ-pipes, speaking machines etc., in: The scientific papers of Sir Charles Wheatstone, Physical Society, London, 1879, pp. 348 - 367

[8] Marcel van den Broecke, Wolfgang von Kempelen's Speaking Machine as a Performer, in: Studies for Antonie Cohen, Foris Publications, Dordrecht, 1983, pp. 9-19

[9] Jean-Silvain Liénard, From speaking machines to speech synthesis, in: Proceedings of the 12th ICPhS Aix-en-Provence, Vol. 1, 1991, pp. 18-27

[10] Péter Nikléczy, Gábor Olaszy, A reconstruction of Farkas Kempelen's speaking machine, in: Proceedings of the Eurospeech 2003, Geneva, pp. 2453-2456

[11] Fabian Brackhane and Jürgen Trouvain, What makes „mama" and „papa" acceptable? Experiments with a replica of von Kempelen's speaking machine, Proceedings of the 8th International Seminar on Speech Production, Strasbourg, 2008, pp. 329-332