What Makes "Mama" and "Papa" Acceptable? – Experiments with a Replica of von Kempelen's Speaking Machine

Fabian Brackhane & Jürgen Trouvain

Institute of Phonetics, Saarland University, Saarbrücken (Germany)
E-mail: Fbrackhane@web.de trouvain@coli.uni-saarland.de

Abstract

The paper reports on experiments with acoustic recordings of a self-built replica of the historic speaking machine of Wolfgang von Kempelen. Several possibilities of the reed as the glottal excitation mechanism were tested. Perception tests with naïve listeners revealed that the machine-generated words 'mama' and 'papa' were partially recognised as an authentic child voice – as it was also the case in von Kempelen's demonstrations in the late 18th century.

1 Introduction

The "Mechanism" by von Kempelen from 1791 [1] is one of the most important historical descriptions of speech production (cf. [2, 3, 4]). In a separate part of the book he provides a more or less detailed explanation of his speaking machine. Unfortunately, the original no longer exists and a machine at the Deutsches Museum in Munich which is attributed to von Kempelen himself does not work any more. According to the inventor himself, the machine had the voice of a child of 3 or 4 years [1] (p. 442), and this was confirmed by various contemporaries [5-10], though the age of the child was also estimated as 5-6 years [8] and there were also critical views on the capability of the machine [11]. However, it is impossible to know exactly how the voice sounded. There are no precise measures of the thickness of the reed and only visual illustrations of it. In addition, von Kempelen's descriptions of how to generate various sounds are imprecise.

Our aim was to replicate the "convincing" quality of the speaking machine with a self-built replica (see Fig. 1) – based on the experiments and other replicas [12-17]. We tested several possibilities with regard to the length, width and thickness as well as of the material of the reed as the glottal mechanism. The words 'mama' and 'papa' were used as test words – as it was done in von Kempelen's demonstrations [5-9].

Figure 1: *Photograph of the inner life of the speaking machine while being played: palm of the hand on the right forming vowel resonances in front of a rubber funnel (vocal tract); the hand on the wooden windchest (thorax) regulating nasal cavity resonances; (invisible) elbow providing pressure on the bellows (lungs). The reed pipe is located within the "nose".*

2 Acoustic recordings

Acoustic recordings of the test words were made using three different reeds as "vocal folds" (see Fig. 2) made of different materials and measures:

1. ivory (as suggested by Kempelen [1] p. 411, from an elephant as used as material for piano keys) with a vibrating length of 33 mm, 13 mm width and 0.36 mm thickness (F0: 222 Hz),
2. brass (as usually used in modern pipe organs and in one other replica [14]), same measures as the ivory reed (F0: 244 Hz),

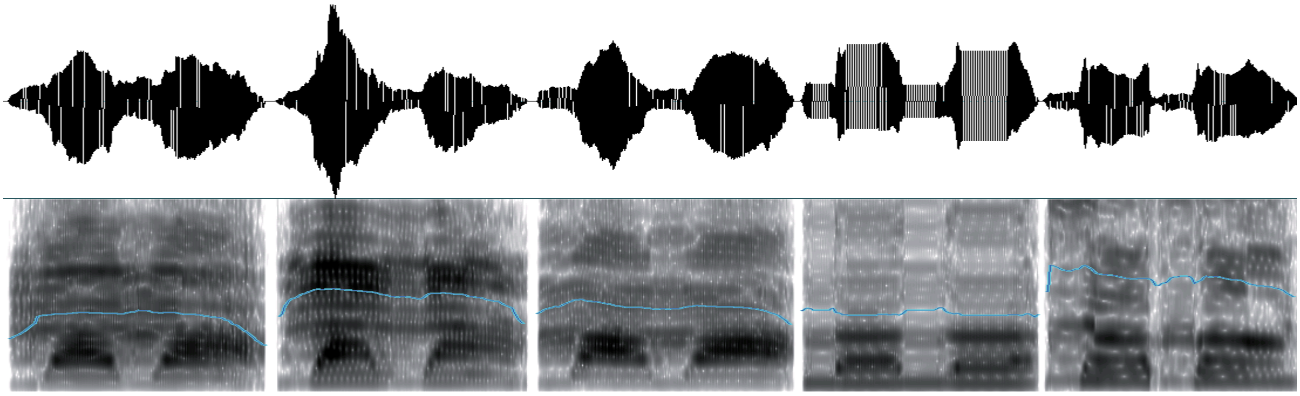
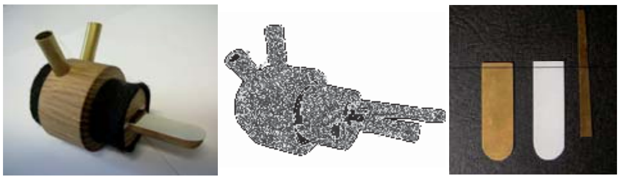


Figure 3: *Waveforms, spectrograms (0-6 kHz) and F0 contours of the five versions of 'mama' (adjusted durations) in the order described in the text.*

3. brass but with a vibrating length of 25 mm, 4-5 mm width, 0.47 mm thickness (F0: 250 Hz). For this smaller reed a separate block (material: lead) and a separate shallot (material: brass) was used which was taken from a dismantled organ. Due to the same superficial measures both blocks (oak and lead) could be used interchangeabily (see Fig. 2a+b).

As a fourth and fifth example, recordings from two different replicas in Budapest [16, 17] were available (F0: 250 Hz and 330 Hz, respectively) also with an ivory reed.



a b c

Figure 2 *a+b: The "nose" of the speaking machine with the ivory reed (a) and a small brass reed (b) as "vocal folds" and two brass tubes as "nostrils" with wood (a) and lead (b). 2 c: a brass reed (left), an ivory reed (middle), a small brass reed (right). The dark line indicates the point of fixation in the pipe.*

The ivory reed was damped with the leather (from a cow) on the upper margin of the shallot as well as on the underside of the reed, so that the reed leather vibrates against the shallot leather as described by von Kempelen [1]. This double usage of leather should result in a special damping effect. In contrast, the brass reeds are not leathered, here only the shallot is leathered as is usual in organ building.

The rubber funnel as the "supra-glottal" resonator shows an effect of "vowel fronting". Comparing the vowels generated with vs. without the resonator (without use of the hand) leaves the auditory impression that the vowel sounds without the "vocal tract" are more "backed", and in case of the small brass reed additionally more "rounded".

Changes of F0 are possible within a range of about four semitones (depending on the material) by increasing or decreasing the air pressure generated in the bellows. A relatively constant F0 was generated for the versions used here (range: half a semitone). Historic witnesses [8, 9] report that there was a prototype of the speaking machine that was able to regulate the effective length of the reed by a wire clamp, so that a control of the fundamental frequency was possible. The version in the Deutsches Museum shows an alternative pitch control mechanism with a lever. In organ building it is possible to apply a tuning wire to a reed with the disadvantage that it is not possible to change F0 while playing (see Fig. 2b but not used here).

In comparison to average F0 values between 272 and 289 Hz for children between 4 and 5 years mentioned in the literature [18], the versions 1-4 show lower pitch (which possibly indicates an older child) but version 5 exceeds these average values by far.

The "user" of the machine needs some practice in playing, so the performance is dependent on the "virtuosity" of the player. The examples used here were performances by the first author. Waveforms and spectrograms of the five versions for 'mama' are illustrated in Fig. 3.

3 Perception tests

Two perception tests were performed to shed more light on the questions (i) whether it is possible to recognise the machine-generate 'mama' and 'papa' as an *authentic child* voice, and (ii) which of the five versions described above is favoured by listeners. All 32 subjects were German native speakers (mean age 25 years, 22 females) and had no immediate personal experience with children's voices.

3.1. Test 1: Recognition

For the first test the authors selected their favourite example of 'mama', to be included in a set of various sounds of ca. 2-4 sec duration (including animal noises, water spilling, laughter, …). The set was presented in randomised order to 10 subjects who wrote down after each sound file what they thought they have heard. The same procedure was repeated with 'papa' with 10 different subjects.

Four of ten subjects described the 'mama' stimulus as spoken by a child whereas the remaining six persons assumed it was a toy or a doll as the sound generator. The 'papa' stimulus was recognised as a child voice by four of ten subjects, two other subjects believed it was a singing voice, and two further persons were unsure whether it was a singing or a speaking voice. For the two remaining subjects the 'papa' stimulus had a mechanical sound.

3.2. Test 2: Naturalness

In the second perception test one example for each of the five versions for 'mama' and 'papa' one was selected (resulting in 10 stimuli). For each word one example was chosen as reference to which the remaining four examples were adjusted with respect to the duration of the four segments: single periods from the stationary part of the vowels were cut out. This resulted in comparable timing of the five 'mama's and the five 'papa's (cp. Fig. 3).

12 listeners (neither participated in test 1) were asked "to which degree does the audio example resemble to the voice of a child age three to six? – Please answer on a 6-point scale from '1' (very good) to '6' (very bad)". Each of the 10 stimuli was presented six times in randomised order via headphones.

Fig. 4 shows for 'papa' a clear preference for both Budapest versions which were rated between '2' and '3' – in contrast to the Budapest 'mama' versions rated between '3' and '5'. Among the Saarbrücken versions the one with the small brass reed was the most preferred with marks between '3' and '4'. The remaining versions were located around '4'. A clear ranking of the three different reeds is not visible so that there is no superiority of one material or one width/length in this small data basis.

The results reveal some large differences between the subjects. Two particular stimuli were indeed consistently judged as child-like by two subjects. However, the same stimuli were also rated '5' to '6' by other subjects.

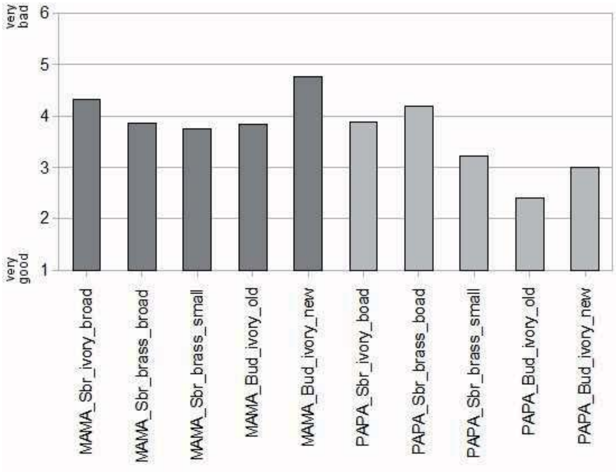


Figure 4: *Mean ratings for the 10 versions.*

As a supplementary result it is noteworthy that some subjects of test 2 mentioned that they thought that some stimuli were spoken by a real child. Those remarks support the partial evidence of test 1 that the machine-generated words were perceived as child-like.

4 Discussion

In test 1 it was shown that naïve humans are able to recognise the machine-generated 'mamas' and 'papas' as *authentic child* voices whereas other subjects believed that the audio files were mechanically generated. This can *partially* explain why von Kempelen was able to give such a convincing performance of his artificial child voice as reported by several contemporaries [12-17]. Other factors to be taken into account are described in [6]: he presented the machine only on request and always immediately after a demonstration of his famous "chess turk" when the public was impressed by his art anyway. He always started the show with the speaking machine with some simple words ('mama' and 'papa') before the audience were allowed to ask him to generate some phrases they wished to hear. This resulted in an auto-suggestive recognition of the words and phrases (a phenomenon which is also informally reported by speech synthesis researchers: after a short time they believe they recognise the phrases they intended to synthesise).

The second test revealed that there is not a great acceptance in terms of naturalness of the stimuli. But probably we would gain similar results if we included samples generated by formant synthesis or articulatory synthesis. A test with those non-concatenative synthesisers should also include some natural reference samples of real child voices as baselines.

Irrespective of the generation method, listening to synthetic speech is a form of listening under adverse conditions. In this light the comment from one of the subjects is illuminating: S/he said if we assume that a healthy child is speaking, the audio impression is bad, but it sounds excellent if we assume that the child has a cold.

Acknowledgements
The authors are grateful to Gábor Olaszy for providing them their audio files and Bill Barry and two anonymous reviewers for their comments on a previous version of the paper.

References

[1] W. von Kempelen. *Über den Mechanismus der menschlichen Sprache. nebst Beschreibung einer sprechenden Maschine*. Wien: J.V. Degen. Facsimile re-print from 1970, Stuttgart: Frommann-Holzboog. 1791.

[2] H. Dudley & T.H. Tarnoczy. The speaking machine of Wolfgang von Kempelen. *Journal of the Acoustical Society of America*, 22: 151-166, 1950.

[3] J.-P. Köster. *Historische Entwicklung von Syntheseapparaten zur Erzeugung statischer und vokalartiger Signale nebst Untersuchungen zur Synthese deutscher Vokale*. Hamburg: Buske. 1972.

[4] B. Pompino-Marschall. Von Kempelen's contribution to the theory of acoustic articulation. *Grazer Linguistische Studien* 62: 137-147, 2004.

[5] K.G. Windisch. *Karl Gottlieb von Windisch's Briefe über den Schachspieler des Herrn von Kempelen nebst drey Kupferstichen die diese berühmte Maschine vorstellen* (herausgegeben von Chr. von Mechel der K. K. und anderer Akademien Mitgliede). Basel: Mechel. 1783.

[6] Anonymous. Ueber Herrn von Kempelens Schach-Spieler und Sprach-Maschine. Zweiter Brief. *Teutscher Merkur*, issue November 1784.

[7] Anonymous. Schreiben über die Kempelische Schachspiel- und Redemaschine. *Hessische Beyträge zur Gelehrsamkeit und Kunst*, 1(3), 1784.

[8] C.F. Hindenburg. *Ueber den Schachspieler des Herrn von Kempelen – nebst einer Abbildung und Beschreibung seiner Sprachmaschine*. Leipzig: Müllersche Buchhandlung, 1784.

[9] J.J. Ebert. *Nachricht von dem berühmten Schachspieler und der Sprachmaschine des K. K. Hofkammerraths Herrn von Kempelen*. Leipzig: Müllersche Buchhandlung, 1785.

[10] J.W. von Goethe. *Goethes Werke*, Herausgegeben im Auftrage der Großherzogin Sophie von Sachsen, IV. Abtheilung, 12. Band: 154, 1893.

[11] Anonymous (Rm.). Review of Windisch, K. G. v.: Briefe über den Schachspieler des Hrn. Von Kempelen. *Allgemeine deutsche Bibliothek*, 58(1): 278-280, 1784.

[12] R. Willis. Ueber Vocaltöne und Zungenpfeifen. *Annalen der Physik und Chemie*, Jg 1832, Drittes Stück: 397-437, 1832.

[13] C. Wheatstone. Reed organ-pipes, speaking machines, etc. Physical Society of London (ed): *The scientific papers of Sir Charles Wheatstone*. London: Taylor & Francis: 348-367, 1879.

[14] M.P.R. van den Broecke. Wolfgang von Kempelen's speaking machine as a performer. In: M. v.d. Broecke, V. van Heuven & W. Zonneveld (eds) *Sound Structures. Studies for Antonie Cohen*. Dordrecht: Foris, 9-19, 1983.

[15] J.-S.Liénard. From speaking machines to speech synthesis. *Proc. 12th Int'l Congress of Phonetic Sciences*, Aix-en-Provence, Vol. 1: 18-27, 1991.

[16] P. Nikléczy & G. Olaszy. A reconstruction of Farkas Kempelen's speaking machine. *Proc. Eurospeech, Geneva*: 2453-2456, 2003.

[17] P. Nikléczy & G. Olaszy. Kempelen's speaking machine from 1791: possibilities and limitations. (Recovering a 200 year-old technology.) *Grazer Linguistische Studien* 62: 111-120, 2004.

[18] D. Ingrisano, G. Weismer & G.H. Schuckers: Sex identification of preschool children's coices. *Folia phoniatrica* 32: 61-69, 1980.