



HAL
open science

An articulated talking face for the iCub

Alberto Parmiggiani, Marco Randazzo, Marco Maggiali, Frédéric Elisei,
Gérard Bailly, Giorgio Metta

► **To cite this version:**

Alberto Parmiggiani, Marco Randazzo, Marco Maggiali, Frédéric Elisei, Gérard Bailly, et al.. An articulated talking face for the iCub. Humanoids 2014 - IEEE-RAS International Conference on Humanoid Robots, Nov 2014, Madrid, Spain. hal-01110293

HAL Id: hal-01110293

<https://hal.science/hal-01110293v1>

Submitted on 27 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An articulated talking face for the iCub

Alberto Parmiggiani¹, Marco Randazzo¹, Marco Maggiali¹, Frederic Elisei², Gerard Bailly² and Giorgio Metta¹

Abstract—Recent developments in human-robot interaction show how the ability to communicate with people in a natural way is of great importance for artificial agents. The implementation of facial expressions has been found to significantly increase the interaction capabilities of humanoid robots. For speech, displaying a correct articulation with sound is mandatory to avoid audiovisual illusions like the McGurk effect (leading to comprehension errors) as well as to enhance the intelligibility in noise. This work describes the design, construction and testing of an animatronic talking face developed for the iCub robot. This talking head has an articulated jaw and four independent lip movements actuated by five motors. It is covered by a specially designed elastic tissue cover whose hemlines at the lips are attached to the motors via connecting linkages.

I. INTRODUCTION

The ability to interact with people in an entertaining, engaging, or anthropomorphic manner is an important part of a robot's functionality [1], [2], [3]. If thanks to its design, a robot adheres to human social expectations, users will find the interaction more natural, enjoyable and useful. Many researchers, are therefore exploring the design space of anthropomorphic robots, trying to endow their creations with characteristics of intentional agents. More and more robots are being equipped with human-specific sensory-motor skills such as multimodal speech production, perception and understanding. Primitive effectors that are used by humans to encode and decode socio-communicative signals, such as eyes, lips, ears and eyebrows, should be patterned to elicit social interaction and user compliance.

For this purpose, the iCub Facility (IIT, Genoa) and GIPSA-Lab (Univ. Grenoble Alpes) jointly developed an animatronic head for the iCub [4], that was completed on October 2013. This paper describes the design and preliminary tests of this system, fondly called “Nina”.

II. ANIMATRONIC TALKING HEADS

An animatronic talking head is a figure that is animated by means of electromechanical devices. Animatronic talking

This work was supported by the French government research program Investissements d'avenir through the Robotex Equipment of Excellence (ANR-10-EQPX-44) and by a UJF/ EmSoC allowance.

¹Alberto Parmiggiani, Marco Randazzo, Marco Maggiali and Giorgio Metta are with the iCub Facility, Fondazione Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy; alberto.parmiggiani@iit.it, marco.randazzo@iit.it, marco.maggiali@iit.it, giorgio.metta@iit.it

²Frederic Elisei and Gerard Bailly are with the GIPSA Lab, Speech & Cognition Dpt., CNRS/Univ. Grenoble Alpes France; Frederic.Elisei@gipsa-lab.grenoble-inp.fr, Gerard.Bailly@gipsa-lab.grenoble-inp.fr

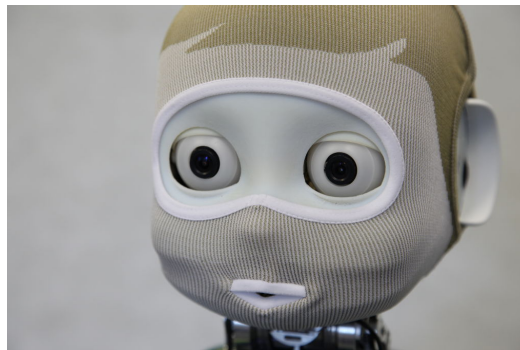


Fig. 1. The figure shows a photograph of the iCub robot with the articulated talking face installed.

heads differentiate in two main categories, which for simplicity, can be called artificial voice production systems and social talking avatars.

Artificial voice production systems are mainly developed to study the bio-mechanics of speech. The idea to produce artificial sounds with physical means that replicate the human vocal tract dates back to the 18th century [5]. This stream of research has been constantly active till very recently: the series of Waseda talkers [6] have human-like organs: lungs, vocal cords, tongue, lips, teeth, nasal cavity and soft palate casted in silicone and rubber. At Sheffield University, Hofe et al. developed Anton [7] consisting of movable tongue and jaw models actuated by servo motors. The motors drive filaments connected to a system of meshes embedded in the silicone tongue body.

Social talking avatars are generally developed to mimic facial movements for articulating speech and displaying emotion. In this category, another distinction can be made between faces that are realistic human-like and those that are not. In this latter class, Kismet [8], [9], [10] was certainly one of the first attempts to develop sociable robots. More recently, two other robotic faces were developed at MIT to study human-robot interaction, namely Mertz [11] and Nexi [12]. Another interesting design approach is the one adopted on the WE-4RII head, later integrated on the Kobian robot [13]. This face is actuated by electric motors that drive the face degrees of freedom (DOF) with timing belt transmissions. Two other noteworthy animatronic faces are iCat [14] and Flobi [15]. Flobi achieves facial displays on a non movable facial surface via a magnetic actuation system. A different approach was adopted by Lin et al. [16] where six required lip postures are obtained with a “minimal” 2DOF mechanism.

Among realistic human-like faces we can cite SAYA [17] proposed by Hashimoto et al.. This face is actuated with McKibben pneumatic actuators, distributed behind the surface of the face to mimic the human facial muscle arrangement. Another interesting animatronic face named Roman [18] has an artificial skin covering a mechatronic head with sophisticated control of the neck, jaw, lips, eyes and brows. Similarly Ishiguro and Nishio [19], Hanson et al [20], [21] developed androids capable of facial expressions with the same layered principle.

III. DESIGN REQUIREMENTS

The standard iCub face has a series of LED, mounted behind the semi-transparent face cover in the mouth and eyebrows area to represent facial expressions (see [4] for details). This system is however insufficient to display complex features of human speech movement; we therefore decided to redesign this subsystem to allow for richer human-robot interactions.

We began the design process by articulating the list of needs of the new mechatronic system. The most important features are listed hereinafter:

- integrate a moving jaw on the head of the iCub;
- integrate mechanically actuated lips capable of replicating the “a”, “i” and “u” lip postures;
- integrate a loudspeaker close to the mouth region;
- integrate a high quality stereo microphones pair on the head;
- decrease the noise levels generated by electronics cooling fans housed inside the head to improve the performance of the audio system;
- modify the eyelids closure mechanism to achieve a complete eyeball closure with a motion of the upper eyelid only.

A. Speech considerations

The necessary DOF should be inspired by the ones of the human face. From a phonetic point of view there are three major movements of the lips: aperture/compression, rounding/spreading and protrusion. The proposed articulated talking head addresses the first two movements while ignoring the third.

The remaining major movements are achieved with five main DOF, defined as follows:

- Jaw_{height} : height of the middle of the tips of the two lower incisors
- X_{right} : horizontal position of the right lip commissure
- X_{left} : horizontal position of the left lip commissure
- Y_{up} : height of the middle of the upper lip
- Y_{down} : height of the middle of the lower lip with reference to Jaw_{height}

Aperture/compression is obtained by three DOF: Jaw_{height} , Y_{up} and Y_{down} . Vertical compression occurs mainly by raising the lower lip while raising the jaw. Jaw oscillation is one of the basic movement for feeding and speech [22], [23]. The “frame then content” hypothesis states that babbling as a precursor of the syllabic structure

of speech emerges from jaw oscillation [24], [25]. This DOF explains more than 50% of the variance of speech-related facial movements [26].

Part of the vertical aperture/compression can also be achieved by intrinsic (orbicularis) as well as extrinsic lip muscles (levator labii, mentalis). This independence between jaw height and lip aperture is quite necessary to properly articulate various sounds: (a) raise the upper lip with raised jaw and lower lip in labiodentals (e.g. [f],[v]) (b) lower the lower lip with raised jaw in alveolars (e.g. [t],[d],[n]) and front vowels ([i], [e]). The latter sounds require a high jaw to ease tongue/palate contact but an open mouth to avoid bilabial contact. Note that the production of post-alveolar fricatives require both raised jaw and open lips.

The horizontal control of the commissures (achieved by the agonist action of X_{right} and X_{left}) is very important for the implementation of rounding/spreading opposition. Like many other languages, French notably exploits a rounded/unrounded contrasts for high-front vowels (e.g. [y], [oe]¹ vs. [i], [e], ²). Note that the production of French post-alveolar fricatives is also produced with rounded lips.

Protrusion is accomplished by pushing the lips forward, making the vocal tract longer. This manoeuvre has the strong global effect of lowering the vocal tract resonances and often complements rounding. Part of this geometric deformation can be achieved by passive biomechanical coupling with lips leaning on teeth. This aspect was not considered on the present system and will be explored in a future version.

B. Using human data

To better suit geometrical and dynamic capabilities of the robot to speech specifications, data recorded from a real human speaker has been exploited [26]. This data gives minimum and maximal displacement of various landmarks around the lips, as well as the jaw displacement, recorded on a female French speaker (See Fig.2). Scaling them to the size of the robot head served as a specification basis for designing the robot new degrees of freedom. The main aspects of this corpus of data are represented in Fig.3 and Fig.4. More in detail the distances in [cm] from the middle of the tips of the two upper incisors of various key points (Fig.3) were used to determine the ranges of motion required for the different joints. The specifications regarding the dynamics of the system (e.g. the required bandwidths of the different joints) were derived from Bode diagrams of the recorded data (Fig 4).

IV. MECHATRONICS

A. Default iCub2 head mechatronics

The standard iCub head houses an on-board PC104 controller, four electronic boards to control the electric motors a gyro sensor and two Dragonfly camera boards. The eyes

¹see http://en.wikipedia.org/wiki/Close-mid_front_rounded_vowel for reference.

²see http://en.wikipedia.org/wiki/Open-mid_front_unrounded_vowel for reference.

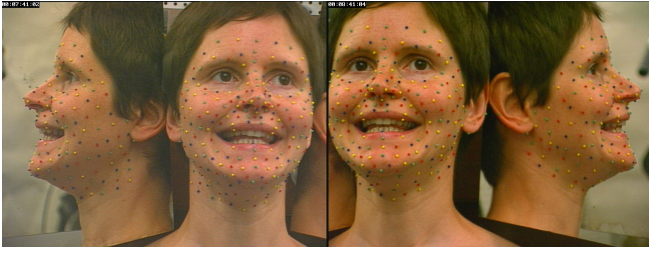


Fig. 2. The figure shows a photograph of a typical recording of a female speaker, with coloured beads marking the moving points.

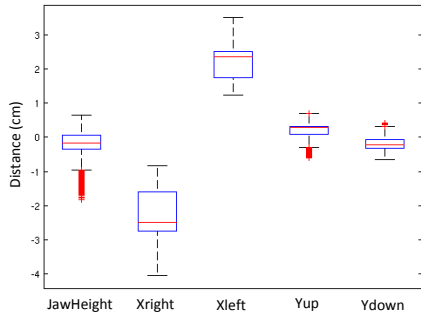


Fig. 3. The figure shows the results of a typical female French speaker recording. The graph shows the distances in [cm] from the middle of the tips of the two upper incisives of various key points. The ranges (difference between 5%-95% percentiles) are: 1.27[cm] for Jaw_{height} , 2.53[cm] for X_{right} , 1.80[cm] for X_{left} (interestingly most people are asymmetrical), 0.68 [cm] for Y_{up} and 0.40[cm] for Y_{up}/Y_{down} . Note the large amount of outliers for Jaw_{height} (due to large mouth openings for some [a] in the corpus) and Y_{up}/Y_{down} (due to complete closures during silences).

have independent pan and tilt DOF and house a video camera used for stereo vision (see [4] for details).

B. Articulated talking head improvements

An articulated jaw with pure rotation was added to the system (see Fig.5), with the motor driving this DOF is housed in the volume behind the neck yaw assembly. Next, the eyes tilt mechanics was completely redesigned to allow the integration of new components; the space was exploited to install a Visaton BF32 loudspeaker [28] and five additional Faulhaber motors [29] (see Fig.6 and Fig.7). On the front

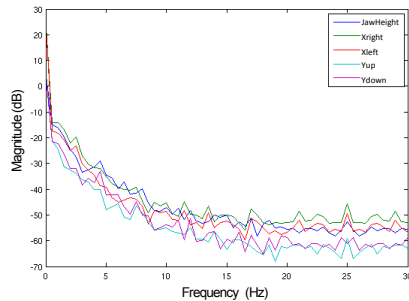


Fig. 4. Power spectrum density extracted from one hour of speech data. The cut-off frequency of lip and jaw movements is around 10-15[Hz]. Ramanarayanan et al. state that 99% of the energy is contained below 21[Hz] for all the articulators [27].

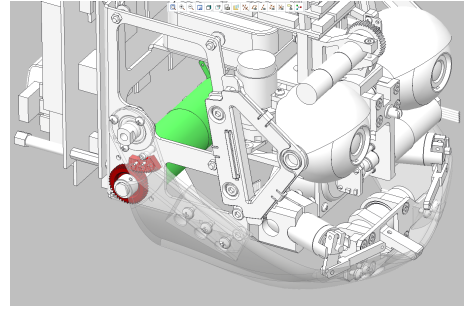


Fig. 5. Jaw mechanism. The figure shows a CAD rendering of the jaw mechanism. The motor actuating the jaw is coloured in green; the gear pair used as the last stage of the transmission is represented in red. Some parts were hidden to clarify the representation.

part of the jaw a motor to control the lower lip vertical displacement via a rod was installed. Three additional motors were fixed to the inner head frame. These control respectively the vertical displacements of the upper lip and the horizontal displacements of the lip corners via three connection linkages. The eyelids were also redesigned in order to obtain their complete closure with the motion of only the upper eyelid (see Fig.8). The main parameters of the facial expression motors are reported in Table I for reference.

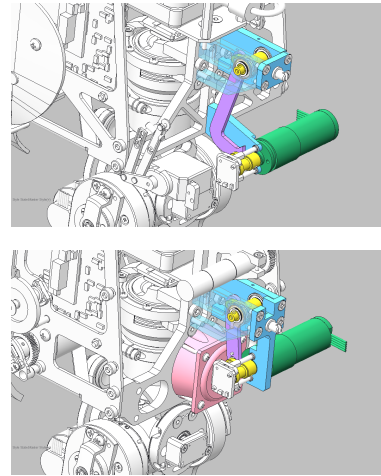


Fig. 6. Eye tilt leverage. The figure shows a CAD rendering of the old and improved eyes tilt leverage lips mechanism. To clarify the representation, corresponding parts in the old and new mechanism have been assigned the same colour. The revision of the eyes tilt leverage allowed to free a large amount of space behind the mouth area. This space was employed to fit the Visaton speaker, and its soundboard (represented in pink).

C. Electronics

Two motor controller boards (each of them capable of controlling up to four brushed motors) were added to the system to control the six additional facial expressions motors. A specific board for the speaker, based on the PCM2704C from Texas Instruments Incorporated (98dB SNR Stereo USB2.0 FS DAC with line-out), was designed and integrated in the assembly. In order to reduce the noise coming from the fans of the pc104 embedded PC, we decided to change the cpu from an Intel[®] Core 2 Duo 2,16 Ghz to an Intel[®]

TABLE I

RANGES OF MOTION. THE TABLE LISTS THE RANGE OF MOTION AND THE LINEAR DISPLACEMENTS OF THE FOUR MOTORS ACTUATING THE LIPS.

| DOF | Jaw | Upper lip up/down movement | Lower lip up/down movement | Left/Right lip side movement |
|--|-------------------|----------------------------|----------------------------|------------------------------|
| Range of Motion | +1°/+11° | -28°/+28° | 0°/+28° | -35°/+35° |
| Corresponding linear displacement [mm] | 18 | 7 | 12 | 13 |
| motor model | Maxon | Faulhaber | Faulhaber | Faulhaber |
| gear ratio | A-Max22-6W+GP22-A | 1319T012SR-IE2-400-14-1 | 1319T012SR-IE2-400-14-1 | 1319T012SR-IE2-400-14-1 |
| max. torque [Nm] | 84:1 × 2.5:1 | 66:1 | 66:1 | 66:1 |
| max. angular velocity [rad/s] | 3 | 0.19 | 0.19 | 0.19 |
| | 4.55 | 20.2 | 20.2 | 20.2 |

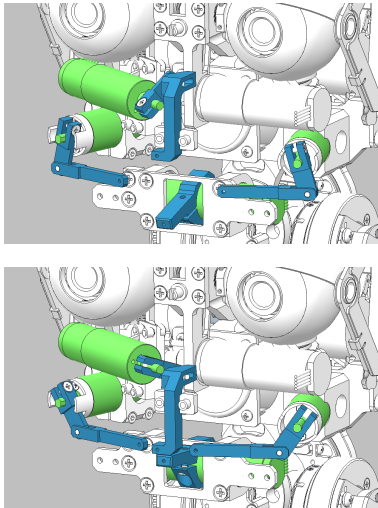


Fig. 7. Lips DOF. The figure shows a CAD rendering of the lips mechanism in an “open” (top) and “closed” (bottom) configuration. The four motors actuating the lips facial expressions are coloured in green; the linkages that convey the motion from the motors to the lips are represented in blue. Some parts were hidden to clarify the representation.

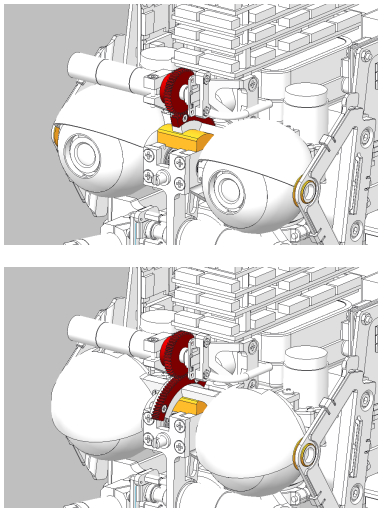


Fig. 8. New eyelids. The figure shows a CAD rendering of the eyelids mechanism in their “open” (top) and “closed” (bottom) configuration. Motion is conveyed with a gear pair represented in red. The eyelids rest on three polymeric bush bearings represented in orange. Some parts were hidden to clarify the representation.



Fig. 9. Textile cover. The figure shows two photographs (rear and front views) of the stretchable textile cover developed for this project.

AtomTMD525. Thanks to the lower power consumption, the heat generated by the electronics was drastically reduced. Consequently we could replace the fans with a low noise version. In the absence of motor commands the signal-to-noise level recorded by the two new high-quality OKM II ear microphones [30] laid above 28dB SPL.

V. TEXTILE COVER

The standard solution to endow mechatronic heads with a “skin” is to install an outer stretchable silicone layer. The moulding process required to manufacture such covers is rather complex, especially with curved surfaces having multiples recesses such as the iCub head covers. The result of these manufacturing operations is not always repeatable. Finally, stretchable silicone-based materials tend to degrade with time; this makes them less suited for systems required to operate for long periods without maintenance.

To address these issues we adopted an alternative solution based on the use of stretchable fabrics. We developed a special Lycra cover for the whole head which is shown in Fig.9. The cover has five openings: the neck, the two eyes, the top back of the skull (for ventilation) and the mouth. The lips thickness was obtained by sewing two layers of stretchy fabric whose corners were connected to the actuating links driven by the four lip motors. The hemline of the mouth was designed to enable a lip protrusion motion when the lips close.

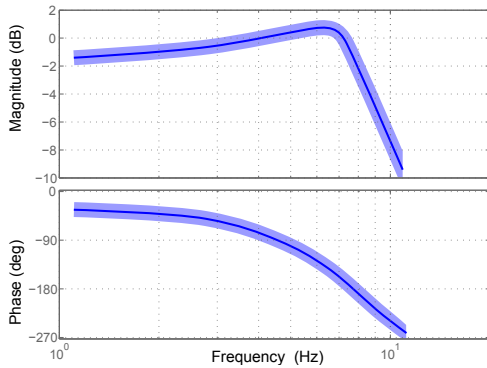


Fig. 10. Jaw Bode diagram. The figure represents an average of the experimental runs conducted to estimate the jaw opening motion bandwidth. The jaw DOF was commanded to track a sinusoidal position reference at frequencies spanning from 1.05 to 10.5 [Hz]. As can be seen, the -3dB intersection occurs approximately at 8.5[Hz]. The banded area represents the average measurement accuracy of the jaw position sensor.

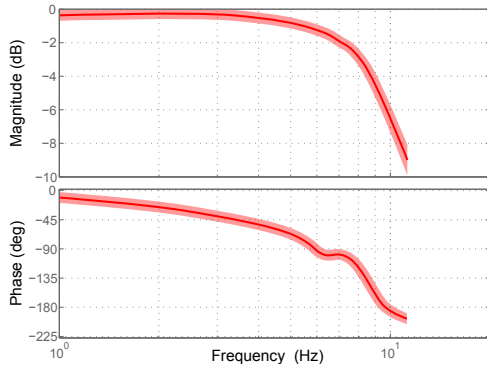


Fig. 11. Lip horizontal opening Bode diagram. The figure represents an average of the experimental runs conducted to estimate the lip horizontal opening motion bandwidth. The lip opening DOF was commanded to track a sinusoidal position reference at frequencies spanning from 1 to 10.5 [Hz]. As can be seen, the -3dB intersection occurs approximately at 8[Hz]. The banded area represents the average measurement accuracy of the lip opening position measurement.

VI. SYSTEM TESTING

To evaluate the performance of the system we carried out a careful analysis of the dynamic properties of the facial expressions mechatronics. The stiffness of the stretchable textile cover was difficult to estimate and to model beforehand; this made it difficult to predict the actual bandwidth of system. We therefore performed an experimental evaluation aimed at assessing the true capabilities of the system. The Bode diagrams of the two most critical DOF, namely the jaw opening motion and the lips horizontal opening motion are shown in Fig.10 and Fig.11.

Indeed both the actual bandwidth of the jaw motion and of the lips opening motion were slightly lower than predicted. Nevertheless, by comparing these data to those shown in Fig.4, we can state that the response of the system should be sufficiently fast to replicate the normal dynamics of human facial movements, including speech articulation.

VII. LESSONS LEARNT

The process of adjusting the functionality of the system to suitable levels of reliability and robustness, was iterative with many lessons learnt along the way; the most important are listed hereinafter.

- **Socket set screws vs. clamped connections.**

In the first iteration of our design all the motors were connected to the lever arms they actuated via pairs of socket set screws acting directly on the motor output shaft. We quickly realized that this type of connection was not suited for the fast switching motions characterizing the jaw and lips movements; as a consequence the pairs of socket set screws often loosened. This often required to unmount the face and head covers, and to tighten the screws to restore the functionality of the mechanism; this, in turn, made operating the system unpractical. All socket set screws connections were therefore replaced with clamped connections which proved to be more reliable.

- **Head cover heat conduction.**

Several modifications were implemented to ease the dissipation of the heat generated by the head, among which the addition of large openings on the head back, and the substitution of the on-board computer with a low power consumption alternative. Nevertheless, the CPU temperature levels reached 90% of the temperature threshold (80°C) during thermal stress testing. This effect was caused by the textile cover that reduced the amount of heat transferred by convection. Although not critical in this case, we recommend not to neglect this aspect if a similar solution is adopted in other systems.

- **Textile cover/mechanics connection.**

Joining the textile cover to the underlying mechanics was a challenging step. The different nature of the materials to be connected caused stress concentrations that tended to damage the cover during operation. After several alternatives were tried, the problem was solved by sewing on the textile cover plastic interface rings, which remained hidden thanks to a double hem line.

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, we presented the design and development of a new robot articulated face for the iCub humanoid robot, capable of performing different facial expressions with five additional DOF, and well suited to display correct articulations while speaking. We evaluated the performance of the new head by investigating the bandwidth of the system. The preliminary results suggest that the facial expressions can be actuated fast enough to replicate typical features of human speech movements. A video of the system in operation is attached to the present work to demonstrate its effectiveness. Similarly to all iCub project material, this project is open hardware and available for download with a LGPL license from the iCub svn repository³.

³<https://svn.robotology.eu/repos/iCubHardware-pub/trunk/mechanics/>



Fig. 12. Facial expressions. The figure four photographs of the expressions achievable with the proposed system.

The next phases for this project are to fine tune the speech synchronization system and conduct user studies, to understand if naive users can understand and interpret the gestures produced by the articulated talking face, correctly (like in [31]).

ACKNOWLEDGEMENTS

The authors would like to thank:

- Roberto Puddu, Marco Pinaffo, Charlie Sangiuneti, Massimiliano Mosto for the construction and assembly of mechanical components.
- Fabrizio Larosa, Davide Gandini, Simone Michelini and Emiliano Barbieri, for the logistic support and the help on wiring the system.
- Maurizia Botti for the painstaking help and support in developing the stretchable textile cover.

REFERENCES

- [1] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 143-166, 2003.
- [2] E. Wade, A. Parnandi, R. Mead, and M. Mataric, "Socially assistive robotics for guiding motor task practice," *Paladyn*, vol. 2, no. 4, pp. 218-227, 2011.
- [3] A. Tapus, M. Maja, B. Scassellatti, et al., "The grand challenges in socially assistive robotics," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, 2007.
- [4] A. Parmiggiani, M. Maggiali, L. Natale, F. Nori, A. Schmitz, N. Tsagarakis, J. S. Viktor, F. Becchi, G. Sandini, and G. Metta, "The design of the iCub humanoid robot," *International Journal of Humanoid Robotics*, vol. 9, 2012.
- [5] J. Riskin, "Eighteenth-century wetware," *Representations*, vol. 83, no. 1, pp. 97-125, 2003.
- [6] K. Fukui, K. Nishikawa, T. Kuwae, H. Takanobu, T. Mochida, M. Honda, and A. Takanishi, "Development of a new human-like talking robot for human vocal mimicry," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 1437-1442, IEEE, 2005.
- [7] R. Hofe and R. K. Moore, "Anton: an animatronic model of a human tongue and vocal tract," in *INTERSPEECH*, pp. 2647-2650, 2008.
- [8] C. L. Breazeal, *Designing sociable robots*. MIT press, 2004.
- [9] C. L. Breazeal, *Sociable machines: Expressive social exchange between humans and robots*. PhD thesis, Massachusetts Institute of Technology, 2000.
- [10] C. Breazeal, "Toward sociable robots," *Robotics and autonomous systems*, vol. 42, no. 3, pp. 167-175, 2003.
- [11] L. Aryananda and J. Weber, "Mertz: A quest for a robust and scalable active vision humanoid head robot," in *IEEE/RAS Int. Conf. on Humanoid Robots (HUMANOIDS)*, pp. 513-532, 2004.
- [12] MIT Media Lab, "Mds head & face." <http://robotic.media.mit.edu/projects/robots/mds/headface/headface.html>, July 2014.
- [13] N. Endo, S. Momoki, M. Zecca, M. Saito, Y. Mizoguchi, K. Itoh, and A. Takanishi, "Development of whole-body emotion expression humanoid robot," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 2140-2145, IEEE, 2008.
- [14] A. van Breemen, X. Yan, and B. Meerbeek, "iCat: an animated user-interface robot with personality," in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pp. 143-144, 2005.
- [15] I. Lutkebohle, F. Hegel, S. Schulz, M. Hackel, B. Wrede, S. Wachsmuth, and G. Sagerer, "The Bielefeld anthropomorphic robot head 'Flobi'," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 3384-3391, IEEE, 2010.
- [16] C.-Y. Lin, L.-C. Cheng, and L.-C. Shen, "Oral mechanism design on face robot for lip-synchronized speech," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4316-4321, IEEE, 2013.
- [17] T. Hashimoto, S. Hitramatsu, T. Tsuji, and H. Kobayashi, "Development of the face robot saya for rich facial expressions," in *SICE-ICASE, 2006. International Joint Conference*, pp. 5423-5428, Oct 2006.
- [18] K. Mianowski, N. Schmitz, and K. Berns, "Mechatronics of the humanoid robot roman," in *Robot Motion and Control 2007*, pp. 341-348, Springer, 2007.
- [19] H. Ishiguro and S. Nishio, "Building artificial humans to understand humans," *Journal of Artificial Organs*, vol. 10(3), pp. p. 133-142, 2007.
- [20] D. Hanson, A. Olney, S. Prilliman, E. Mathews, M. Zielke, D. Hammons, R. Fernandez, and H. Stephanou, "Upending the uncanny valley," in *Proceedings of the national conference on artificial intelligence*, vol. 20, p. 1728, 2005.
- [21] D. Hanson, "Exploring the aesthetic range for humanoid robots," in *Proceedings of the ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science*, pp. 39-42, 2006.
- [22] C. A. Moore and J. L. Ruark, "Does speech emerge from earlier appearing oral motor behaviors?," *Journal of Speech, Language, and Hearing Research*, vol. 39, no. 5, pp. 1034-1047, 1996.
- [23] J. R. Green, C. A. Moore, and K. J. Reilly, "The sequential development of jaw and lip control for speech," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 1, pp. 66-79, 2002.
- [24] B. L. Davis and P. F. MacNeilage, "The articulatory basis of babbling," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 6, pp. 1199-1211, 1995.
- [25] P. F. MacNeilage, "The frame/content theory of evolution of speech production," *Behavioral and brain sciences*, vol. 21, no. 04, pp. 499-511, 1998.
- [26] P. Badin, G. Bailly, L. Reveret, M. Baciuc, C. Segebarth, and C. Savariaux, "Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images," *Journal of Phonetics*, vol. 30, no. 3, pp. 533-553, 2002.
- [27] V. Ramanarayanan, L. Goldstein, and S. S. Narayanan, "Spatio-temporal articulatory movement primitives during speech production: Extraction, interpretation, and validation," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1378-1394, 2013.
- [28] VISATON, "BF 32 loudspeaker data-sheet." http://www.visaton.com/en/chassis_zubehoer/breitband/bf32_8.html, July 2014.
- [29] Faulhaber, "Faulhaber DC motors data-sheets." https://fmcc.faulhaber.com/type/PGR_13813_13801/PGR_13818_13813/en/, July 2014.
- [30] Soundman, "OKM II stereo microphone data-sheet." <http://www.soundman.de/en/products/okm-ii-studio-rock/>, July 2014.
- [31] R. Beira, M. Lopes, M. Praga, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Salfarén, "Design of the robot-cub iCub head," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 94-100, 2006.