

Image-based Assessment of Jaw Parameters and Jaw Kinematics for Articulatory Simulation: Preliminary Results

Ajish K. Abraham, V. Sivaramakrishnan, N. Swapna, N. Manohar

All India Institute of Speech and Hearing, Mysuru, India

ajish68@aiishmysore.in, sivaramakrishnanv7@gmail.com, nsn112002@yahoo.com,
manu.aiish@gmail.com

Abstract

Correcting the deficits in jaw movements have often been ignored in assessment and treatment of speech disorders. A robotic simulation is being developed to facilitate Speech Language Pathologists to demonstrate the movement of jaw, tongue and teeth during production of speech sounds, as a part of a larger study. Profiling of jaw movement is an important aspect of articulatory simulation. The present study attempts to develop a simple and efficient technique for deriving the jaw parameters and using them to simulate jaw movements through inverse kinematics.

Three Kannada speaking male participants in the age range of 26 to 33 years were instructed to produce selected speech sounds. The image of the final position of the jaw during production of each speech sound was recorded through CT scan and video camera. Angle of ramus and angle of body of mandible were simulated through inverse kinematics using RoboAnalyzer software. The variables for inverse kinematics were derived through kinematic analysis. The Denavit-Hartenberg (D-H) parameters required for kinematic analysis were obtained from still image. Angles simulated were compared with the angles obtained from CT scan images. No significant difference was observed.

Index Terms: Jaw kinematics, jaw movement simulation, articulatory simulation, image - based jaw measurement.

1. Introduction

Identifying the movements of jaw required for precise articulation of speech sounds and correcting the deficits have often been ignored in assessment and treatment of speech disorders [1]. Lack of dissociation and grading affects the production of major speech sounds [1]. The jaw, composed of upper jaw (maxilla) and lower jaw (mandible), must move fluidly during speech sound production, in synchrony with the appropriate movement of the lips and tongue. This movement is determined by the simultaneous activities of the temporomandibular joint (TMJ). In TMJ, opening and closing are the main movements with respect to the mandible; rotation and translation movements also occur during the production of speech sounds [2]. When there is an abnormality in the jaw structurally and functionally, one cannot produce sounds appropriately, which negatively affects the communication and hence, the quality of life [3].

The role of jaw movement in articulation has been investigated by several researchers. For phonemes like /b/, mandible is the primary mover [4]. Jaw position for stop

consonants were different depending upon the speech sound involved [5]. The effect of an increase in speaking rate was reflected by an overall decrease in the jaw displacement for vowels [6]. A study of jaw opening, position, displacement, velocity, duration for movement and jaw closing action during the production of a variety of vowels and consonants at normal and fast rate was conducted on normal American English speaking adults. The study concluded that faster speaking rate resulted in smaller displacements and lower velocities for jaw opening whereas jaw closing velocities were somewhat higher [7]. Jaw position is known to affect vowel height and also plays a role in the manner of consonantal articulation [8]. The above studies revealed that jaw related parameters vary with the speech sound, rate of speech, phonetic context and languages.

Various methods used for tracking jaw movements were reviewed by Madhavan et al. [9]. Tracking of jaw movements, focused towards assessment of speech disorders, include the electromagnetic articulography, ultrasound imaging, palatography and computerized tomography (CT) scan. Setting up the articulograph is a complex job, the process involved is time consuming and involves high equipment cost. On the other hand, ultrasound imaging and palatography does not provide information on the jaw parameters. Though CT is the best method, CT scan does carry risks, due to the radiation exposure and it is not advisable to go for repeated imaging. Speech-language pathologists (SLPs) do not attempt measurements of jaw parameters, despite their importance, due to the complexity and non affordability of the currently used equipments and techniques. Hence, there is a need to develop simple, efficient, affordable and non intrusive methods for jaw measurements with comparable accuracy of CT scan measurements.

A system with two cameras was developed for tracking jaw movements with a tracking error < 0.5 mm [10]. This method used reflective markers, required calibration of cameras and did not provide any jaw measurements. A system using Microsoft connect was developed to track the jaw movements without markers [11]. This system showed tracking errors in between 2.4 mm and 9 mm and was investigated with non speech tasks. A video-based method utilizing a depth sensor and face tracking algorithm was developed [12] without sensors or markers. This method was also complex.

SLPs correct the articulation disorders through speech therapy, by using several techniques such as manual demonstration of producing sounds, modeling of the sounds etc. 2D and 3D models of articulatory simulations of jaw

movements were implemented by several researchers. These include an early 2D model by Sanguineti et al. [13], a 3D rigid body jaw model [14], a rigid-body jaw-hyoid model [15], [16], a 3D jaw motion model [17] and a physiological articulatory model [18]. To win the child's attention on these 2D and 3D displays on the computer screen and wanting the child to participate in the speech therapy session is a challenge. A robotic simulation may overcome these challenges. Robotic jaw simulation implemented by Flores et al. [19] was for dental applications and the one implemented by Sawada et al. [20] was for voice simulation, which could not depict the jaw movements. A robotic model which can simulate the articulation movements to facilitate SLPs for better demonstration of the production of speech sounds is being developed by the authors of the present study, as part of a larger study. Profiling of jaw movement is an important aspect of developing a robotic simulator.

Analytical study of the motion of a robot manipulator is dealt by forward and inverse kinematics. In forward kinematics, homogenous transformations based on 4×4 real matrices are used to represent rotation of a robot manipulator [21]. A general transformation between two joints requires four parameters known as Denavit-Hartenberg (D-H) parameters [22]. The movement of the upper jaw and lower jaw can be evolved using D-H parameters. RoboAnalyzer software was introduced by Othayoth et al. [23] which can model the robot's input - output motion characteristics using D-H parameters. Using these parameters, the angle of ramus (θ_{1CT} , Figure 2) and angle of body of mandible (θ_{2CT} , Figure 2) are simulated in the present study. The accuracy of the simulation is obtained by comparing with the values obtained through CT scan. It is expected that, the present study will be able to develop an easily implementable method to measure the jaw parameters. Through the proposed model of jaw kinematics, the study would simulate the jaw movements and thereby aid in the development of a robotic jaw, which would further help in the implementation of robotic systems for articulation training.

The objectives of the study were: 1) To obtain the angle of ramus (θ_1) and angle of body of mandible (θ_2) from CT scan image and θ_1 , θ_2 , ramus length (a_1) and length of body of mandible (a_2) from still image, representing final position of production of selected speech sounds, 2) To derive the coordinates of symphysis through kinematic analysis using the jaw parameters obtained from still image, 3) To simulate the angles of jaw movement θ_1 and θ_2 through inverse kinematics using the coordinates of symphysis derived from still image and 4) To compare the simulated angles with θ_1 and θ_2 obtained through CT scan image.

Rest of the paper is organized as follows. The materials used and the procedure followed for data acquisition is described in Section 2. Experimental results are shown in Section 3, followed by the discussion in Section 4 and conclusion in Section 5.

2. Materials and Methods

2.1. Participants

Three Kannada speaking male participants in the age range of 26-33 years (mean age = 29 years) were included in the study. Only those adults with no history of sensory, neurological,

communicative, academic, cognitive, intellectual, emotional or oro-facial abnormalities were included. All ethical procedures were followed. The purpose and procedures of the study were explained to the participants and an informed verbal and/or written consent was obtained.

2.2. Speech Material

The participants were instructed to produce speech sounds /t/, /l/, /V/, /ʃ/, /k/ and /a/, each sound representing different place of articulation. The speech sounds selected for robotic simulation in the larger study were used in the present study. Each participant was asked to produce only two speech sounds (Table 1), each sound repeated three times, to restrict the exposure to radiation within safe limits.

Table 1: *Speech sounds produced by the participants.*

Participants	Speech Sounds
Participant 1	/a/ and /l/
Participant 2	/k/ and /ʃ/
Participant 3	/t/ and /V/

2.3. Data acquisition

The schematic in Figure 1 shows the procedure followed for data acquisition and simulation.

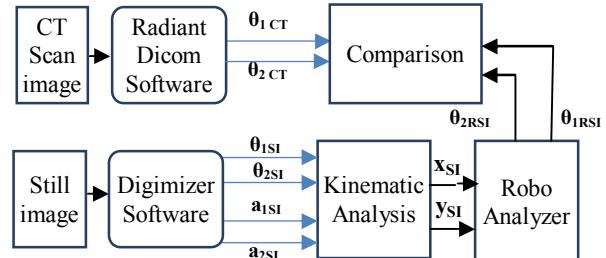


Figure 1: *Procedure for data acquisition.*

2.3.1. Obtaining the jaw parameters from CT scan images

Images of the jaw region at final position during production of each speech sound were obtained using Optima 660 128 multi slice CT scan with slice thickness of 0.625 sec. The participant was informed to hold the final position of production of the required speech sound for 20 seconds to record the image. Radiant DICOM Viewer software version 3.5 [24] was used to obtain the jaw parameters from the CT scan images (Figure 2). The parameters obtained were:- 1) Angle θ_{1CT} : Angle between the ramus line to the reference line (vertical line from the Condyle) and 2) Angle θ_{2CT} : Angle between the ramus line extension and the mandible line.

2.3.2. Obtaining the jaw parameters from still images

The participants were informed to sit straight in an acoustically treated room at a distance of 20 cm from the camera. They were instructed to utter each of the sounds (Table 1) three times, one after the other, facing the camera. The sound production was video recorded using a Samsung camera S5K5E8 (resolution 4000 x 3000), with frame rate of 30 fps. Still images representing final position during production of each sound were taken from the captured videos and incorporated into the Digimizer software [25] version 4.3. Ramus length (a_{1SI}): Length between the most posterior point

of the condyle and the intersection point of the lower mandible of the ramus; and Length of body of mandible (a_{2SI}): Linear distance between the symphysis and the intersection point of the ramus length, were measured from the image. The mandibular condyle (point B in figure 3) was identified on the image by the midpoint of the tragus (point A in figure 3) of the ear. The symphysis (point C in figure 3) of the mandible was identified by the soft tissue in the region of the chin. a_{1SI} was identified as the distance between points B and D, whereas, a_{2SI} was the distance between points C and D. The angles θ_{1SI} and θ_{2SI} were also identified from still image as shown in Figure 3.

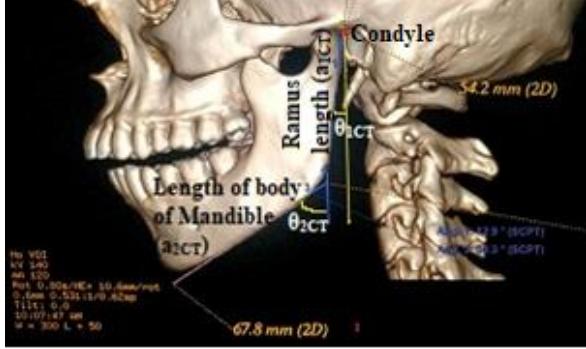


Figure 2: Jaw measurements from CT scan.

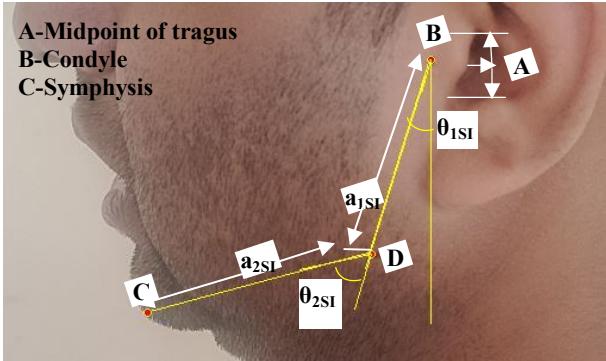


Figure 3: Jaw measurements from still image.

2.3.3 Kinematic analysis of jaw movement with 2 degrees of freedom (dof) to derive x and y coordinates of symphysis

Kinematic representation of a 2 dof robotic jaw (rest position) is shown in Figure 4.

Table 2: Denavit-Hartenberg (D-H) parameter for 2 degrees of freedom (θ -Joint angle in degree, d -Link offset in mm, a -link length in mm, α -Twist angle in Degree).

	θ	d	a	α
1	θ_1	0	a_1	0
2	θ_2	0	a_2	0

By using the D-H parameter (Table 2), the homogeneous transformation matrix is found out [22]. The D-H convention is a product of four basic transformations to represent the homogeneous transformation and is denoted by A_i . The A matrix is a homogenous 4x4 transformation matrix which describe the position of a point of simphysis and the

orientation of the object in a three dimensional space. In D-H convention, each homogeneous transformation matrix A_i is represented as a product of four basic transformations as follows.

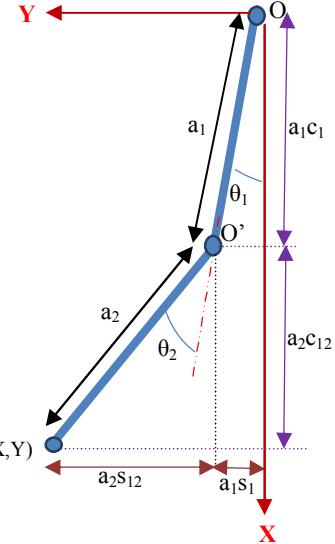


Figure 4: Revolute joint jaw manipulator with 2 degrees of freedom. (a_1 - Ramus length, a_2 -Mandible length).

Transformation matrices for joint 1 (O in Figure 4) and joint 2 (O' in Figure 4) are,

$$A_1 = \begin{bmatrix} c_1 & -s_1 & 0 & a_1 c_1 \\ s_1 & c_1 & 0 & a_1 s_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} c_2 & -s_2 & 0 & a_2 c_2 \\ s_2 & c_2 & 0 & a_2 s_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\text{Overall matrix } A_i = A_1 \cdot A_2 \quad (1)$$

Substituting A_1 and A_2 in (1)

$$A_i = \begin{bmatrix} c_{12} & -s_{12} & 0 & a_1 c_1 + a_2 c_{12} \\ s_{12} & c_{12} & 0 & a_1 s_1 + a_2 s_{12} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

From (2), coordinates x and y are found as

$$x = a_1 c_1 + a_2 c_{12} \quad (3)$$

$$y = a_1 s_1 + a_2 s_{12} \quad (4)$$

From Figure 4,

$$c_1 = \cos\theta_1, \quad (5)$$

$$c_{12} = \cos(\theta_1 + \theta_2), \quad (6)$$

$$s_1 = \sin\theta_1, \quad (7)$$

$$s_{12} = \sin(\theta_1 + \theta_2) \quad (8)$$

x, y coordinates are obtained for still image by substituting values of θ_1 and θ_2 in (5), (6), (7) and (8) and applying to (3) & (4) respectively along with substitution of values of a_1 and a_2 in (3) and (4).

2.3.4. Simulating θ_1 and θ_2 using RoboAnalyzer software

Using x and y values derived from still images through kinematic analysis, θ_1 and θ_2 were simulated using the RoboAnalyzer software [23].

2.4. Analysis

The values of θ_1 and θ_2 simulated through RoboAnalyzer and the actual values obtained from CT scan image were compared using Friedman's Test. Percentage error is calculated as

$$\% \text{ error} = \frac{\theta_{\text{CT}} - \theta_{\text{RSI}}}{\theta_{\text{CT}}} \times 100 \quad (9)$$

3. Results

Values of jaw parameters obtained from CT scan images and still images are shown in Table 3. These parameters were used to derive x_{SI} and y_{SI} coordinates (Table 4) through kinematic analysis of jaw movement with 2 degrees of freedom. The RoboAnalyzer software simulated θ_{IRSI} and θ_{RSI} through inverse kinematics using the x and y coordinates derived through kinematic analysis (Table 4).

Table 3: Jaw parameters obtained from CT scan and still images.

Sounds	CT scan		Still image			
	$\theta_{1\text{CT}}$ (degrees)	$\theta_{2\text{CT}}$ (degrees)	$a_{1\text{SI}}$ (mm)	$a_{2\text{SI}}$ (mm)	$\theta_{1\text{SI}}$ (degrees)	$\theta_{2\text{SI}}$ (degrees)
/a/	7.3	53.6	52.0	67.3	8.1	56.2
/t/	5.0	54.0	47.6	76.9	5.5	56.5
/l/	10.0	54.0	46.3	73.2	9.2	55.1
/v/	11.1	54.7	52.0	62.0	10.6	58.7
/ʃ/	12.9	60.3	50.0	63.9	14.0	62.1
/k/	13.0	52.0	50.0	63.0	14.0	56.3

Table 4: Coordinates (x,y) computed from the kinematic derivation and angles (θ_{IRSI} and θ_{RSI}) simulated through the RoboAnalyzer software.

Sounds	Coordinates from Kinematic derivation		Simulated angles (in degrees)	
	x_{SI}	y_{SI}	θ_{IRSI}	θ_{RSI}
/a/	80.6	67.9	7.98	56.41
/t/	83.3	72.4	5.37	56.75
/l/	77.4	73.3	9.14	55.15
/v/	72.9	67.6	10.57	58.86
/ʃ/	63.8	74.1	13.94	62.22
/k/	69.7	71.4	13.96	56.39

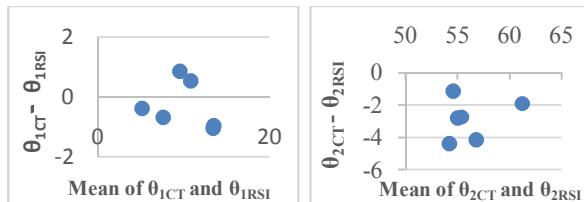


Figure 5: Bland-Altman plot of comparison between actual and simulated values of θ_1 and θ_2 .

No significant difference ($p > 0.05$) was observed when the simulated values (θ_{IRSI} and θ_{RSI}) of still image (Table 4)

were compared with the values ($\theta_{1\text{CT}}$ and $\theta_{2\text{CT}}$) recorded in the CT scan (Table 3), using Friedman's Test. The agreement between the two is shown in Figure 5. Percentage error (Table 5) was calculated for simulated values of θ_{IRSI} and θ_{RSI} comparing with the corresponding values ($\theta_{1\text{CT}}$ and $\theta_{2\text{CT}}$) from CT scan, using equation (9).

Table 5: Percentage error of the simulated angles.

Sounds	Percentage error	
	$\theta_{1\text{RSI}}$ (%)	$\theta_{2\text{RSI}}$ (%)
/a/	9.31	5.24
/t/	7.40	5.09
/l/	8.60	2.12
/v/	4.77	7.60
/ʃ/	8.06	3.18
/k/	7.38	8.44

4. Discussion

Information regarding jaw parameters is an essential input while developing the robotic jaw manipulator. The present study simulated angle of ramus (θ_{IRSI}) and angle of body of mandible (θ_{RSI}) through inverse kinematics using RoboAnalyzer software. Figure 5 shows the Bland-Atman plots which describe the agreement between the simulation values of θ_{IRSI} and θ_{RSI} with the corresponding values obtained from CT scan. The cross correlation between the two (simulated v/s actual) was found to be good. The maximum percentage error was found to be 9.31% (Table 5). This is comparable with the errors observed in a similar study [11].

A simple model of jaw kinematics has been proposed in this study. The actual physiological behavior of the TMJ is quite complex, with lateral and protrusion motions [26]. However, during speech production, the jaw movements occur mostly in the sagittal plane [10]. In the present study, the video recording for obtaining still image was focused on the sagittal plane. To achieve more repeatability and precision, the still image based jaw measurement system requires calibration for each user [10], which will be considered in the future. A comparison between various cameras in terms of distance and video resolution will also be taken up [12]. CT scan imaging was necessary to get the base values. The radiation exposure involved in CT scan imaging restricted the number of participants in the study, which is a major limitation of the study.

5. Conclusion

This study provided preliminary data on the jaw movements of Kannada speakers and demonstrated a technique for simulation of jaw movements. The data adds to the much needed corpus of characterization of vowels and consonants in Indian languages. Such data may be useful in building therapeutic models for adults with speech sound disorders. The current study was limited to simulating jaw movements pertaining to selected speech sounds. Future studies can include stimuli of different syllable structure.

6. Acknowledgements

The work was supported by Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India.

7. References

- [1] S. Rosenfeld-Johnson and J. Gray, "The role of the jaw for feeding and speech," *Advance Magazine*, vol. 18, no. 35, 2008.
- [2] A. Smith, "The Control of Orofacial Movements in Speech," *Critical Reviews in Oral Biology and Medicine*, vol. 3, no. 3, pp. 233-267, 1992.
- [3] M. G. Esther Bianchini, Guiovaldo Paiva and R. F. Cláudia de Andrade, "Mandibular Movement Patterns During Speech in Subjects with Temporo mandibular Disorders and in Asymptomatic Individuals," *The Journal of Craniomandibular & Sleep Practice*, vol. 26, no. 1, pp. 50-58, 2008.
- [4] 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*, 2002.
- [5] T. Gay, "Jaw movements during speech: A cineradiographic investigation," *Haskins Laboratories, Status Report on Speech Research*, vol. 39, no. 40, pp. 219-229, 1974.
- [6] T. Gay, "A cineradiographic study of vowel production," *Journal of Phonetics*, vol. 2, no. 4, pp. 255-266, 1974.
- [7] K. Oshimat, and V. L. Gracco, "Mandibular contributions to speech production," In *Second International Conference on Spoken Language Processing*, 1992.
- [8] C. Mooshammer, P. Hoole, and A. Geumann, "Jaw and order," *Language and Speech*, vol. 50, no. 2, pp. 145-176, 2007.
- [9] Sanjay Madhavan, M. Dhanraj and R. Ashish Jain, "Methods of recording mandibular movements - A review," *Drug Invention Today*, vol. 10, no. 7, pp. 1254-1259, 2018.
- [10] Y. Feng and L. Max, "Accuracy and precision of a custom camera-based system for 2-D and 3-D motion tracking during speech and nonspeech motor tasks," *Journal of Speech, Language, and Hearing Research*, vol. 57, pp. 426-438, 2014.
- [11] Y. Tanaka, T. Yamada, Y. Maeda and K. Ikebe, "Markerless three-dimensional tracking of masticatory movements," *Journal of Biomechanics*, vol. 49, pp. 442-449, 2016.
- [12] B. Andrea, N. Aravind and Y. Yana, "Video-based tracking of jaw movements during speech: Preliminary results and future directions," *Proc. Interspeech 2017*, pp. 689-693, 2017.
- [13] V. Sanguineti, R. Laboissiere and D. J. Ostry, "A dynamic biomechanical model for neural control of speech production." *Journal of the Acoustical Society of America*, vol. 103, no.3, pp. 1615-1627, 1998.
- [14] Q. Fang, S. Fujita, X. Lu and J. Dang, "A model-based investigation of activations of the tongue muscles in vowel production." *Acoust. Sci. and Tech*, vol. 30, no. 4, pp. 277-287, 2009.
- [15] I. Stavness, J. Lloyd, Y. Payan and S. Fels, "Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics." *International Journal of Numerical Methods in Biomedical Engineering*, vol. 27, pp. 367-390, 2010.
- [16] I. Stavness, J. Lloyd, Y. Payan and S. Fels. "Towards Speech Articulation Simulation with a Dynamic Coupled Face-Jaw-Tongue Model." *ISSP'2011, Canada*, pp.1-4, 2011.
- [17] E. Vatikiotis-Bateson and D. J. Ostry, "Analysis and modeling of 3D jaw motion in speech and mastication," *IEEE Xplore Conference: Systems, Man, and Cybernetics*, vol. 2, 1999.
- [18] J. Dang and K. Honda, "A physiological articulatory model for simulating speech production process," *Acoust. Sci. and Tech.* vol. 22, no.6, pp. 415-425, 2001.
- [19] E. Flores, S. Fels and E. Vatikiotis-Bateson, "Chew on this: Design of a 6DOF anthropomorphic robotic jaw," *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, Jeju, Korea (South), pp. 648-653, 2007.
- [20] H. Sawada, M. Kitani and Y. Hayashi, "A robotic voice simulator and the interactive training for hearing-impaired people." *BioMed Research International*, 2008.
- [21] S. Kucuk and Z. Bingul, "Robot kinematics: Forward and inverse kinematics." *INTECH Open Access Publisher*, 2006.
- [22] J. Denavit and R. S. Hartenberg, "A kinematic notation for lower-pair mechanisms based on matrices." *Journal of Applied Mechanics*, vol. 1, pp. 215-221, 1955.
- [23] R. S. Othayoth, R. G. Chittawadigi, R. P. Joshi and S. K. Saha, "Robot kinematics made easy using RoboAnalyzer software," *Computer Applications in Engineering Education*, vol. 25, no. 5, pp. 669-680, 2017.
- [24] Radiant DICOM Viewer. (n.d.). "Radiant DICOM Viewer.". <https://www.radiantviewer.com/>
- [25] Digimizer. (n.d.). "Easy-to-use image analysis software.". <https://www.digimizer.com/>
- [26] W. Yang, N. Marshak, Sýkora, D. et al. Building anatomically realistic jaw kinematics model from data. *Vis Comput*, vol. 35, pp. 1105-1118, 2019.