Estimating the Distribution of Oral Presentation Skills in an Educational Institution: A Novel Methodology

Federico Domínguez[1,2][a], Leonardo Eras[1][b], Josué Tomalá[2][c] and Adriana Collaguazo[2][d]

[1] *Information Technology Center, Escuela Superior Politécnica Del Litoral, ESPOL, Guayaquil, Ecuador*

[2] *Faculty of Electrical and Computer Engineering, Escuela Superior Politécnica Del Litoral, ESPOL, Guayaquil, Ecuador*

Keywords: Oral Presentation Skills, Human Pose Identification, Feedforward Neural Network, Automatic Presentation Feedback.

Abstract: Mastering oral presentation skills is of paramount importance for new graduates as they navigate the competitive job market of the 21st century. Consequently, procuring the effective development of these skills in students is an essential task for higher education institutions (HEIs). We developed a technological solution that facilitates oral presentation skills learning by providing automatic and immediate feedback using machine learning algorithms on audiovisual recordings of oral presentations. We have been using this tool to record novice students' presentations since 2017 and, by using the resulting data corpus, developed a methodology to accurately detect and evaluate posture and gaze in oral presentations. This article presents this methodology and its application on more than 3,000 recordings from more than 2,000 different students across all study programs at our university. Preliminary results provide a glimpse of the prevalence and distribution of oral presentation skills across several demographic variables. Statistically significant patterns point to possible oral communication deficiencies in engineering programs at our HEI, highlighting the potential of our methodology to serve as a diagnostic tool for communication skills learning strategies.

1 INTRODUCTION

Effective oral communication is one of the core competencies for higher educated professionals and one of the main skills needed to succeed in the 21st century society (Trilling and Fadel, 2009; van Ginkel et al., 2015). For this reason, teaching oral presentation skills in higher education institutions, whether in specialized communication courses or embedded in disciplinary curricula, is becoming increasingly important. In this context, plenty of practice and timely feedback have been identified as key components to the effective development of oral presentation competence (De Grez et al., 2009). The limited time available in typically crowded classrooms has prompted the development of technological solutions that facilitate practice and automate feedback of oral presentations and current research findings point to their effectiveness in improving oral communication skills in higher education students (Ochoa and Dominguez, 2020; van Ginkel et al., 2019).

The Automatic Feedback Presentation system (RAP for its Spanish acronym) is one of those systems. Developed at ESPOL University in Guayaquil, Ecuador, the RAP system provides an immersive environment for oral presentation practice, recording, and automatic feedback delivery. The system uses a recording of the presentation and, optionally, the presentation slides file to extract basic oral communication features such as the presenter's posture, gaze, voice volume, use of filled pauses, and slides legibility to generate a feedback report (Ochoa et al., 2018). Since 2017, the system has been experimentally integrated in the educational activities of several courses, logging more than 3,000 recordings from over 2,000 students across all programs at ESPOL University (Domínguez et al., 2021).

The deployment of the RAP system in a large real learning scenario provided us with a unique opportunity to explore the prevalence of oral presentation skills across different disciplines in a relatively large sample of students and academic staff. Moreover, current state-of-the-art computer vision algorithms for

[a] https://orcid.org/0000-0002-3655-2179
[b] https://orcid.org/0000-0002-3594-9289
[c] https://orcid.org/0000-0001-9406-3296
[d] https://orcid.org/0000-0002-0707-0226

Domínguez, F., Eras, L., Tomalá, J. and Collaguazo, A.
Estimating the Distribution of Oral Presentation Skills in an Educational Institution: A Novel Methodology.
DOI: 10.5220/0011853900003470
In *Proceedings of the 15th International Conference on Computer Supported Education (CSEDU 2023) - Volume 2*, pages 39-46
ISBN: 978-989-758-641-5; ISSN: 2184-5026

human pose estimation allow us to detect and classify posture and gaze in oral presentation recordings with increasing detail and at relatively low computational costs. Some of these techniques were not used in the initial deployment of the RAP system, however it is possible to retrofit them to the system's previous output (presentation recordings) to gain insights into oral presentation skills in higher education students and professionals.

This article presents a methodology to extract and evaluate posture and gaze from oral presentation recordings and explores the use of these two features as a proxy measurement for oral presentation competence in a sample of 2191 users. Section 2 briefly summarizes the state-of-the-art techniques used to detect and classify human pose in monocular images and section 3 presents the methodology to detect, classify, and evaluate posture and gaze in a RAP recording. The results of applying this methodology in 3726 recordings is presented in section 4. Discussions and conclusions stemming from these results are presented in sections 5 and 6.

2 STATE-OF-THE-ART IN HUMAN POSE IDENTIFICATION

Human pose identification can be defined as the combined process of estimating the configuration of the body (pose) from a single, typically monocular, image and then classifying the pose within the context of the image. In the context of an oral presentation, we define posture as the configuration of the torso, arms, and legs; and gaze as the orientation of the eyes and head. Therefore, the human pose identification process is used to extract both features, posture and gaze, from an oral presentation. Aside from automatic evaluation of oral presentations, applications of human pose identification range from fall detection in health care and industry (Tran et al., 2021; Hasib et al., 2021; Ren et al., 2020; Liu et al., 2022) to workout guidance in sports (Hung et al., 2020). In most implementations, human pose identification is subdivided in two distinct processes: estimation and classification.

2.1 Human Pose Estimation

The human pose estimation process outputs the coordinates of the most important joints of the human body (e.g., elbow, wrist, knee) detected in an image. This process is key in applications such as human activity recognition, human-computer interaction, animation, marker-less motion capture, and more (Sun et al., 2019; Sigal, 2014). Performance of estimation models on these application domains depends on several factors such as occlusions and truncations of the human body in the image, lighting and contrast, and noise (Andriluka et al., 2014).

The state-of-the-art in human pose estimation changes quickly as it is a relatively new technology powered typically by Convolutional Neural Networks (CNN) and the Deep Learning revolution; however two solutions stand out for their performance, real-time capabilities, and ease of deployment: OpenPose from Carnegie Mellon University and Mediapipe BlazePose from Google (Mroz et al., 2021; Bazarevsky et al., 2020; Cao et al., 2019). OpenPose relies heavily on GPU power to produce accurate results while BlazePose trades-off accuracy for faster runtime performance (Mroz et al., 2021). BlazePose accuracy is relatively high and close to state-of-the-art while its real-time performance makes it suitable for mobile applications (Mroz et al., 2021; Bazarevsky et al., 2020). Moreover, BlazePose has been found to outperform OpenPose in images with human self-occlusion (Liu et al., 2022).

2.2 Human Pose Classification

The classification or annotation of the action represented by a specific human pose is a difficult problem in itself. Human limbs have several degrees of freedom and in a typical image are occluded with objects or parts of the body. Classification through a mathematical model of the position of the human limbs has been done in the original version of the RAP system but scales poorly and it is usually discarded as a difficult problem (Ren et al., 2020).

Several machine learning (ML) algorithms have been proposed to classify postures: fuzzy logic with Support Vector Machines (SVM) (Ren et al., 2020), CNNs (Hasib et al., 2021), and k-means with YOLO (Tran et al., 2021). K-means, an unsupervised ML algorithm, allows for automated annotation but with mixed results. Most techniques, however, rely on supervised ML algorithms that require manual annotation for training, a typically onerous and expensive task.

3 METHODOLOGY

Versions 1 and 2 of the RAP system, deployed in 2017 and 2019 respectively, used OpenPose for human pose estimation and a mathematical model to classify postures as either open (one or both arms open) or close (hands down closer to the body). This methodology had an accuracy rate of 84% for correctly identifying an open/close posture and a slightly higher accuracy rate for correctly identifying gaze to the audience in an oral presentation. Tests on the field revealed that, while the system was accurate enough to generate an effective feedback report, the amount of false positives alienated some users. For this reason, for version 3 of the RAP system, deployed in 2022, we improved the extraction of the posture and gaze features by changing both the human pose estimation and human pose classification algorithms. As the system is able to identify presentation postures with finer granularity, it was also necessary to revisit the automatic posture evaluation algorithm. The next subsections detail the methodology used to improve these algorithms.

3.1 Pose Estimation in an Oral Presentation

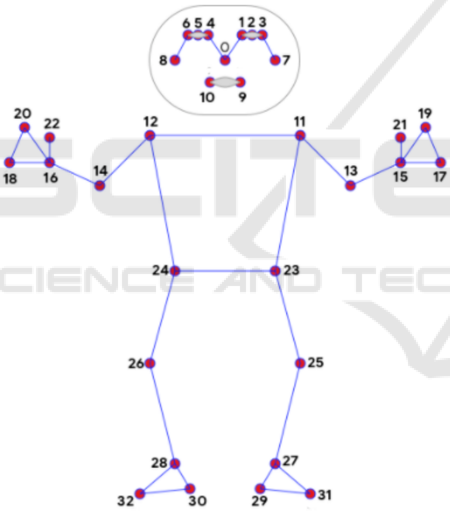


Figure 1: BlazePose is capable of detecting a human pose in an image and outputs the x,y,z coordinates (landmarks) of the 33 most important joints of the human body (Bazarevsky et al., 2020).

For the 3rd version of the RAP system we decided to switch the human pose estimation library from OpenPose to BlazePose because BlazePose:

- uses, as part of Google's Mediapipe, the Apache 2.0 open source license which provides greater flexibility with fewer limitations;
- does not require a GPU and can therefore run on a mobile phone, aligning with the future road map of the RAP system;
- detects more body landmarks in the torso, arms, and legs enabling finer posture classification; and

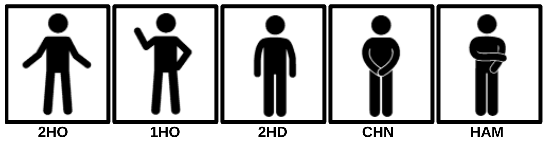


Figure 2: Five classification targets for postures were identified and are defined as such: 2HO - two hands open, 1HO - 1 hand open, 2HD - two hands down, CHN - closed hands, HAM - hand in arm.

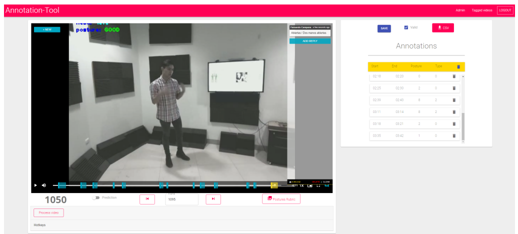


Figure 3: Our annotation tool allows taggers to seek for a specific time-lapse in the video and label the posture manually.

- has a mature Python API that facilitates its integration into currently existing systems.

In the RAP system, BlazePose takes as input an uncompressed video frame of 1500 x 1000 pixels and outputs a data frame of 33 body landmarks (see Figure 1). These landmarks are then normalized and fed to the posture and gaze classification module.

3.2 Posture and Gaze Classification in an Oral Presentation

Using human pose landmarks to classify the presenter's posture and gaze requires, regardless of the classification algorithm, the previous identification of expected classification targets. For gaze it is relatively simple, the presenter's gaze is either aimed at the front (audience), left or right (slides), or the back. For posture however, subtle differences in body configuration can imply totally different postures spawning a larger number of classification targets. Posture and gaze are therefore classified separately using independently trained algorithms.

3.2.1 Posture

The previous posture classification algorithm used a mathematical model to discriminate between two classification targets: open and closed posture. This algorithm was sensitive to different video and room configurations, affecting its accuracy which reached a maximum of 84%. By 2019, the system had accumulated 3726 recordings from 2191 users, therefore, with the aim to improve accuracy and include new

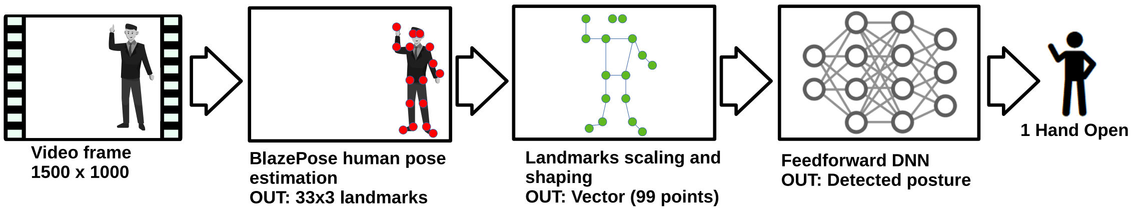


Figure 4: Human pose identification pipeline example of a single video frame where the input is the uncompressed frame and the output is the detected posture.

classification targets, we explored more flexible machine learning algorithms that benefited from a large data set.

We identified 11 distinct postures employed by users of the RAP system using a random sample of video recordings in an initial exploratory analysis. Of these 11 postures, only five were frequent enough to be able to use them reliably in a training data set. These five postures are: two hands open, one hand open, two hands down, closed hands, and hand in arm (see Figure 2).

After the identification of classification targets for postures, the next step was to manually tag video frames in RAP recordings to procure a ground truth data set that can be used for training and evaluation of ML algorithms. We organized an internal tagging campaign by recruiting volunteers, students and instructors, and providing them with a web annotation tool to facilitate and structure the video tagging task. This tool, written in Javascript and based on the videojs-annotation-comments plugin (Contently, 2022), serves each volunteer a random RAP video recording and an intuitive user interface to tag postures in a specific section of the video (see Figure 3).

Seven taggers (two professors, one research staff, and four students) annotated 301 RAP recordings during two tagging campaigns. To ensure a substantial inter-rater consistency (Fleiss' Kappa > 0.6), all taggers were trained on the correct identification of all five postures and five pre-annotated videos were used to quantify agreement between taggers. The resulting annotated data set consisted of more than 180 thousand video frames, each frame consisting of one identified posture.

Several ML algorithms were tested using the annotated data set: support vector machines (SVMs), k-nearest neighbors (kNN), logistic regression, and a feedforward deep neural network (Feedforward DNN). Best results were obtained with a feedforward DNN of 6 hidden layers (see Figure 7 for the detailed architecture) with a maximum accuracy of 95.5%. The resulting network architecture was obtained empirically and Figure 5 shows a confusion matrix with the accuracy results per category target using 145,450 frames for training, 9,331 frames for validation, and 26,339 frames for testing.

Figure 4 details the resulting human pose identification pipeline where a video frame from a RAP recording is first processed by BlazePose to extract the presenter's body landmarks; these landmarks are then normalized and fed to the feedforward DNN which predicts the presenter's posture.

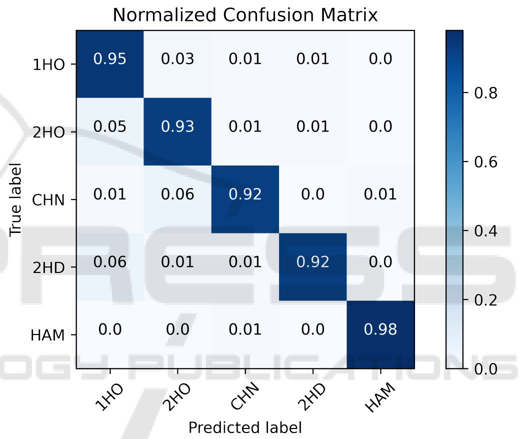
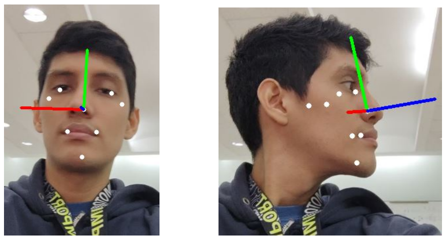


Figure 5: Confusion matrix for five posture classification targets using 145,450 frames for training, 9,331 frames for validation, and 26,339 frames for testing.

3.2.2 Gaze

For gaze we used a kNN classification algorithm that can use either a set of eight face points or a set of 3 angles as input to solve a PnP (Perspective-n-Point) problem in order to get the normal to the face spanning from the nose. Seven of these eight points were already extracted using BlazePose and the last one, the chin, had to be calculated from the rest. Pitch, yaw, and roll correspond to the 3 angles that can be obtained after calculating the normal vector (see Figure 6).

The accuracy of the gaze prediction varies slightly when using one of the two methods described. Using the eight face points produced a 96.42% accuracy, while using the normal vector angles produced a 94.11% accuracy.



(a) Looking at the camera (b) Looking away

Figure 6: The normal vector of the face (blue line) is used to estimate the direction of the gaze. Here an example of an individual with the gaze to the center and to the right side of the video camera.

We identified six distinct gazes: center, back, left, right, up, and down. Only two gazes were frequent enough to be able to use them reliably, these are: center and back. Of those, center is the only gaze related to a positive score as it represents a person looking directly at the audience. Back is related to the person looking directly at the presentation slides. Up and down gazes almost never occur in a real presentation, only during testing. Left and right gazes, while detected in a few presentations, did not appear enough to be useful in this study.

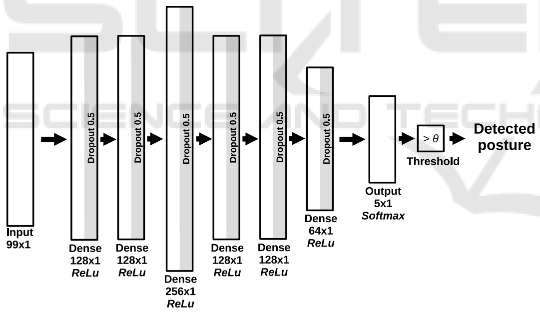


Figure 7: Architecture of the Feedforward Artificial Neural Network with 6 hidden layers used as the posture classifier in the final stage of the human pose identification pipeline.

3.3 Posture and Gaze Evaluation in an Oral Presentation

Posture and gaze evaluation is the process of discerning how appropriate a given posture or gaze is in the context of an oral presentation. For example, during an oral presentation direct eye contact with the audience is preferred than constantly looking at the slides or any form of gaze aversion (Gordon et al., 2006). In a RAP recording, a gaze to the front/camera is highly correlated with eye contact with the audience and a gaze to the side or back is highly correlated with the presenter looking at the slides. Therefore, during the evaluation process a gaze to the front is rewarded while a gaze to the side or back is penalized.

The presenter's posture is considered a form of body language which typically conveys an unintentional and unconscious message to the audience (Dittmann, 1987). It has been found that certain postures can unambiguously convey a positive or negative message, for example, an open posture communicates receptivity and openness to the audience (van Ginkel et al., 2019; Bull, 1987) while a closed posture communicates discomfort, nervousness, or disinterest (Sheth, 2017). Moreover, an open posture has been found to have a positive effect on the persuasiveness of the presentation (Bull, 1987). Therefore, open postures such as two hands open and one hand open are rewarded and closed postures such as closed hands and hand in arm are penalized.

After a RAP presentation, the system's evaluation process produces a presentation score for each feature (i.e. posture, gaze, filled-pauses, etc.). In the case of posture, points are assigned to each video frame depending on the identified posture. The overall score is the sum of all points per frame divided by the total number of frames. Table 1 shows the amount of points given to each posture resulting in presentations dominated with open postures receiving higher scores. Presentations where hands down or hand in arm dominate – usually due to fidgeting, communicating nervousness and stiffness (Sheth, 2017; Gordon et al., 2006) – will consequently have lower scores. The same process is done for gaze, table 2 shows the amount of points given to each detected gaze.

Table 1: Points assigned to each identified posture during the evaluation process.

Posture	Points
Two hands open (2HO)	+2
One hand open (1HO)	+1
Closed hands (CHN)	-1
Two hands down (2HD)	-2
Hand in arm (HAM)	-3

Table 2: Points assigned to each identified gaze during the evaluation process.

Gaze	Points
To the audience (center)	+2
To the left	-1
To the right	-1
To the back	-3

4 RESULTS

We applied the methodology described in the previous section in the implementation of version 3 of the

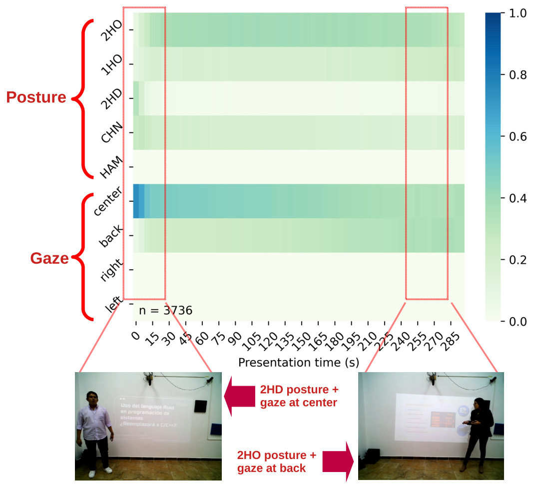


Figure 8: This heatmap shows the average frequency usage of posture and gaze for all users, represented with color bands in 5-seconds resolution, in every stage of the RAP presentation's 5-minute span. It shows that on average users start their presentation looking at the front with their hands down and move on to stare at the slides as the presentation progresses.

RAP system and on 3726 RAP recordings made using versions 1 and 2 of the system from 2017 till 2019. These recordings were made mostly as course assignments by students but also as part of training exercises by administrative staff and professors. Most of these recordings came from Communication I and Physics I courses, which are mandatory for almost all students, providing a large sample containing students from all study programs.

We used the evaluation process for posture and gaze described in section 3.3 to quantify a subset of the oral presentations skills of a RAP user. While we are aware that presentation skills have several additional dimensions not covered by posture and gaze, results obtained in controlled experiments of the RAP system point to a positive correlation between these two features and a novice student overall presentation skills, please refer to (Ochoa and Dominguez, 2020; Domínguez et al., 2021) for details. Quantification is presented here using the following tools: a presentation heatmap, which shows the use of postures and gaze over the presentation time, and the mean score per feature.

A presentation heatmap is drawn by calculating the frequency of each posture and gaze in five-seconds intervals for the entire five minute presentation; the color of each band represents how frequently was a specific posture or gaze detected during that interval of the presentation (Figure 8). Averaging heatmaps across demographic categories allow us to quickly visualize differences in presentation style and behaviour. Findings stemming from visual differences between heatmaps were corroborated using a Welch t-test for the means of each heatmap band in each category, a Bonferroni correction was applied to all p-values.

Figure 9 shows the average heatmaps of students' recordings per study program. Two programs stand out, Social studies and Design and Communication, as their students clearly spend more time looking at the audience than the rest (*center* band in both programs is darker than the rest, $P < .001$). Less salient than gaze to the audience, students in both programs also use less the one hand open (1HO) posture ($P < .001$ for all 1HO pairings except with Maritime and Life sciences). As expected the average and median gaze score for students in both study programs is higher than their peers (Figure 10), the same effect was not observed in the average posture score.

Figure 11 compares presentation performance between students in the two extremes of academic performance. Students in the 90th percentile tend to look more to the audience than students in the 10th percentile ($P = .0015$ for center). Statistical significance was not observed within intermediate percentiles. Figure 12 compares presentation performance between all students and professional staff (high school and university professors, administrative staff, and research staff). Professional staff tend to perform better as they maintain their gaze to the audience during the entire presentation ($P < .001$) and tend to use more both hands ($P < .001$ for 2HO). Students tend to use more the one hand open posture ($P < .001$ for 1HO).

Figure 13 compares presentation performance for students who used the system twice in the same semester. In the second attempt, students tend to look more at the audience, use more both hands, and use less the one hand open posture ($P < .001$ for all). As in previous comparisons, as users tend to look more to the audience their use of the one hand open posture (1HO) decreases because typically this posture is used to point at the slides while the user is looking away from the audience.

5 DISCUSSIONS

By combining academic and demographic metadata with our human posture and gaze evaluation methodology in oral presentation recordings a few important patterns emerged from the results in the previous sections. First, at least in our academic institution, oral presentation skills are not equally distributed among study programs, with engineering students lagging

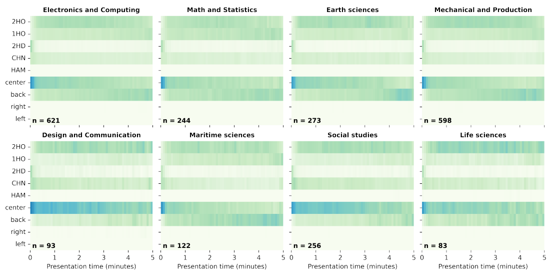


Figure 9: Average presentation heatmap by study program: Students on social sciences and design programs tend to look more at their audience.

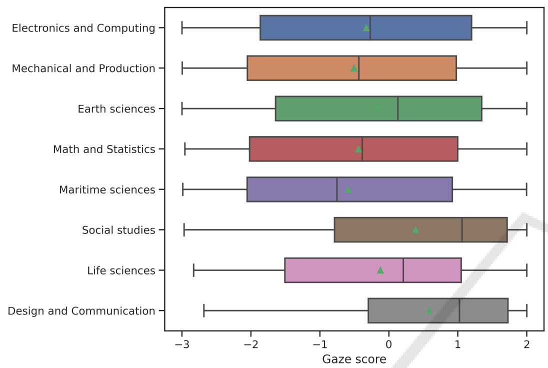


Figure 10: Distribution of student's gaze score by study program.

behind. Not surprisingly, top students present better than their peers, professional staff perform better than novice students, and students tend to improve their presentation skills after using the RAP system. No differences in presentation skills were observed by gender or social economic status.

Considering that oral presentation skills are an essential tool for all students and that international engineering accreditation programs such as ABET emphasize and evaluate oral communication competency, it is important to evaluate how differences in study program curricula affect intake of these skills. How much time is allocated to oral presentation skills exercises and feedback in different study programs? Nevertheless, the bulk of students that used our system are novice students in their second or third semester, therefore confounding factors such as extroverts disproportionately selecting study programs in social and communication sciences are still important and must be taken into account.

6 CONCLUSIONS

This article presents a methodology to evaluate human posture and gaze in oral presentation recordings

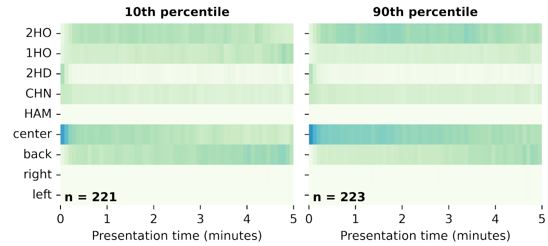


Figure 11: Average presentation heatmap by academic performance: Top students look more at the audience and use more open gestures than students in the tenth percentile.

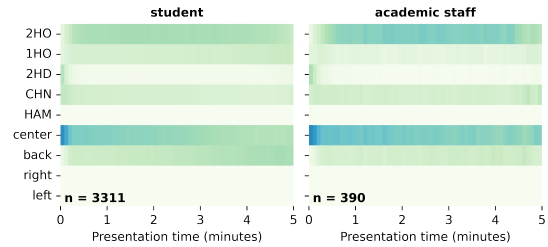


Figure 12: Average presentation heatmap for students and university staff: Professional staff tend to perform better presentations than students.

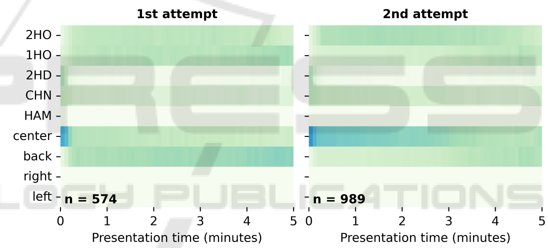


Figure 13: Average presentation heatmap for students who used the system twice: On the second attempt, students focused less on the slides.

and use this evaluation as a proxy measurement for oral presentation skills in students in a higher education institution. One of our main findings, based on 3726 recordings from 2191 subjects, is that oral presentation skills in engineering students lag behind students in the social sciences and design and communication study programs. This result can be used as a motivation to evaluate how oral communication skills are learned across study programs in our institution. Therefore, our methodology together with our technological tool, the RAP system, can be used not only as a tool to learn basic presentation skills but also to diagnose the effectiveness of communication skills learning strategies across study programs in an educational institution.

ACKNOWLEDGEMENTS

We would like to thank Fernando Campaña, Karen Bermúdez, Kelly Castro, Karina Ortega, Ricardo Salazar, and Juan Francisco Quimi for their invaluable help during the execution of this project.

REFERENCES

Andriluka, M., Pishchulin, L., Gehler, P., and Schiele, B. (2014). 2D human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3686–3693. IEEE.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., and Grundmann, M. (2020). BlazePose: On-device Real-time Body Pose tracking. In *Fourth Workshop on Computer Vision for AR/VR*, Seattle, WA, USA.

Bull, P. (1987). *Posture & Gesture*. Elsevier.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172 – 186.

Contently (2022). Videojs annotation comments. https://contently.github.io/videojs-annotation-comments/. Accessed: 2022-10-31.

De Grez, L., Valcke, M., and Roozen, I. (2009). The impact of goal orientation, self-reflection and personal characteristics on the acquisition of oral presentation skills. *European Journal of Psychology of Education*, XXIV:293–306.

Dittmann, A. (1987). The Role of Body Movement in Communication. In Siegman, A. and Feldstein, S., editors, *Nonverbal Behavior and Communication*, pages 37 – 64. Psychology Press, second edition edition.

Domínguez, F., Ochoa, X., Zambrano, D., Camacho, K., and Castells, J. (2021). Scaling and Adopting a Multimodal Learning Analytics Application in an Institution-Wide Setting. *IEEE Transactions on Learning Technologies*, 14(3):400–414.

Gordon, R. A., Druckman, D., Rozelle, R. M., and Baxter, J. C. (2006). Non-verbal behaviour as communication. In Hargie, O., editor, *The Handbook of Communication Skills*. Routledge.

Hasib, R., Khan, K. N., Yu, M., and Khan, M. S. (2021). Vision-based Human Posture Classification and Fall Detection using Convolutional Neural Network. In *2021 International Conference on Artificial Intelligence, ICAI 2021*, pages 74–79, Islamabad, Pakistan.

Hung, J. S., Liu, P. L., and Chang, C. C. (2020). A Deep Learning-based Approach for Human Posture Classification. In *MSIE 2020: Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering*, pages 171–175, Osaka, Japan. ACM.

Liu, W., Liu, X., Hu, Y., Shi, J., Chen, X., Zhao, J., Wang, S., and Hu, Q. (2022). Fall Detection for Shipboard Seafarers Based on Optimized BlazePose and LSTM. *Sensors*, 22(14).

Mroz, S., Baddour, N., McGuirk, C., Juneau, P., Tu, A., Cheung, K., and Lemaire, E. (2021). Comparing the Quality of Human Pose Estimation with BlazePose or OpenPose. In *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, pages 1–4. IEEE.

Ochoa, X. and Dominguez, F. (2020). Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting. *British Journal of Educational Technology*, 51(5):1615–1630.

Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., and Castells, J. (2018). The RAP System : Automatic Feedback of Oral Presentation Skills Using Multimodal Analysis and Low-Cost Sensors. In *LAK'18: International Conference on Learning Analytics and Knowledge*, pages 360–364, Sydney, Australia. ACM.

Ren, W., Ma, O., Ji, H., and Liu, X. (2020). Human Posture Recognition Using a Hybrid of Fuzzy Logic and Machine Learning Approaches. *IEEE Access*, 8:135628–135639.

Sheth, T. (2017). Non-verbal Communication: A Significant Aspect of Proficient Occupation. *Journal of Humanities and Social Science*, 22(11):69–72.

Sigal, L. (2014). Human Pose Estimation. In Ikeuchi, K., editor, *Computer Vision*, chapter Human pose, pages 362–370. Springer, Boston, MA, USA.

Sun, K., Xiao, B., Liu, D., and Wang, J. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5693–5703, Long Beach, CA, USA. IEEE.

Tran, T. H., Nguyen, D. T., and Phuong Nguyen, T. (2021). Human Posture Classification from Multiple Viewpoints and Application for Fall Detection. In *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, pages 262–267, Phu Quoc Island, Vietnam. IEEE.

Trilling, B. and Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times*. John Wiley & Sons.

van Ginkel, S., Gulikers, J., Biemans, H., and Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. *Educational Research Review*, 14:62–80.

van Ginkel, S., Gulikers, J., Biemans, H., Noroozi, O., Roozen, M., Bos, T., van Tilborg, R., van Halteren, M., and Mulder, M. (2019). Fostering oral presentation competence through a virtual reality-based task for delivering feedback. *Computers and Education*, 134(July 2017):78–97.