EchoMotion: Enhancing Exercise Analysis with Acoustic Sensing

Mohammed Mosuily
School of Electronics and Computer Science
University of Southampton
Southampton, UK
mtm1g19@soton.ac.uk

Jagmohan Chauhan
School of Electronics and Computer Science
University of Southampton
Southampton, UK
j.chauhan@soton.ac.uk

***Abstract*—EchoMotion revolutionizes exercise monitoring with its innovative acoustic-based smart speaker system, designed to perform human pose estimation using inaudible acoustic signals. Leveraging a combination of acoustic features, deep learning techniques, and a custom loss function, the system transforms acoustic reflections from the human body into precise 3D pose estimations. Ground truth data were recorded using the Microsoft Azure Kinect DK, a depth-sensing camera used for evaluation. Data were collected from 22 participants performing five fast-movement cardio exercises in both home and lab environments, yielding over 11 hours of synchronized acoustic and ground truth data. EchoMotion achieved a low Mean Absolute Error (MAE) of 0.59 mm, demonstrating superior accuracy for fast movement exercises compared to the reported MAE range of 2.8 mm to 96 mm in SOTA works, which also focus on slow movements. Our system is non-invasive, cost-effective, respects privacy, and is capable of performing in various acoustic conditions, making it an ideal tool for home-based exercise monitoring and feedback. EchoMotion's ability to analyze the exercise pose estimations provides valuable insights for users, trainers, and clinicians, enhancing the quality of remote exercise programs.**

***Index Terms*—Acoustic sensing, pose estimation, and home-based monitoring.**

I. Introduction

Cardiovascular diseases and other health-related conditions are a major concern worldwide [1]. Exercise plays a critical role in maintaining health and reducing the risk of these conditions [2]. However, there is a significant gap in the feedback personal trainers and clinicians receive from patients or individuals performing cardio exercises at home. Current feedback systems include live video streams of exercises (impacting privacy), providing images of the exercises, or requiring individuals to manually enter information into a diary for review by an expert. Additionally, the general healthcare cost for countries such as the United States has escalated to approximately $3 trillion per year [3]. Remote exercise monitoring has the potential to bridge these gaps due to the low cost of current technology and its ease of use. Thus, developing accessible and remote home-based solutions is crucial.

We developed EchoMotion to address the aforementioned limitations. Our system analyzes the reflections of active acoustic signals from the human body in motion, converting these signals into a human pose estimation, as illustrated in Figure 1. This approach significantly advances the integration of technology within healthcare and general fitness, offering a novel method for analyzing exercise in remote monitoring sessions. It enables the analysis of critical aspects of cardio exercises, such as form, and other aspects, such as duration and repetition. Our system successfully performed detailed motion tracking for five exercises with 22 participants. In addition, we conducted studies into the impact of distances, noise, and cross-subject training to test the robustness of the system.

While vision-based systems like Kinect offer robust pose tracking, they suffer from privacy concerns, sensitivity to lighting, and high hardware requirements. Similarly, RF and WiFi-based systems often require complex signal processing and custom hardware. In contrast, EchoMotion provides a cost-effective, privacy-preserving alternative using readily available microphones and inaudible acoustic signals, making it ideal for home-based monitoring.

EchoMotion's contributions can be summarized as follows:

- EchoMotion introduces the first non-invasive, acoustic-based smart speaker system designed to enhance the analysis of cardio exercises. By utilizing deep learning techniques, it accurately interprets acoustic signals to generate precise human pose estimations. The system also offers a cost-effective solution by leveraging readily available sensors such as microphones and speakers.
- A comprehensive acoustic dataset was collected from 22 individuals performing five cardiovascular exercises. Each session, including repetitions and rest periods, lasted 6 minutes per exercise. This dataset contains approximately 11 hours of data and 880,000 frames of exercises.
- EchoMotion achieved a low average Mean Absolute Error (MAE) of 0.59 mm, significantly lower than the MAE range of 2.8 mm to 96 mm reported in the literature. This was accomplished through a novel combination of acoustic features and a custom loss function, underscoring our technical contribution. Exercises were conducted at a distance of 1 meter under normal room acoustic conditions (40 dB). Additionally, our system demonstrates adaptability to various environments (home and lab) and conditions (1 m and 2 m), as well as different noise levels (60 dB and 80 dB), while maintaining high accuracy.

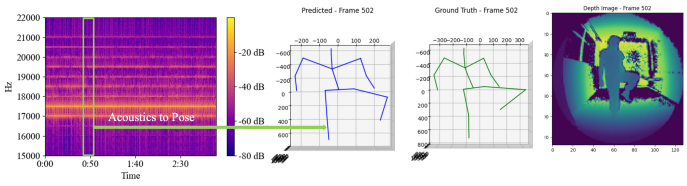

Fig. 1. EchoMotion: One Frame of High Knee Exercise

II. Related Works

This paper focuses on cardio exercises due to their ease for home-based workouts. Current methods for human motion sensing in terms of exercise or pose estimation are either, (1) Wearable, (2) Vision, (3) RF/WiFi, or (4) Acoustics.

(1) Sensor-Based Monitoring: Traditional monitoring methods in cardio exercises typically involve wearable sensors such as Inertial Measurement Units (IMUs) [4], [5] and Surface Electromyography (sEMG) [6]. While effective, these technologies can be intrusive or inconvenient. **(2) Vision-Based Monitoring:** Recent advances in computer vision have led to the use of systems like Kinect for exercise monitoring. [7]–[10]. However, despite the Kinect's robustness, it faces challenges such as a limited sensor range, the need for optimal lighting conditions, and privacy concerns [11]. **(3) RF and WiFi-Based Monitoring:** Radio Frequency (RF) and WiFi-based human motion sensing present privacy-conscious solutions that are not affected by light or temperature variations. Technologies like EMAS [12] GoPose [13] and MoRe-Fi [14] highlight the capabilities of these approaches, but often require specialized devices. **(4) Acoustics-Based Monitoring:** To the best of our knowledge, EchoMotion is the first method of cardio exercise form analysis utilizing acoustics. Current methodologies incorporating acoustics include Hear-Your-Action paper [15], HearFit [16], PoseSonic [17], and LoEar [18], each has their own limitations. For instance, Hear-Your-Action focuses solely on the classification of movement types. Additionally, the limited dataset's small size and lack of diversity restrict the model's ability to generalize across broader, real-world scenarios. while PoseSonic, a wearable smart glasses-based acoustic system, is limited to upper body detection. Hearfit, on the other hand, lacks capabilities in human posture detection and form analysis, primarily focusing on exercise classification.

A study by Shibata et al. [19] investigates the use of audible acoustic chirp signals for human pose estimation. However, their approach encounters limitations due to the reliance on loud, audible-range frequencies, which are impractical in real-world settings. Additionally, their method requires an expensive ambisonic microphone (priced at $1,299) and high-performance computing to support their large models, potentially making it inaccessible to the general public. Moreover, they present only abstract conceptualizations of the technology without offering practical applications. In contrast, EchoMotion achieves superior human posture estimation using readily available hardware, similar to commercial smart speakers, which makes it cost-effective. Furthermore, by utilizing inaudible acoustic signals and an efficient model for human pose estimation, EchoMotion reduces the need for high-performance computing.

III. Methodology

1) Data Collection and Recording Setup: Since there are no existing datasets for cardio exercise at home using acoustics, we collected a new dataset. Data was collected from 22 participants (16 males, 6 females), aged 18-44, representing seven ethnic backgrounds, with heights ranging from 164 to 187 cm, weights from 50 to 95 kg, and Body Mass Indexes (BMIs) from 18.3 to 32.4. The majority of participants wore standard sports attire, while a few opted for slightly loose clothing. The data was captured in: (a) a living room (6.62 $\times$ 3.6 m) for 13 subjects (Figure 3), and (b) a laboratory (6 $\times$ 2.7 m) for 9 subjects (Figure 5). In consultation with a personal trainer and medical professionals, we selected exercises, Air Punches, Seated Leg Extensions, Jumping Jacks, Squats, and High Knees, to evaluate EchoMotion's performance in typical cardio routines [20]. These exercises target various regions of the body, including the upper/lower body, seated/standing positions, and large motion movements, demonstrating that our system is capable of adapting to a wide range of exercises.

The recording setup involved the MiniDSP UMA-8 USB microphone array V2.0 with 7 microphones to record acoustic signals, a Bose Companion 2 speaker for tone generation, and the Azure Kinect DK for collecting ground truth 3D pose data [21]. The reason for selecting the Kinect over more expensive options (e.g., the $12k OptiTrack Mocap) was its affordability, availability, and prior successful applications [13]. Ground truth data consisted of the X, Y, and Z coordinates of 13 key body joints, including the pelvis, shoulders, elbows, wrists, hips, knees, and ankles. Ultrasonic tones, generated at frequencies ranging from 17k to 21k Hz, were recorded at a 44.1 kHz sample rate with 1024 frames per audio buffer. These continuous tones enabled action recognition by capturing reflections from human movements while minimizing interference from environmental noise. To evaluate the system's robustness, we introduced background music during select recording sessions using a secondary speaker positioned 1 meter behind the participant. The same exercises were then repeated while music was played at controlled noise levels of 60 dB and 80 dB, calibrated using a sound level meter to simulate realistic home environments. The signals were recorded and stored in WAV format for further analysis. Figure 6 illustrates the acoustic signal processing pipeline. Each recording session captured five exercises, lasting approximately 6 minutes each, with 30 seconds of rest between sets. We recorded around 11 hours of data, yielding approximately 880,000 frames. Ethics approval was obtained for our study from the University of Southampton Ethics and Research Governance Committee under reference number 80815 (ERGO II). The dataset and code are publicly available at https://github.com/MohammedMosuily/EchoMotion.

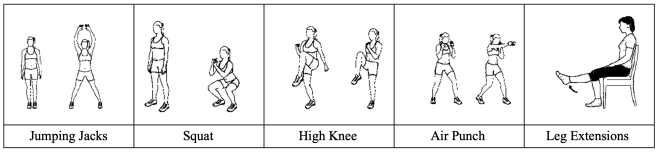

Fig. 2. List of Exercises (Jumping Jacks, Squat, High Knee, Air Punch, Leg Extensions)

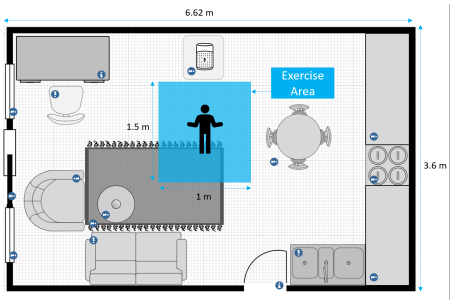

Fig. 3. Data Collection in the living room

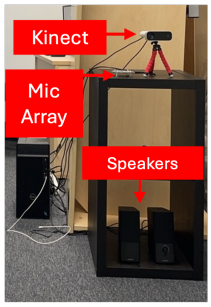

Fig. 4. Hardware setup: Kinect, mic array, and speaker positions.

2) Feature Extraction and Model Architecture: For feature extraction, we employed two key methods the Mel-spectrogram [19] and spectral contrast. First, the Mel-spectrogram converts the short-time Fourier transform (STFT) of the signal into a Mel-scaled representation, which reflects how humans perceive sound frequencies, even for ultrasonic signals. It is computed as follows:

$$\text{Mel-Spectrogram}(f,t) = \log\left(\sum_{k=0}^{K} |X(k,t)|^2 H_m(k)\right) \quad (1)$$

where $X(k,t)$ represents the magnitude of the STFT at frequency bin k and time t, and $H_m(k)$ is the Mel-filter bank applied to the magnitude. The Mel-spectrogram emphasizes frequency bands that are relevant for capturing changes in human movement from the reflected ultrasound signals. Second, spectral contrast measures the difference in amplitude between peaks (high-energy components) and valleys (low-energy components) in the frequency spectrum. This helps distinguish between various human actions based on the structure of the reflected ultrasound waves. The spectral contrast for each frequency band at time t is computed as:

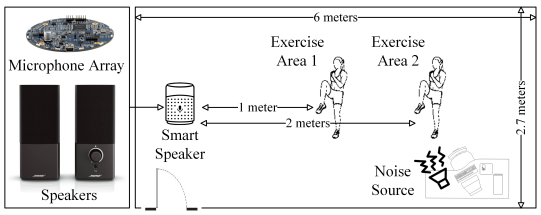

Fig. 5. Data Collection Setup in the lab

$$\text{Spectral Contrast}(t) = \log\left(\frac{\max_{k\in[f_{\text{low}},f_{\text{high}}]} X(k,t)}{\min_{k\in[f_{\text{low}},f_{\text{high}}]} X(k,t)}\right) \quad (2)$$

where $X(k,t)$ is the magnitude of the STFT at frequency bin k and time t, and $[f_{\text{low}}, f_{\text{high}}]$ denotes the frequency range of interest for each band. This log-scaled ratio of peak to valley magnitude provides insight into the tonal characteristics of the sound across different frequency ranges. By combining both these features, we obtained features that capture both the time-varying and frequency-varying characteristics of the signals, which were used as inputs to the deep learning model.

3) Training Procedure and Optimization: For this study, we employed a sequential neural network model to predict 3D pose data based on the extracted audio features. The sequential model consisted of 5 dense layers with batch normalization and dropout layers to prevent overfitting. The final layer of the model produced the 3D coordinates of human joints using a linear activation function. The custom loss function was designed to minimize prediction errors for key joints such as the elbows and wrists, which are critical for accurate pose estimation. This loss function is based on mean squared error (MSE) with additional weighting applied to specific joint coordinates, and is defined as:

$$\mathcal{L}(y_i,\hat{y}_i) = \text{MSE}(y_i,\hat{y}_i) \times \text{Weighted Joint Loss}(y_i,\hat{y}_i) \quad (3)$$

where:

$$\text{Weighted Joint Loss}(y_i,\hat{y}_i) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \quad (4)$$

Where y_i is the actual value of the observation (Kinect), and $\hat{y}_i$ is the predicted value (EchoMotion) for the same observation, across all measurements.

This ensures the model focuses more on key joints during training, improving the overall pose prediction accuracy. Moreover, the dataset was split into training and testing sets using an 80/20 ratio, and both the input features and target labels were normalized. We trained the model using a batch size of 64 for up to 100 epochs with the Adam optimizer, starting with a learning rate of 0.001, and used an RTX 3070 GPU. Early stopping was employed to halt training when no validation loss improvement was detected for 10 epochs.

IV. Experiments and Results

To evaluate the accuracy of our methodology, we employ Equation 5 to calculate the Mean Absolute Error (MAE) for both single-subject and cross-subject analysis in pose estimation between EchoMotion (Acoustics) and Kinect (Camera-Based), measured in millimetre (mm) and we used Equation 6 for the Root Mean Squared Error (RMSE). The ideal value for MAE and RMSE should be zero. In addition, Figures 7 and 8 show that varying noise levels (40 dB, 60 dB, 80 dB) and participant distances (1 m, 2 m) do not affect EchoMotion, as the MAE remains steady at 0.6 mm across five participants.

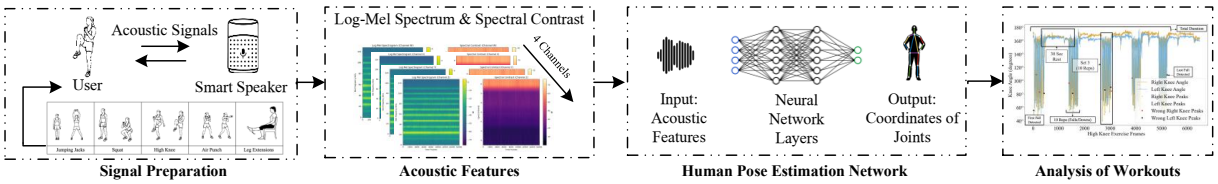

Fig. 6. Flowchart representing acoustic signal analysis and processing pipeline for EchoMotion

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

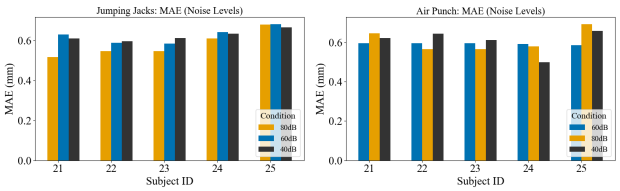

Fig. 7. Impact of noise at 40dB, 60dB, and 80dB (5 participants)

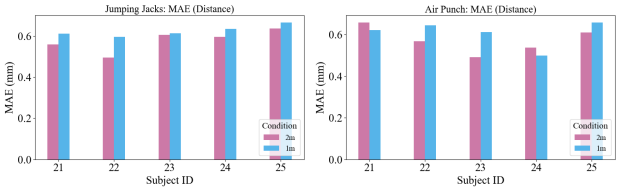

Fig. 8. Impact of distance at 1 m and 2 m (5 participants)

Where y_i is the actual value of the observation (Kinect), and $\hat{y}_i$ is the predicted value (EchoMotion) for the same observation, across all measurements.

1) Single Subject Human Pose Estimation: The exercises were recorded at a distance of 1 meter under normal room acoustics (40 dB). Table I summarizes the performance for five exercises: Air Punch (AP), High Knees (HK), Jumping Jacks (JJ), Leg Extensions (LE), and Squats (SQ). In the living room, Leg Extensions (LE) achieved the lowest MAE of 0.551 mm, while Jumping Jacks (JJ) and High Knees (HK) had slightly higher MAEs of 0.582 mm and 0.622 mm, respectively. In the lab, LE again had the lowest MAE at 0.506 mm, with the other exercises showing MAEs between 0.598 mm and 0.625 mm. The RMSE ranged between 0.785 mm and 0.908 mm, demonstrating consistent accuracy across exercises and environments.

2) Cross-Subject Human Pose Estimation: We utilized Equations (5) and (6) to evaluate our results in cross-subject training. To accommodate differences in body dimensions between subjects, we fine-tuned our model. Using a leave-one-out approach, we trained on all subjects except one, using the remaining subject for testing. Table II presents the results: MAE values ranged from 0.612 mm to 1.280 mm across various exercises in the living room environment, with the lowest MAE for High Knees (HK) at 0.612 mm and the highest for Leg Extensions (LE) at 1.280 mm. In the lab environment, MAE values ranged from 0.911 mm to 2.717 mm, with the lowest error for High Knees (HK) and the highest for Air Punch (AP). The RMSE values ranged from 0.787 mm to 1.696 mm in the living room, and from 1.242 mm to 3.200 mm in the lab. The higher errors in the lab environment reflect greater variations in body types, clothing, and environmental factors between subjects.

TABLE I
POSE ESTIMATION RESULTS FOR MAE AND RMSE IN MM, ACROSS LIVING ROOM AND LAB SETTINGS.

Living Room for subjects (1-13)					
Statistic	**Exercise**				
	LE	**JJ**	**HK**	**AP**	**SQ**
μ(MAE)	0.551	0.582	0.622	0.603	0.620
σ^2(MAE)	0.002	0.005	0.003	0.005	0.009
μ(RMSE)	0.810	0.890	0.896	0.871	0.895
σ^2(RMSE)	0.002	0.004	0.002	0.004	0.006
Lab for subjects (14-22)					
μ(MAE)	0.506	0.598	0.622	0.625	0.618
σ^2(MAE)	0.001	0.001	0.005	0.003	0.005
μ(RMSE)	0.785	0.908	0.903	0.878	0.908
σ^2(RMSE)	0.002	0.002	0.002	0.003	0.002

TABLE II
RESULTS OF POSE ESTIMATION FOR CROSS-SUBJECTS, MAE AND RMSE.

Exercise	**Living Room**		**Lab**	
	MAE (mm)	RMSE (mm)	MAE (mm)	RMSE (mm)
JJ	1.067	1.354	1.378	1.968
HK	0.612	0.787	0.911	1.242
AP	0.894	1.138	2.717	3.200
SQ	1.060	1.296	1.036	1.437
LE	1.280	1.696	1.133	1.573

3) Comparison to Existing Systems: Existing work on pose estimation is shown in Table III. EchoMotion achieves a lower MAE with a higher number of participants for single-subject pose estimation across all exercises compared to other baseline systems utilizing different modalities. While the comparison is not entirely fair due to differences in participant numbers and exercise types, it still provides an indication of EchoMotion's

performance relative to existing systems. The closest work to ours is by Shibata et al. [19], which achieved the lowest reported MAE of 2.8 mm. However, their system relied on expensive microphones, a MoCap suit, and an audible low-frequency chirp signal, making it difficult to directly compare with our system. In contrast, EchoMotion uses inaudible tones in the 17kHz-21kHz range for human pose estimation, making it a novel and more practical approach using acoustics. To the best of our knowledge, no existing work utilizes this frequency range for predicting human pose estimation, and EchoMotion's performance with an MAE of 0.59 mm demonstrates competitive accuracy at fast body movement.

TABLE III
COMPARISON OF ECHOMOTION WITH REPORTED PERFORMANCE AND SENSING TECHNIQUES FOR POSE ESTIMATION BY MOVEMENT SPEED

Method	Modality	Subjects	MAE (mm)	Speed
GoPose [13]	RF/Wifi	10	47	Medium
BodyTrak [10]	RGB	12	69	Slow
PoseSonic [17]	Audio	22	61	Slow
Shibata et al. [19]	Audio	8	2.8	Slow
(Ours)	**Acoustics**	**22**	**0.59**	**Fast**

V. CONCLUSION

In conclusion, EchoMotion marks a significant advancement in acoustic-based human pose estimation, introducing a non-invasive smart speaker system that leverages acoustics and deep learning for precise exercise analysis. We achieved a low error for human pose estimation using our custom loss function, with an overall performance of MAE 0.59 mm compared to the reported MAE of 2.8 mm to 96 mm in other systems. EchoMotion's adaptability to various environments, including both living room and lab settings, highlights its potential for widespread application in home-based exercise monitoring. The system's ability to provide feedback and adapt exercises to individual needs can significantly enhance the quality of life for users. Future work includes refining the predicted pose estimation frames.

REFERENCES

[1] M. Vaduganathan, G. A. Mensah, J. V. Turco, V. Fuster, and G. A. Roth, "The Global Burden of Cardiovascular Diseases and Risk," *Journal of the American College of Cardiology*, vol. 80, no. 25, pp. 2361–2371, 2022. [Online]. Available: https://www.jacc.org/doi/abs/10.1016/j.jacc.2022.11.005

[2] L. Ciumărnean, M. V. Milaciu, V. Negrean, O. H. Orășan, S. C. Vesa, O. Sălăgean, S. Iluț, and S. I. Vlaicu, "Cardiovascular Risk Factors and Physical Activity for the Prevention of Cardiovascular Diseases in the Elderly," *International Journal of Environmental Research and Public Health*, vol. 19, no. 1, 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/1/207

[3] C. Garvey, J. P. Singer, A. M. Bruun, A. Soong, J. Rigler, and S. Hays, "Moving Pulmonary Rehabilitation into the Home: A CLINICAL REVIEW," *Journal of Cardiopulmonary Rehabilitation and Prevention*, vol. 38, no. 1, pp. 8–16, 1 2018.

[4] L. E. Osborn, M. M. Iskarous, and N. V. Thakor, "Chapter 22 - Sensing and Control for Prosthetic Hands in Clinical and Research Applications," in *Wearable Robotics*, J. Rosen and P. W. Ferguson, Eds. CA, United States: Academic Press, 2020, ch. 22, pp. 445–468. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128146590000229

[5] G. Spina, G. Huang, A. Vaes, M. Spruit, and O. Amft, "COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 597–606. [Online]. Available: https://doi.org/10.1145/2493432.2493454

[6] Y. Zhou, Y. Fang, K. Gui, K. Li, D. Zhang, and H. Liu, "sEMG Bias-Driven Functional Electrical Stimulation System for Upper-Limb Stroke Rehabilitation," *IEEE Sensors Journal*, vol. PP, p. 1, 2 2018.

[7] Z. Zhang, "Microsoft Kinect Sensor and Its Effect," *IEEE MultiMedia*, vol. 19, no. 2, pp. 4–10, 2012.

[8] F. Mortazavi and A. Nadian-Ghomsheh, "Continues online exercise monitoring and assessment system with visual guidance feedback for stroke rehabilitation," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 32 055–32 085, 2019. [Online]. Available: https://doi.org/10.1007/s11042-019-08020-2

[9] X. Yu and S. Xiong, "A dynamic time warping based algorithm to evaluate Kinect-enabled home-based physical rehabilitation exercises for older people," *Sensors (Switzerland)*, vol. 19, no. 13, 7 2019.

[10] H. Lim, Y. Li, M. Dressa, F. Hu, J. H. Kim, R. Zhang, and C. Zhang, "BodyTrak: Inferring Full-body Poses from Body Silhouettes Using a Miniature Camera on a Wristband," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 3, 9 2022. [Online]. Available: https://doi.org/10.1145/3552312

[11] M. Tölgyessy, M. Dekan, Chovanec, and P. Hubinský, "Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2," *Sensors (Switzerland)*, vol. 21, no. 2, pp. 1–25, 1 2021.

[12] D. Fan, X. Yang, N. Zhao, L. Guan, and Q. H. Abbasi, "Exercise Monitoring and Assessment System for Home-Based Respiratory Rehabilitation," *IEEE Sensors Journal*, vol. 22, no. 19, pp. 18 890–18 902, 2022.

[13] Y. Ren, Z. Wang, Y. Wang, S. Tan, Y. Chen, and J. Yang, "GoPose: 3D Human Pose Estimation Using WiFi," in *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 6, no. 2. New York, NY, USA: Association for Computing Machinery, 7 2022, pp. 1–25. [Online]. Available: https://doi.org/10.1145/3534605

[14] T. Zheng, Z. Chen, S. Zhang, C. Cai, and J. Luo, "MoRe-Fi: Motion-robust and Fine-grained Respiration Monitoring via Deep-Learning UWB Radar," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 111–124. [Online]. Available: https://doi.org/10.1145/3485730.3485932

[15] R. Tanigawa and Y. Ishii, "Hear-Your-Action: Human Action Recognition by Ultrasound Active Sensing," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 7260–7264.

[16] Y. Xie, F. Li, Y. Wu, and Y. Wang, "HearFit: Fitness Monitoring on Smart Speakers via Active Acoustic Sensing," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*. Vancouver, CA: IEEE, 6 2021, pp. 1–10.

[17] S. Mahmud, K. Li, G. Hu, H. Chen, R. Jin, R. Zhang, F. Guimbretière, and C. Zhang, "PoseSonic: 3D Upper Body Pose Estimation Through Egocentric Acoustic Sensing on Smartglasses," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, no. 3, 9 2023. [Online]. Available: https://doi.org/10.1145/3610895

[18] L. Wang, W. Li, K. Sun, F. Zhang, T. Gu, C. Xu, and D. Zhang, "LoEar: Push the Range Limit of Acoustic Sensing for Vital Sign Monitoring," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 3, 9 2022.

[19] Y. Shibata, Y. Kawashima, M. Isogawa, A. Kimura, and Y. Aoki, "Listening Human Behavior: 3D Human Pose Estimation with Acoustic Signals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 8 2023, pp. 13 323–13 332.

[20] Asthma Lung UK, "Pulmonary Rehabilitation," https://www.asthmaandlung.org.uk/living-with/keeping-active/pulmonary-rehabilitation, 2023.

[21] T. Tran, D. Ma, and R. Balan, "Remote Multi-Person Heart Rate Monitoring with Smart Speakers: Overcoming Separation Constraint," *Sensors*, vol. 24, no. 2, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/2/382