Report on Signal Analyses of Swallowing Events

1st Zihang Huang
Department of Engineering
Kings College London
London, United Kingdom
k21094939@kcl.ac.uk

2nd Tengjun Jiao
Department of Engineering
Kings College London
London, United Kingdom
k22038903@kcl.ac.uk

3rd Beiwen Tan
Department of Engineering
Kings College London
London, United Kingdom
k22017497@kcl.ac.uk

4th Chengyue Zhang
Department of Engineering
Kings College London
London, United Kingdom
k22017304@kcl.ac.uk

5th George Taylor
Department of Engineering
Kings College London
London, United Kingdom
k22028968@kcl.ac.uk

***Abstract*—This report details a study on analyzing swallowing audio signals using microphones to collect 500 instances across 5 participants, involving water and food swallowing events. It focuses on time-domain signals, signal size, frequency spectra, bandwidth estimation, and event correlation.**

I. Introduction

The concept of swallowing, pivotal for fluid and food intake, encompasses a complex sequence of events critical for nutritional absorption and overall health. Swallowing involves multiple phases: the oral phase, where the bolus is prepared and moved from the oral cavity to the pharynx; the pharyngeal phase, ensuring the safe transfer of the bolus through the pharynx to the esophagus; and the esophageal phase, characterized by the peristaltic movement of the esophagus to transport the bolus into the stomach [1]. This intricate process, largely involuntary but also under voluntary control, underscores the body's remarkable ability to ensure that sustenance reaches the stomach safely for digestion, highlighting the significance of swallowing in daily survival and quality of life.

Signal analysis of swallowing event's audio signals is an innovative approach to understanding and improving the diagnostic capabilities related to dysphagia and other swallowing disorders. By capturing and analyzing the acoustic signatures of swallowing events, researchers and clinicians can uncover patterns and anomalies that may not be evident through traditional diagnostic methods. This report details a comprehensive study where audio signals of 100 swallowing events from each of five students were meticulously recorded, filtered, and analyzed. Through the application of advanced signal processing techniques and statistical analysis, the report presents findings that enhance our understanding of the swallowing process. It demonstrates how filtering methods can isolate relevant features from background noise and how statistical data can elucidate the characteristics and variability of swallowing sounds. This analysis is crucial for developing non-invasive, cost-effective diagnostic tools and tailored rehabilitation strategies to address swallowing difficulties, thereby contributing significantly to the fields of speech pathology and rehabilitative medicine.

II. Method of Data Collection

A. Hardware Selection

The hardware chosen for recording the audio signals of swallowing events includes standalone microphones capable of attaching to the throat area, ensuring close proximity to the source of the sound. To address the challenge of noise reduction, particularly noise created by movement of the throat during swallowing, adhesive tape was employed. This method is effective in minimizing extraneous sounds without interfering with the integrity of the swallowing sounds being recorded. For capturing the nuanced sounds of swallowing with minimal background noise, omnidirectional microphones such as the MAONO PD400XS and Razer Seiren V3 Cardioid Microphones are highly regarded within the field recording community for their low noise level and high sensitivity, making them ideal for this purpose.

In selecting the materials for the study, water was chosen as the liquid for participants to swallow due to its neutrality and consistency. Water does not alter the viscosity and sound characteristics of swallowing in the way thicker liquids might, providing a standardized baseline for comparison across all participants. For solid foods, hard surface items were selected to amplify the significant sounds associated with the swallowing process.

B. Signal Selection

The selection of signals was conducted with a rigorous standard to ensure the clarity and relevance of the data collected. Only signals with minimal background noise were accepted for analysis. This necessitated a process wherein any recordings that contained a significant amount of noise or interference were identified and excluded from the dataset. Participants were then asked to re-record these events under more controlled conditions to meet the study's stringent requirements for signal purity. By controlling for ambient noise, the integrity of the audio signals is preserved, allowing for a more precise analysis. Participants are instructed to find a quiet room where external sounds are minimized, ensuring that the microphones capture only the intended swallowing sounds. This meticulous approach to recording in a noise-free setting is critical for the success of the study, as it significantly reduces the need for extensive filtering and cleaning of the audio data post-recording.

III. Signal Analyses Methods

A. Low-Pass Filtering

Low-Pass Filtering (LPF) plays a crucial role in analyzing audio signals from swallowing events, especially in isolating the relevant low-frequency components essential for accurate assessment. A low-pass filter effectively allows signals below a specific cutoff frequency to pass through while attenuating frequencies above this threshold [2]. For a LPF, the transfer function is given by $H(f) = \frac{1}{1+j\frac{f}{f_c}}$. In which f is the frequency of input signal, f_c is the cutoff frequency [3].

Through rigorous testing and analysis, a cutoff frequency of 6000 Hz has been established for the Low-Pass Filtering (LPF) of swallowing event audio signals. This decision ensures that the LPF only permits signals with frequencies below this threshold, effectively isolating the crucial low-frequency components that are most indicative of the swallowing process, while significantly reducing the higher-frequency noise and artifacts that may not be relevant to the assessment.

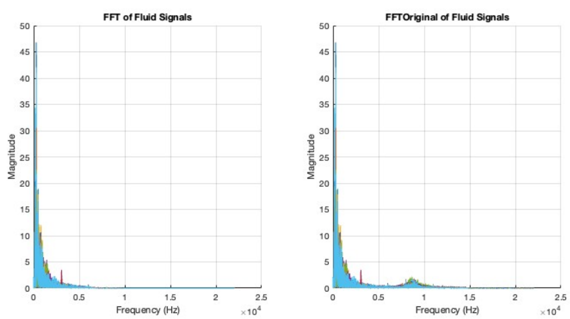

Fig. 1. Filtered FFTs Vs. Original FFTs

B. Signal Quantification Techniques

1) Signal Energy and Power: Energy and power, fundamental concepts in physics and engineering, serve distinct yet interconnected roles in quantifying physical systems. Energy, measured in joules, encapsulates the capacity to perform work or produce heat, embodying various forms such as kinetic, potential, thermal, and electrical energy. Power, on the other hand, quantified in watts, represents the rate at which energy is transferred or converted per unit of time. Essentially, power measures the speed at which work is done or energy is expended [4].

In the context of analyzing swallowing events, energy and power metrics are pivotal for interpreting the dynamics and efficiency of the swallowing process. For instance, the energy in the audio signals of swallowing events can highlight the intensity and duration of muscle contractions involved, offering insights into the physiological aspects of swallowing. Similarly, analyzing the power of these signals could provide information on how quickly energy is being used during swallowing, indicating the effort or efficiency of the process.

For discrete signal represented as $x[n]$, where n is the discrete time index, the total energy E is calculated by summing the squares of the signal's amplitude at each point over all n, expressed as $E = \sum_{n=-\infty}^{\infty} |x[n]|^2$. The power can be calculated by averaging the squared amplitudes over one period N or over the segment's length: $P = \frac{1}{N}\sum_{n=0}^{N-1} |x[n]|^2$.

C. Frequency-Domain Analysis

1) Fourier Transform: The Fourier Transform (FT) is a powerful mathematical tool used to analyze the frequency components of signals, transforming them from the time domain into the frequency domain [5]. For a discrete signal $x[n]$, where n represents discrete time indices, the Discrete Fourier Transform (DFT) is given by the equation: $X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N}$, where $X[k]$ is the frequency domain representation of $x[n]$, N is thetotal number of samples, k is the discrete frequency indices, and j is the square root of -1.

This transformation allows us to view the signal in terms of its constituent frequencies, providing insights into its spectral properties.

Analyzing swallowing signals with FT is particularly beneficial due to the complex nature of these signals, which can exhibit variations in amplitude, phase lags, and frequency content due to the physiological differences in swallowing mechanisms among individuals. The FT helps in decomposing these signals into their frequency components, making it easier to identify specific characteristics such as the dominant frequencies of swallowing sounds, which may vary due to factors like texture of food or liquid being swallowed, or the presence of swallowing disorders.

Moreover, by using the FT, researchers can effectively filter out noise and focus on relevant frequency bands, enhancing the analysis and interpretation of swallowing sounds. This frequency domain analysis facilitates the comparison of swallowing events across different conditions and individuals by standardizing the assessment based on spectral content rather than raw time-domain signals, which may be more susceptible to variations and noise.

2) Essential Bandwidth: The concept of essential bandwidth refers to the frequency range within which the majority of a signal's power is concentrated, encapsulating the most significant components of the signal's energy. This metric is particularly useful in signal processing for distinguishing the critical portions of a signal from those containing negligible information or noise [6]. In the analysis of swallowing audio signals, identifying the essential bandwidth allows researchers to focus on the frequency ranges most relevant to the physiological aspects of swallowing.

The calculation typically involves integrating the signal's power spectral density over frequency and identifying the minimum and maximum frequencies that enclose the defined percentage of the signal's power. This process highlights the frequency components that are crucial for the signal's characterization, thereby guiding interventions or further analyses aimed at understanding or modifying swallowing function.

IV. Statistical Result Analyses

A. Difference in Energy and Power

In signal analysis, particularly when dealing with varied and potentially erratic datasets like swallowing audio signals, the median value of the energy of each signal is chosen over the average due to its robustness against outliers. Outliers can significantly skew the average, leading to a distorted representation of the data's central tendency. The median, being the middle value of a dataset when ordered by magnitude, is less affected by extreme values, providing a more reliable measure for central tendency in the presence of outliers.

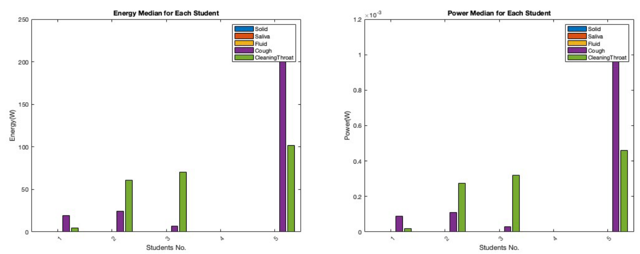

Fig. 2. Median of Energy and Power for Different Swallowing Event For Each Student

The plot comparing the median of energy and power for each swallowing event among 5 samples illustrates this point clearly. It's evident that coughing and throat clearing generate substantially more energy and power compared to other swallowing events. This could be attributed to the increased effort and muscle activity involved in these actions, resulting in higher amplitude signals.

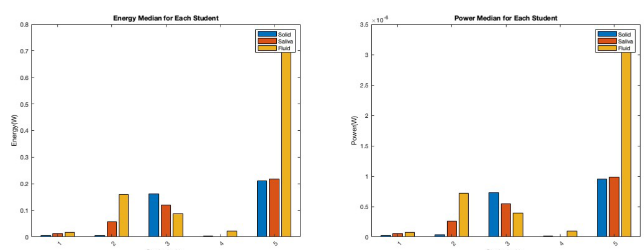

Fig. 3. Median of Energy and Power for Different Swallowing Event For Each Student, Excluding Coughing and Clearing Throat

When removing these two values from the chart and focusing solely on the other three types of swallowing events, we observe a general similarity in amplitude across these events for each student. However, an exception is noted in sample 5, where swallowing fluids is marked by significantly higher amplitudes. This outlier could suggest an individual characteristic of the student's swallowing technique, possibly a tendency to swallow water more forcefully, thus producing louder swallowing sounds.

B. Difference in Essential Bandwidth

Student / Event	**Solid**	**Saliva**	**Fluid**	**Cough**
1	1642.4 Hz	1590.4 Hz	856.1 Hz	2048.7 Hz
2	1374.8 Hz	1023.3 Hz	484.7 Hz	1668.8 Hz
3	3108.4 Hz	2936.5 Hz	2730.3 Hz	3746.8 Hz
4	4179.7 Hz	2683.0 Hz	2349.7 Hz	6280.9 Hz
5	1472.2 Hz	1495.0 Hz	1452.1 Hz	1353.5 Hz

TABLE I
Median Value of Essential Bandwidth of Swallowing Event Signals for Five Students

In analyzing the essential bandwidth data from the five students' samples for various swallowing events, we observe a range of values that reflect the diversity in the acoustic properties of each event type. The bandwidth for cough events generally exhibits higher values compared to other events, which may indicate a broader range of frequencies involved in the cough reflex, a more forceful and variable action. The outliers, particularly the cough event from the fourth student showing a bandwidth of 6280.9 Hz, and the cleaning throat event from the second student with 2799.4 Hz, suggest significantly broader frequency engagement compared to their peers. These anomalies could result from various factors, such as individual differences in anatomy, a more forceful execution of the event, or differences in the recording setup.

C. The Need for Normalisation

Moreover, the noticeable amplitude variation between samples underscores the impact of using different recording devices, each with its distinct default gains. This variance highlights the technical aspect of signal collection that can influence the data analysis, emphasizing the importance of standardization or calibration in comparative studies to account for equipment-induced discrepancies. Such differences in recording setup can lead to varying baseline levels of energy and power in the captured signals, complicating direct comparison without appropriate normalization or adjustment for device-specific characteristics.

Therefore, an additional normalized FFT signal is calculated by dividing each frequency's FFT signal magnitude by the largest magnitude to normalize. The normalization process in signal analysis is a crucial step for ensuring that comparisons between different signals focus on the structural aspects of the data rather than their absolute magnitudes.

D. Frequency Domain Graphs

Analyzing the FFT (Fast Fourier Transform) graphs of swallowing signals across different students, it's evident from the frequency domain representation that the same type of swallowing event produces a similar spectral shape. The consistency in the frequency content for the same event type

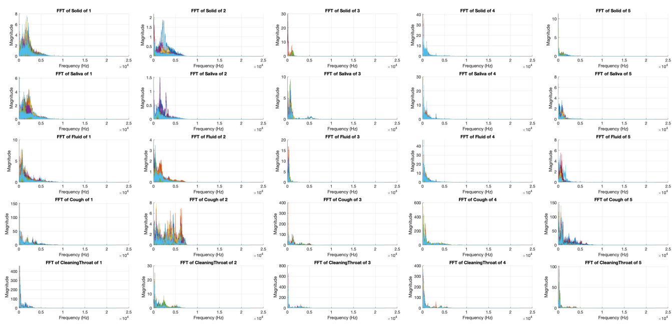

Fig. 4. FFT Plots of 5 Students

reflects the physiological uniformity of the swallowing process among individuals.

When examining the cough signals, we observe considerable variability between different students. This variation is likely due to the personal nature of the cough reflex, which is influenced by factors such as lung capacity, throat muscle strength, and even individual habits or manners of coughing. Each person's cough is as distinct as their voice, carrying unique characteristics that can cause significant differences in the frequency and magnitude of the FFT representation.

The consistent patterns in the FFT graphs for other types of swallowing events suggest that these actions are more uniform across individuals, possibly due to their less voluntary nature compared to coughing. For example, the act of swallowing fluid tends to follow a more standardized sequence of muscle contractions and relaxations, leading to more predictable acoustic signatures.

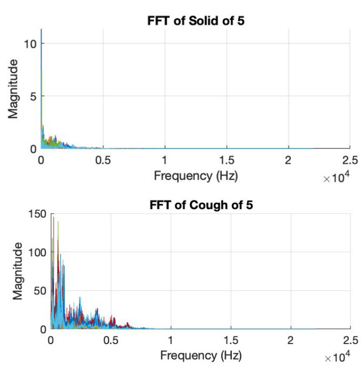

Fig. 5. Comparing Swallowing Solid vs Coughing for Student No.5

When visualizing the frequency domain analysis of repetitive swallowing samples from the same student, using a different color for each trial, we can expect a few insights. The plots of solid and fluid swallowing events would likely show consistent frequencies and magnitudes across the 20 samples. In contrast, coughing events could display a wide variance in the spectral content from one instance to another. This variability may arise from the spontaneous and less controlled nature of coughing, which is often a reflexive response to clear the throat or airways.

E. Correlation Analyses

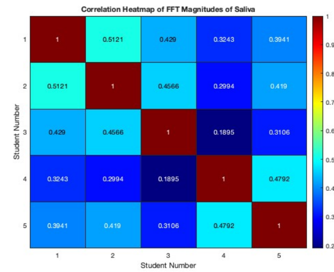

Fig. 6. Correlation Heatmap of FFTs for Same Event

The correlation heatmap of FFT magnitudes for the saliva swallowing event across five different students is a tool to examine the similarities or dissimilarities in the frequency domain characteristics of the event as recorded by each student. The heatmap presents correlation coefficients between the FFT outputs, where a value of 1 would indicate perfect correlation and 0 no correlation.

The highest correlation value between any two students' saliva swallowing events is 0.5666. This indicates that while there's some degree of similarity in the spectral content of the events between those two students, the correlation isn't exceptionally strong. The varied correlation values suggest that each student's swallowing sound has unique spectral characteristics, which could be influenced by individual anatomical features, the recording environment, or microphone placement.

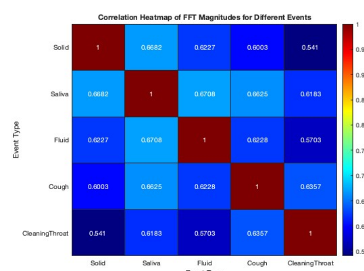

Fig. 7. Correlation Heatmap of FFTs for Different Event

The correlation for the different swallowing events of the same student, student number 5, shows values that are closer to 0, indicating a nearly non-existent correlation between the normalized FFT signals of different events. This low correlation suggests that each type of swallowing event—solid, saliva, fluid, cough, and clearing throat—has a distinct spectral signature.

This lack of correlation is meaningful; it implies that the physiological and acoustic properties involved in each event type are unique enough to be distinctly recognized and separated from one another through FFT analysis. Such differentiation is crucial in biomedical signal processing, as it allows for the identification and classification of various sounds related to human health.

PEER ASSESSMENT OF CONTRIBUTION

Member	Tasks	Contribution
Zihang Huang	Group leader, overseeing data collection, developing analytical strategies, writing MATLAB code, creating statistical tables and plots, leading report composition	50%
George Taylor	Researching data collection protocol, authoring the report's data collection section, verifying data quality	12.5%
Beiwen Tan	Developing MATLAB codes for Fourier transforms and signal quantization	12.5%
Tengjun Jiao	Developing MATLAB codes for Fourier transforms and signal quantization	12.5%
Chengyue Zhang	Creating statistical plot codes for graphical presentations	12.5%

TABLE II
TEAM MEMBER CONTRIBUTIONS

REFERENCES

[1] B. M. Popkin, K. E. D'Anci, and I. H. Rosenberg, "Water, hydration, and health," *Nutrition Reviews*, vol. 68, no. 8, pp. 439–458, 08 2010. [Online]. Available: https://doi.org/10.1111/j.1753-4887.2010.00304.x

[2] SoundScapeHQ. (2024, 03) What is a low-pass filter? [Online]. Available: https://soundscapehq.com/what-is-a-low-pass-filter/

[3] R. Keim, "Understanding low-pass filter transfer functions," 2019, accessed: 2023-03-28.

[4] Tutorials Point, "Signals and systems: Energy and power signals," 2021.

[5] J. James, *A Student's Guide to Fourier Transforms*, 3rd ed., ser. Student's Guides. Cambridge University Press, 2011.

[6] A. Mojahed, L. A. Bergman, and A. F. Vakakis, "Generalization of the concept of bandwidth," *arXiv preprint arXiv:2110.06770*, 2021. [Online]. Available: https://ar5iv.org/pdf/2110.06770