Perilaryngeal-Cranial Functional Muscle Network Differentiates Vocal Tasks: A Multi-Channel sEMG Approach

Rory O'Keeffe, Seyed Yahya Shirazi, Sarmad Mehrdad, *Graduate Student Member, IEEE*, Tyler Crosby, Aaron M. Johnson, and S. Farokh Atashzar, *Senior Member, IEEE*

Abstract*—*Objective: **Objective evaluation of physiological responses using non-invasive methods for the assessment of vocal performance and voice disorders has attracted great interest. This paper, for the first time, aims to implement and evaluate perilaryngeal-cranial functional muscle networks. The study investigates the variations in topographical characteristics of the network and the corresponding ability to differentiate vocal tasks.** ***Method:*** **Twelve surface electromyography (sEMG) signals were collected bilaterally from six perilaryngeal and cranial muscles. Data were collected from eight subjects (four females) without a known history of voice disorders. The proposed muscle network is composed of pairwise coherence between sEMG recordings. The network metrics include (a) network degree and (b) weighted clustering coefficient (WCC).** ***Results:*** **The varied phonation tasks showed the median degree, and WCC of the muscle network ascend monotonically, with a high effect size ($|r_{rb}| \sim 0.5$). Pitch glide, singing, and speech tasks were significantly distinguishable using degree and WCC ($|r_{rb}| \sim 0.8$). Also, pitch glide had the highest degree and WCC among all tasks (degree> 0.7, WCC> 0.75). In comparison, classic spectrotemporal measures showed far less effectiveness (max $|r_{rb}| = 0.12$) in differentiating the vocal tasks.** ***Conclusion:*** **Perilaryngeal-cranial functional muscle network was proposed in this paper. The study showed that the functional muscle network could robustly differentiate the vocal tasks while the classic assessment of muscle activation fails to differentiate.** ***Significance:*** **For the first time, we demonstrate the power of a perilaryngeal-cranial muscle network as a neurophysiological window to vocal performance. In addition, the study also discovers tasks with the highest network involvement, which may be utilized in the future to monitor voice disorders and rehabilitation.**

***Index Terms*—Voice disorder, intermuscular coherence, surface electromyography, neurophysiology.**

Manuscript received 12 September 2021; revised 3 April 2022 and 26 April 2022; accepted 6 May 2022. Date of publication 20 May 2022; date of current version 22 November 2022. This work was supported by the National Science Foundation under Awards 2037878 and 2031594. *(Corresponding author: S. Farokh Atashzar.)*

S. Farokh Atashzar is with the Department of Electrical and Computer Engineering, and Department of Mechanical and Aerospace Engineering, New York University, New York, NY 11201 USA, and also with NYU Center for Urban Science and Progress and NYU WIRELESS Center, USA (e-mail: f.atashzar@nyu.edu).

Rory O'Keeffe, Seyed Yahya Shirazi, and Sarmad Mehrdad are with the Department of Electrical and Computer Engineering, New York University, USA.

Aaron M. Johnson is with the Department of Otolaryngology-Head and Neck Surgery, NYU Grossman School of Medicine, USA, and also with the Department of Rehabilitation Medicine, New York University, USA.

Tyler Crosby is with the Department of Otolaryngology-Head and Neck Surgery, NYU Grossman School of Medicine, USA.

Digital Object Identifier 10.1109/TBME.2022.3175948

I. Introduction

MORE than 17 million people in the United States are estimated to suffer from dysphonia (a voice disorder) each year [1], [2]. Excessive voice use and maladaptive compensatory muscle tension in response to underlying neurological or physiological laryngeal disease are considered to be potential roots for voice disorders such as muscle tension dysphonia (MTD) [3]. The laryngeal muscles are internal to the neck and require invasive access for direct examination. However, it is suggested that the perilaryngeal, cervical and cranial muscles may show activity alterations in subjects with dysphonia [4], which can help with the diagnosis using surface electromyography (sEMG).

The current methods for monitoring the function of laryngeal muscles include intramuscular EMG, external laryngeal palpation, and laryngeal endoscopy [5]–[11]. Although these methods have provided much information about muscle activation and function during voicing, they are invasive, uncomfortable, and subjective. Intramuscular EMG requires inserting small wires into the muscle using a needle to measure relative neuromuscular activity. Laryngeal endoscopy involves placing a flexible endoscope through the nose or a rigid endoscope through the mouth to visualize the gross anatomy and movements of the vocal folds. It requires a trained specialist to perform and subjectively interpret the findings. Manual palpation of the larynx and perilaryngeal musculature is easy to perform but does not provide any quantitative or standardized measure of muscle tension. Additionally, these evaluation methods can disturb the normal function of the muscle during the examination (for example, due to the pain), which can affect the accurate assessment.

Recording sEMG is a non-invasive technique that can potentially provide objective information about perilaryngeal muscle activity during voicing based on temporal and spectral characteristics of the muscle signal measured at the skin surface. It should be mentioned that discriminative differences have been suggested in the classic literature [4], [12] when comparing spectrotemporal features of sEMG for patients with MTD and healthy controls, yet recent studies with larger patient populations failed to show significant differences in the classic spectrotemporal sEMG metrics [9], [13]. As a potential reason for this inconsistency, Van Houtte *et al.* suggested that the earlier studies may have included patients with other comorbidities or with secondary illnesses affecting the results, and that could be why in recent studies, when controlling only for MTD, the discriminative power of classic features of sEMG dropped [13]. This calls for more advanced functional measures that can provide a holistic analysis of the distributed motor control on perilaryngeal muscles.

In neuroscience literature, coherence analysis has been used in the context of brain connectivity to detect how different regions of the brain are synchronized (or functionally coupled) during different tasks, and this measure has been used for detecting the degrees of several central nervous system conditions, such as Parkinson's Disease [14], [15]. More recently, using coherence analysis, the fluency of corticomuscular connectivity has also been investigated to understand how the central nervous system communicates with the peripheral nervous system [16], [17]. Similarly, the functional muscle network is an emerging concept that uses simultaneous multi-channel sEMG to decode how various muscle groups are synergistically synchronized during various motor tasks [18]–[20]. Intermuscular coherence networks have been recently used to holistically investigate the muscular system during various gait tasks and uniquely discriminate subtle differences in lower limb functions in non-disabled adults [21], [22].

To the best knowledge of the authors, the concept of functional muscle networks has not been used at the perilaryngeal and cranial levels. Some efforts have been conducted to assess beta-band (15-35 Hz) coherence between two anterior neck muscles during voicing, which showed some discriminative power to indicate hyperfunction and differences between control subjects and patients with vocal nodules [23], [24]. Expanding from a single coherence measurement in specific frequency bands to a wideband intermuscular coherence network increases the possibility for monitoring motor functions or impairments due to the wider spectral and spatial distribution of the analysis. Thus, it is imperative to understand the power of the perilaryngeal-cranial muscle network and the corresponding relationship with various vocal functions.

The purpose of this study is to quantify the perilaryngeal-cranial muscle network characteristics of a series of vocal tasks for healthy subjects. We hypothesize that in non-disabled subjects increasing the loudness and pitch (i.e., vocal frequency) will change the network connectivity in a manner that can be registered using topographical characteristics of the network, such as degree and clustering coefficient. In this study, to conduct a comparative analysis, the classical spectrotemporal features are also quantified to determine if the tasks with stronger muscle networks also consistently elicit statistically distinguishable spectrotemporal muscle activity. We show that the muscle network provides robust and statistically consistent discrimination for increasing loudness and pitch, suggesting that the perilaryngeal-cranial muscle network can indeed be used to detect subtle differences in vocal tasks, while the conventional spectrotemporal features fail to function accordingly.

II. Methods

Eight healthy subjects (four males, four females, 33.38 ± 9.32 years) participated in the study. The institutional review board of the New York University Grossman School of Medicine approved the study, and subjects provided their written consent after they received the study description. Subjects denied any history of dysphonia or neck and cervical-related injuries.

A. Experimental Procedure

Subjects performed a series of vocal tasks, each of a different type or while varying tonal parameters (Fig. 1(b)). The first group of tasks involved making a maximally sustained /a/ sound at a constant pitch and volume. With two levels of loudness and two levels of pitch, in total, there were four varied phonation (/a/ sound) tasks: 1) habitual loudness, habitual pitch, 2) elevated loudness, habitual pitch, 3) habitual loudness, high pitch and 4) elevated loudness, high pitch. Subjects were instructed to sustain the /a/ sound for as long as was comfortable. Subjects performed three trials of each of loudness and pitch combinations before moving to the next tasks. The second group of tasks included single repetition vocal exercises, namely (i) pitch glide, (ii) spontaneous speech, and (iii) singing. Pitch glide involved starting to intone at a low pitch and smoothly increasing to a final high pitch [25]. The spontaneous speech task involved responding to the prompt, "tell me how to make a peanut butter and jelly sandwich," in a typical conversational voice, while the singing task involved singing 'Happy Birthday' in a comfortable key chosen by the participant. The third group of tasks involved reading the first full paragraph of The Rainbow Passage [26], a standard reading passage used to evaluate the voice, at three levels of loudness: habitual, elevated, and whispering.

In this work, sEMG signals were recorded from twelve sensors, using the wireless Trigno sEMG system (Delsys Inc., Natick, MA), with a sampling frequency of 1259 Hz and an on-board 2nd-order high-pass filter at 20 Hz (Fig. 1). Four bipolar Trigno Mini sensors were used for the inner cervical muscles (inferior and superior infrahyoid, bilaterally), while eight bipolar Trigno Avanti sensors were used for Masseter, Superior Sternocleidomastoid, Inferior Sternocleidomastoid, and Trapezius. With regard to palpation, subjects were instructed to (i) clench their teeth to identify masseter, (ii) look left and right to identify lower and upper sternocleidomastoid, (iii) look up and down to identify the infrahyoid muscles, (iv) move shoulders forwards and backward before staying a neutral position to identify trapezius muscles. The skin surface was thoroughly wiped prior to sensor placement. Sensors were placed parallel to the direction of the muscles. In order to minimize the noise content of the recorded signals, subjects were instructed not to move their head during the task.

Following the recording, signals were pre-processed using MATLAB R2020b (MathWorks Inc. Natick MA). The first and last 1 s of all trials were clipped out, and other trials were clipped further in the case of a head movement at the beginning or at the

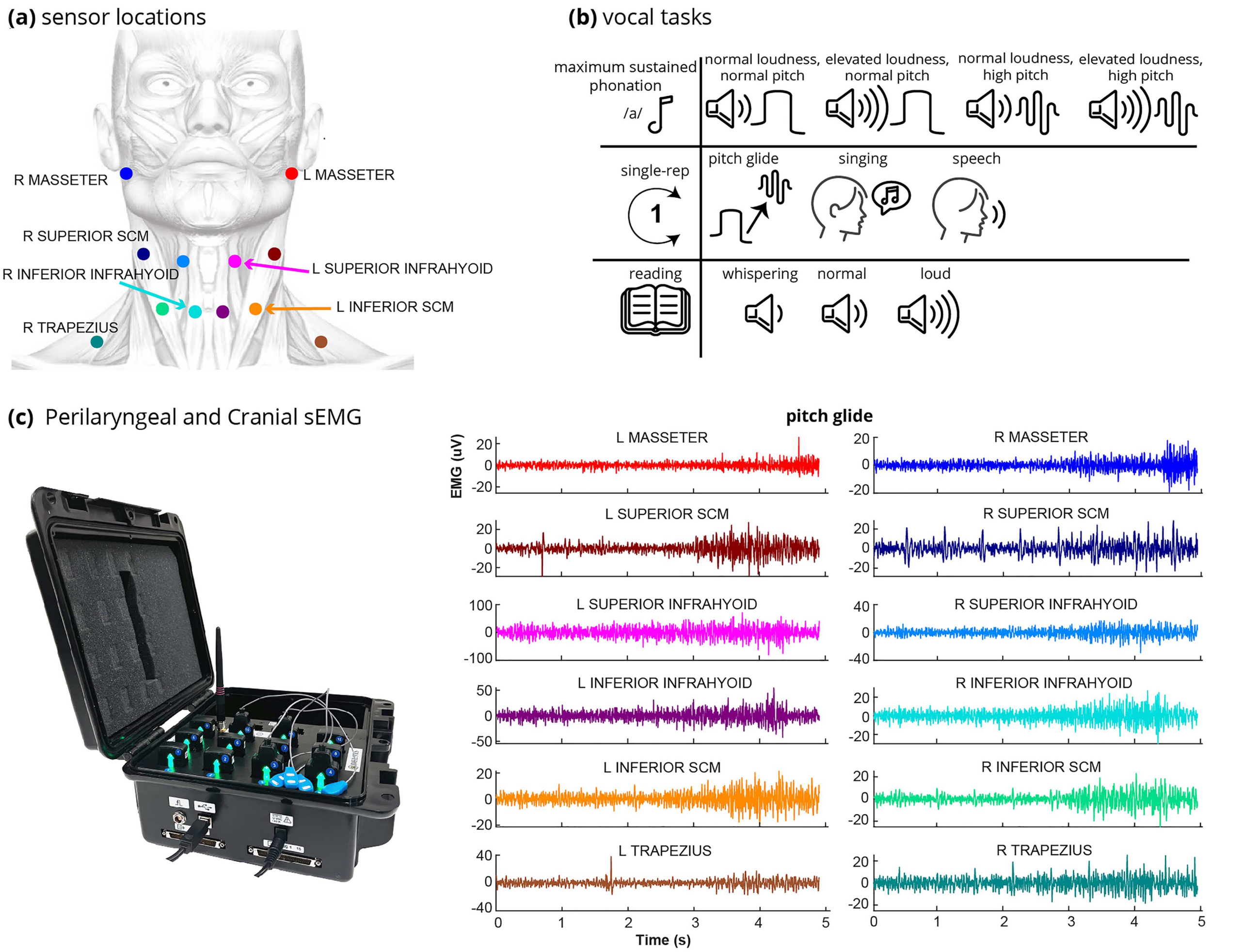


Fig. 1. (a) Six sensors were placed on each side of the neck at Masseter, Superior Sternoclediomastoid (Superior SCM), Superior Infrahyoid, Inferior Infrahyoid, Inferior Sternocleidomastoid (Inferior SCM) and Trapezius. (b) Subjects performed vocal tasks, classified as varied phonation, single repetition, or reading. Varied phonation tasks involved intoning /a/ at defined loudness and pitch levels. Single repetition tasks included pitch glide, singing, and speech. Reading tasks involved reading a passage at different loudness levels. (c) sEMG was recorded using the wireless Trigno system (Delsys Inc., Natick, MA) with eight Avanti and four Mini sensors (blue head). An exemplar recording from all perilaryngeal and cranial muscles is shown for the pitch glide task. Note how the amplitude level increases as the task develops for the infrahyoid and SCM muscles in particular.

end. Afterward, the signals were filtered with a high-pass filter at 20 Hz, a band-stop filter at 57.5–62.5 Hz for power-line noise, and a low-pass filter at 100 Hz. All filters were Butterworth 4th order zero-phase. For all of the analyses, we considered the 20–100 Hz range since the frequency bands of interest for intermuscular coherence networks generally include beta (14–30 Hz) and gamma (30–100 Hz) bands. Exemplar perilaryngeal and cranial muscle signals during pitch glide are shown in Fig. 1.

B. Muscle Signal Analysis

Muscle networks were constructed for all tasks, using coherence. Magnitude squared coherence, C_{xy} between two signals $x(t)$ and $y(t)$ is:

$$C_{xy} = \frac{|P_{xy}(f)|^2}{P_{xx}P_{yy}} \tag{1}$$

where P_{xx} and P_{yy} are the power spectral densities (PSDs) and P_{xy} is the cross power spectral density (CPSD). To compute the coherence, Welch's overlapped averaged periodogram method [27] was utilized with a Hamming window of 2048 samples (1.63 ms) and 50% overlap. The maximum coherence component in the was selected for each sensor pair. Using this maximum coherence value, muscle networks were constructed for each trial. Each node in the network represents a muscle, and the width of each line illustrates the pairwise muscle coherence. In the case of tasks that had multiple trials, the median network across trials was computed.

The degree of each node, D_i, is the average of all edges connected to the node. If the muscle network is represented by adjacency matrix A, D_i is defined as:

$$D_i = \left(\frac{1}{N-1}\right) \sum_{j=1, j\neq i}^{N} A_{ij}, \tag{2}$$

where N is the number of nodes. A node that has no connectivity with other nodes will have a degree, $D_i = 0$, while a node that is perfectly connected to other nodes will have $D_i = 1$.

A node's weighted clustering coefficient (WCC_i) gives the measure of how well that node is connected to its neighbors. The weighted clustering coefficient is defined as:

$$WCC_i = \frac{\sum_{j\neq i}, \sum_{k\neq i, j\neq k} A_{ij}A_{ik}A_{jk}}{\sum_{j\neq i}, \sum_{k\neq i, j\neq k} A_{ij}A_{ik}} \tag{3}$$

A node that is not connected to its neighbors will have a weighted clustering coefficient, $WCC_i = 0$, while a node that is very well connected to its neighbors will have $WCC_i = 1$.

Global efficiency is directly proportional to how well the network is connected overall. The efficiency (E) of a network is defined as:

$$E = \frac{1}{N\left(N-1\right)} \sum_{i\neq j} \frac{1}{L_{ij}}. \tag{4}$$

where L_{ij} is the shortest path between nodes i and j [28]. The efficiency (E) is normalized by the ideal efficiency (E_{id}) to give the global efficiency, GE:

$$GE = \frac{E}{E_{id}}, \tag{5}$$

which is bounded between 0 and 1. A network with perfect connectivity will have $GE = 1$, while one with no connectivity will have $GE = 0$.

In order to provide a comparison between the muscle coherence network and the conventional spectrotemporal metrics, muscle activations were quantified in the time and frequency domains. The time-domain activation was quantified by finding the root mean square (RMS) value across the trial duration. With regard to the spectral domain, PSD was computed using Welch's method [27] and the median PSD across 20–100 Hz was computed. Furthermore, the median frequency was computed for each task. The median frequency is defined as the frequency at which the area under the PSD graph is divided in two. The median value across trials was used for PSD, median frequency, and RMS when there were multiple trials.

C. Statistical Analysis

In order to evaluate the statistical trends observed in coherence muscle networks, a coherence distribution was constructed for each task (as later will be discussed in Fig. 5). Each distribution consisted of degree and weighted clustering coefficient for all nodes across all subjects' muscle networks, giving $n = \#subjects \times \#nodes = 8 \times 12 = 96$. Similarly, distributions were constructed for RMS, PSD, and median frequency ($n = 96$ for all). The Kolmogorov-Smirnov test for normality rejected the normal distribution hypothesis for the coherence, RMS, PSD, and median frequency distributions. Therefore, nonparametric statistical tests were used in our analysis. The Friedman test was used to compare tasks in each group (i.e., varied phonation, single-repetition, reading). The Wilcoxon signed-rank test was used as a posthoc test if the Friedman test revealed significance. The significance level, α, for all tests was initially set at 0.05. To adjust for multiple comparisons, the Bonferroni correction was applied, dividing α by the number of comparisons.

Finally, by using the rank-biserial correlation, the effect size of the non-normal distributions was quantified [29]. In this regard, $|r_{rb}|$ was used for measuring the rank-biserial correlation. A higher value means that the effect size is larger. For example, as can be seen in Fig. 5(b), the coherence degree for single repetition tasks has a very high effect size ($|r_{rb}| = 0.83$), and the difference between tasks is even visually clear. On the other hand, the coherence degree for reading tasks has a low effect size ($|r_{rb}| = 0.1$), as there is not a clear relationship between coherence and reading task loudness.

III. Results

A. Coherence Networks

The median network across subjects displays a visible difference between vocal tasks, e.g., between pitch glide and speech (Fig. 2). Similarly, Fig. 3 shows that the mean degree of the network changed by the task for all subjects. Interestingly, the mean degree showed a monotonic increasing trend in response to both raised loudness and pitch for the varied phonation tasks, and pitch glide appears to have the highest coherence of all 10 tasks (Fig. 3). Mean degree showed a monotonically decreasing trend from pitch glide to singing to speech. Looking at the adjacency matrices corresponding to the intermuscular coherence networks confirms this observation and shows that there appears to be little difference in the observed network for reading tasks with different loudnesses (Fig. 4).

In order to support the initial observations of the coherence network differences between the vocal tasks, coherence distributions were constructed by including all nodes in the subjects' intermuscular network, measured using degree and weighted clustering coefficient (Fig. 5). For the varied phonation tasks, the task-wise network degree and weighted clustering coefficient median were monotonically ascending with increasing pitch and loudness, and **all** tasks were statistically different from each other (*Friedman* $Chi^2_{(3,285)} > 136.62$, $p < 0.05$, *posthoc Wilcoxon signed-rank test:* **all six** *pairwise comparisons* $p < 0.001$). The network's global efficiency showed a trend of monotonically increasing coherence with pitch and loudness for 5 out of 8 subjects. Moreover, the effect size of the network metrics indicated quite a high value (degree: $|r_{rb}| = 0.48$, weighted clustering coefficient: $|r_{rb}| = 0.5$). For the single repetition tasks, the median of network degree and weighted clustering coefficient was decreasing monotonically, with the pitch-glide having the highest network degree and weighted clustering coefficient at both greater than 0.7 (*Friedman* $Chi^2_{(2,190)} > 186.18$, $p < 0.05$, *post-hoc Wilcoxon signed-rank test for* **all three** *pairwise comparisons,* $p < 0.001$). Furthermore, the global efficiency trend was consistent and decreasing across all eight subjects. The rank biserial correlation of the network metrics also indicated a very high effect size (degree: $|r_{rb}| = 0.83$, weighted clustering coefficient: $|r_{rb}| = 0.85$). The degree and weighted clustering coefficient of habitual were higher than whispered reading ((*Friedman* $Chi^2_{(2,190)} > 4.77$, $p < 0.05$, *post-hoc Wilcoxon signed-rank test:* $p < 0.008$) while the weighted

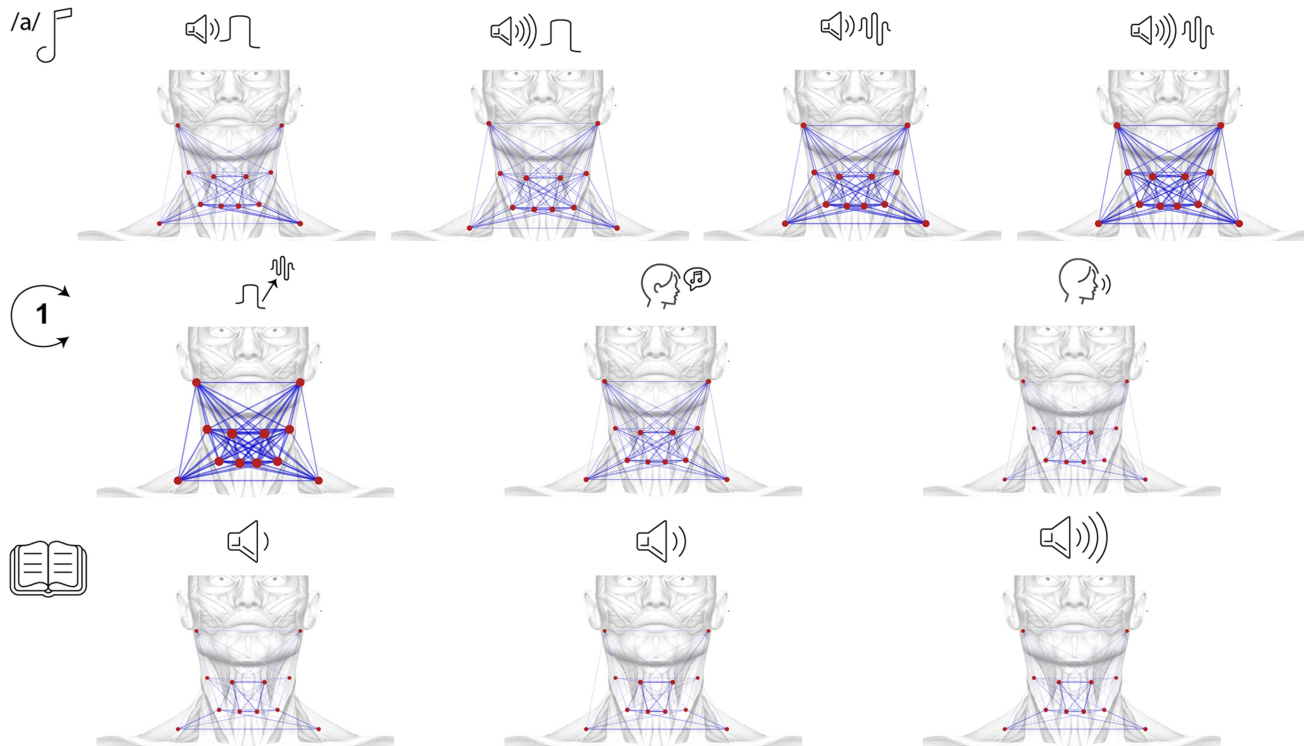


Fig. 2. Coherence muscle networks for each vocal task were created from twelve sEMG sensors placed bilaterally on the neck area. The width of each line denotes the pairwise coherence between the two connected muscles, which is equal to the maximum coherence component in the 20–100 Hz range. In the case of tasks that had multiple trials, the median network across trials is shown. Each node radius is equal to the degree (mean of coherences involving that node). The line widths and node radii seem largest for /a/ with elevated loudness, high pitch, and pitch glide tasks.

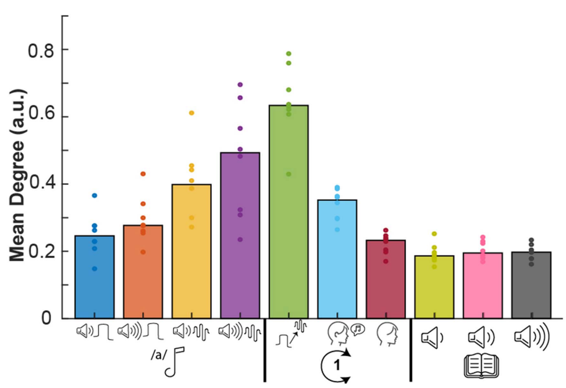


Fig. 3. Network mean degree for each task. The bar depicts the median value across subjects. Individual subject values are denoted by dots. For the subject median, the mean degree shows an increasing trend starting from the first task (habitual loudness, habitual pitch) and continuing incrementally until pitch glide. Network connectivity then shows a decreasing trend from pitch glide to singing to speech. Finally, reading tasks seem to have little difference between each other.

clustering coefficient of habitual was higher than loud reading ($p = 0.001$). However, other task pairs failed to show significant differences for the post-hoc test. Global efficiency did not indicate a consistent trend amongst subjects for the reading tasks. Moreover, for the reading task set, the effect size was low for the network metrics (degree: $|r_{rb}| = 0.08$, weighted clustering coefficient: $|r_{rb}| = 0.1$).

Since one of the secondary aims of this study is to suggest a suitable candidate task(s) for monitoring the effect of therapy, the tasks which had produced the highest network metrics were identified. For this, three tasks were selected, which had resulted in the highest response in Fig. 5, when compared within their categories. The selected three tasks are (i) the varied phonation task with elevated loudness and high pitch, (ii) pitch glide, and (iii) reading at habitual loudness, and results are given in Fig. 6. Both network degree and weighted clustering coefficient were significantly different from each other for **all** comparisons (*Friedman* $Chi^2_{(2,190)} > 149.08$, $p < 0.05$, *post-hoc Wilcoxon signed-rank test:* **all three** *pairwise comparisons* $p < 0.001$). Moreover, pitch glide had the highest network degree (median value ~ 0.7) and weighted clustering coefficient (median value ~ 0.75), even higher than the varied phonation task with elevated loudness and high pitch (degree: median value ~ 0.55, weighted clustering coefficient: ~ 0.6). Reading at habitual loudness had a lower degree (median value ~ 0.21) and weighted clustering coefficient (median value ~ 0.22) than the other two tasks. This suggests that pitch glide generates the maximum response of the network, which can be considered potentially the most responsive and suitable task for identifying abnormalities.

B. Spectrotemporal Metrics

To compare the ability of spectrotemporal metrics to distinguish different tasks, statistical analyses on RMS, PSD, and

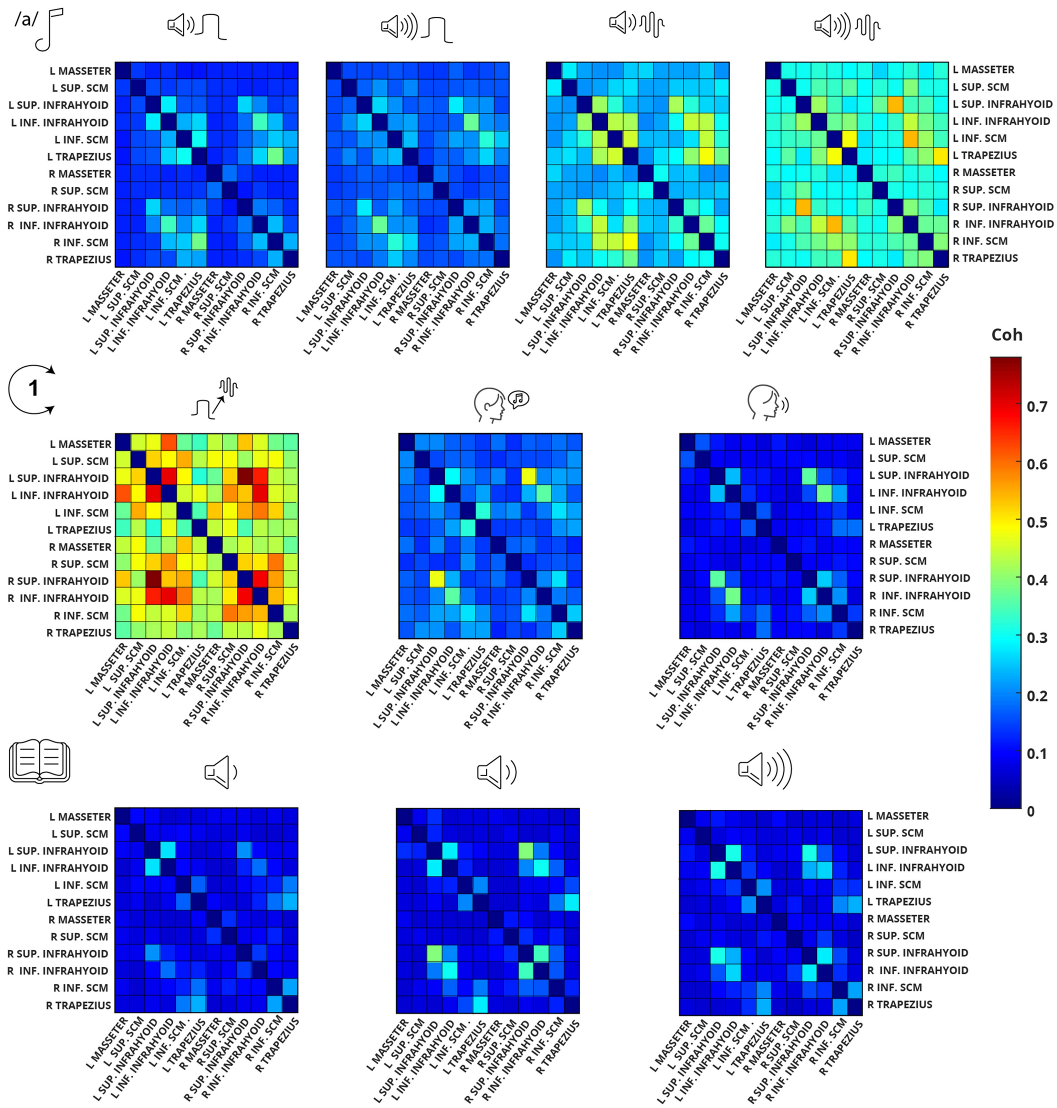


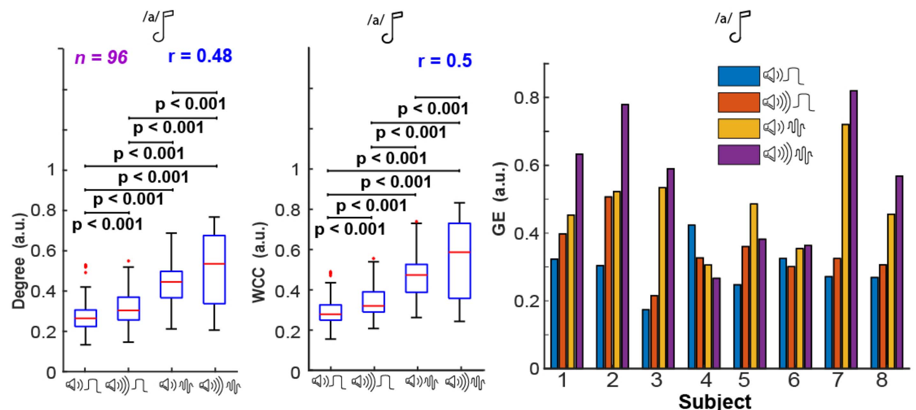
Fig. 4. The median intermuscular coherence network adjacency matrix (12 rows $\times$ 12 columns $=$ 144 squares) across subjects was computed for each vocal task. In the case of tasks that had multiple trials, the median adjacency matrix across trials was first computed. Each square represents the pairwise maximum coherence between the two muscles from the corresponding row and column and the color is proportional to the pairwise coherence. Since the connectivity is undirected, the adjacency matrix is symmetrical. The highest pairwise coherences are observed for the elevated loudness, high pitch and pitch glide tasks.

median frequency of sEMG were conducted. Distributions for each of the three aforementioned quantities were constructed by considering all nodes across all subjects (n = *#subjects x #nodes* = *96*). The mean value across each distribution was computed for network and spectrotemporal metrics to form the comparative Table I. The level of cervical muscle activation as measured by RMS was low by sEMG standards, with a range of $4 - 11\mu$V for the task mean values (Fig. 7(a), Table I).

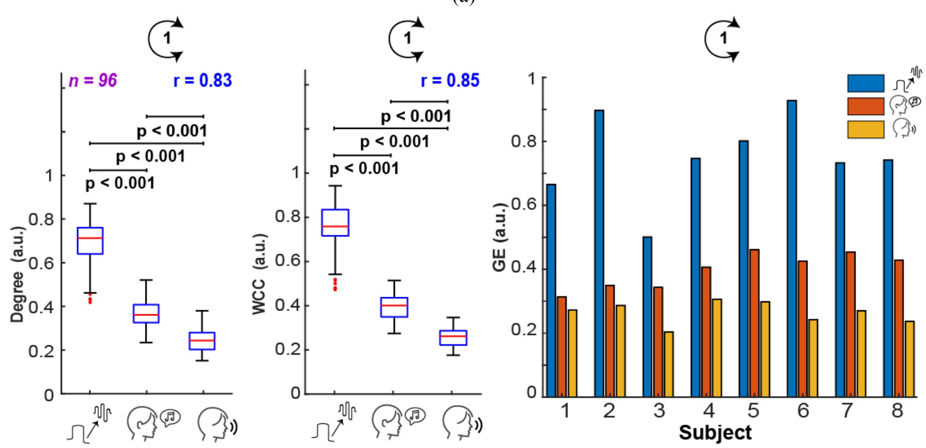
For the varied phonation tasks, the median PSD of louder tasks was higher than habitual loudness tasks ($p < 0.008$). However, the overall effect size of varied phonation tasks (RMS: $|r_{rb}| = 0.08$, PSD: $|r_{rb}| = 0.07$, median frequency: $|r_{rb}| = 0.06$) was quite small. For single repetition tasks, RMS and PSD did not show a clear trend (Fig. 5(b)). However, median frequency of pitch glide was higher than other tasks ($p < 0.001$), $|r_{rb}| = 0.12$. PSD and RMS showed increased values for loud reading (Fig. 5(c)). All distributions were different from each other (RMS: $p < 0.015$, PSD: $p < 0.01$) but the rank biserial correlation, $|r_{rb}| = 0.12$, suggests a weak effect size.

IV. Discussion

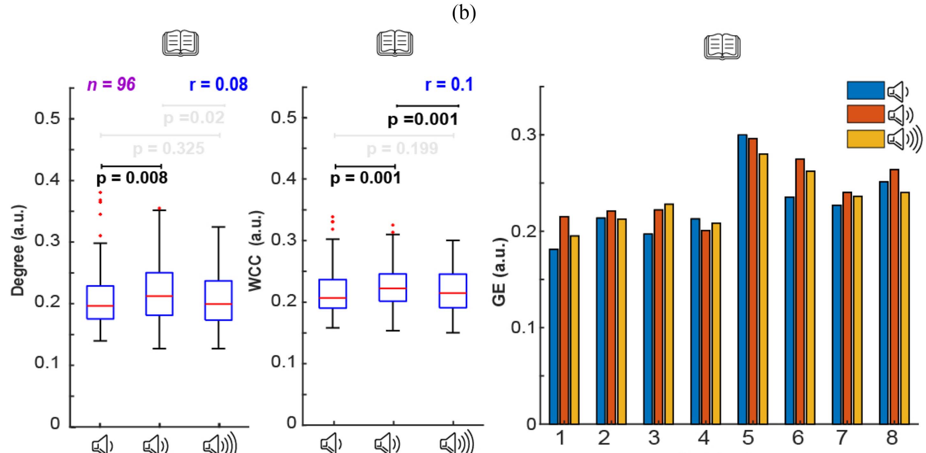
The results show that the perilaryngeal-cranial intermuscular coherence network can distinguish both changes in vocal parameters (i.e., loudness and pitch) and different vocal tasks. The muscle network quantifies the spectral synchrony and can robustly capture small changes associated with vocal output. The strong performance of the muscle network is demonstrated initially by the network visualization, and then statistical analysis using the network metrics (degree and weighted clustering coefficient) confirmed the observed trends. With regard to the vocal tasks, two important statistically robust trends are identified as follows:



(a)



(b)



(c)

Fig. 5. Statistical results for network metrics of the three groups of tasks. The coherence muscle network was constructed for each task, and produces 12 node values for each of degree and weighted clustering coefficient (WCC). For each distribution of degree and weighted clustering coefficient, all node values for all subjects are included (12 nodes $\times$ 8 subjects $=$ 96 data points). For global efficiency, a singular value was obtained for each subject's muscle network. (a) For **varied phonation** tasks, intermuscular network degree and weighted clustering coefficient increase monotonically with raised loudness and pitch, with statistical significance (*all six taskwise comparisons:* p $<$ 0.001). The adjusted significance level was $\alpha' = 0.05/6 = 0.0833$. The effect size ($|r_{rb}|$) is quite high for both degree ($|r_{rb}| = 0.48$) and weighted clustering coefficient ($|r_{rb}| = 0.5$). (b) For **single repetition** tasks, intermuscular network degree and weighted clustering coefficient decrease monotonically from pitch glide, to singing, to speech, with statistical signficance (*all three taskwise comparisons:* p $<$ 0.001). The adjusted significance level was $\alpha' = 0.05/3 = 0.0167$. The effect size ($|r_{rb}|$) is very high for both degree ($|r_{rb}| = 0.83$) and weighted clustering coefficient ($|r_{rb}| = 0.85$). All eight subjects' global efficiency bar plots follow the monotonically decreasing pattern from pitch glide to singing to speech. (c) For **reading** tasks, there are no visible trends in degree, weighted clustering coefficient or subject-wise global efficiency. The degree and weighted clustering coefficient were higher for habitual vs loud reading ($p < 0.009$). The adjusted significance level was $\alpha' = 0.05/3 = 0.0167$. For weighted clustering coefficient, there is a significant difference between reading at habitual and elevated loudness ($p = 0.001$). Effect size for degree ($|r_{rb}| = 0.08$) and weighted clustering coefficient ($|r_{rb}| = 0.1$) is low.

TABLE I
MEAN VALUES OF NETWORK (LEFT) AND SPECTOTEMPORAL (RIGHT) METRICS FOR EACH TASK

Task	*Metric* Degree (a.u.)	WCC (a.u.)	GE (a.u.)	*Metric* RMS (uV)	PSD (dB)	MDF (Hz)
varied phonation						
$\rightarrow$ loudness, $\rightarrow$ pitch	0.27	0.28	0.29	3.97	-128.61	51.48
$\uparrow$ loudness, $\rightarrow$ pitch	0.32	0.34	0.35	4.64	-127.57	52.57
$\rightarrow$ loudness, $\uparrow$ pitch	0.44	0.47	0.48	3.97	-128.72	50.58
$\uparrow$ loudness, $\uparrow$ pitch	0.51	0.55	0.56	4.66	-127.59	51.22
single repetition						
pitch glide	0.70	0.76	0.76	7.52	-125.10	55.45
singing	0.36	0.39	0.40	6.17	-124.75	51.80
speech	0.24	0.25	0.27	7.26	-123.66	52.02
reading						
whisper	0.21	0.22	0.23	6.66	-124.79	51.85
normal	0.22	0.23	0.24	7.68	-123.96	52.67
loud	0.21	0.22	0.23	11.10	-121.61	53.81

$\rightarrow$ loudness = habitual loudness, $\uparrow$ loudness = elevated loudness, $\rightarrow$ pitch = habitual pitch, $\uparrow$ pitch = high pitch

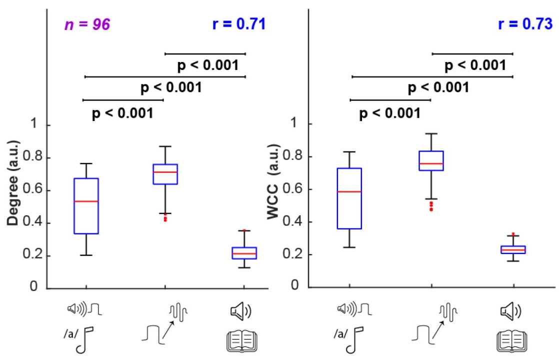


Fig. 6. Comparison of tasks with highest network metrics from each group: 1) Pitch glide has both the highest degree and weighted clustering coefficient (WCC), followed by 2) varied phonation task with elevated loudness, high pitch, followed by 3) habitual reading, and all tasks are different from one another (*for both degree and weighted clustering coefficient, all three taskwise comparisons:* $p < 0.001$). The adjusted significance level was $\alpha' = 0.05/3 = 0.0167$. Reading has $\sim 1/3$ the median degree and weighted clustering coefficient of the varied phonation task and pitch glide.

(i) network degree and weighted clustering coefficient increase monotonically with loudness and pitch in the varied phonation tasks and (ii) network degree and weighted clustering coefficient are both the highest for the pitch glide task. The ability of the intermuscular coherence network to distinguish vocal parameters and tasks was far superior to the conventional node-wise metrics for sEMG, such as RMS and PSD. These results suggest that both the varied phonation and single repetition tasks in combination with the perilaryngeal-cranial myographic network are sensitive to various vocal features and thus can be potentially considered as candidates for measuring the efficacy of therapy for vocal disorders. The current study showed very strong statistics supporting the use of the proposed network measures while the conventional spectrotemporal metrics fail to provide the needed sensitivity to the vocal features.

Perilaryngeal-cranial intermuscular coherence ascends monotonically with loudness and pitch in the varied phonation tasks (Figs. 2—5(a), Table I). The high differentiation of the varied phonation and single repetition group tasks with high to very high effect size (varied phonation: $|r_{rb}| \sim 0.5$, single repetition: $|r_{rb}| \sim 0.8$) robustly supports the hypothesis that perilaryngeal-cranial intermuscular coherence is generally proportional to loudness and pitch. To the best of the authors' knowledge, this is the first study that shows perilaryngeal-cranial intermuscular coherence is conclusively correlated to loudness and pitch.

In contrast to the coherence network, conventional node-wise metrics such as RMS, PSD, and median frequency failed to efficiently discriminate varied phonation tasks (Fig. 5(a)). Although there were some differences between louder and habitual loudness tasks, the effect size ($|r_{rb}|$) was much smaller for these metrics than for coherence (RMS: 0.08, PSD: 0.07, versus network degree: 0.48 and weighted clustering coefficient: 0.5), i.e., the average difference between the task pairs was less pronounced compared to the median of the tasks. Despite inconclusive shifts in low cervical muscle activations (Table I, Fig. 7), significant relative changes from task to task were captured by network metrics (Fig. 5(a), (b)). The superior performance of coherence over node-wise metrics may arise from the fact that network analysis allows us to conduct a holistic neurophysiological analysis of the functional synchrony and synergistic co-modulation of the muscles needed for successful conduction of the tasks. Thus, in the context of voice, the authors believe that the synchronous behavior of the muscles has higher discriminative power (for separating vocal features) and potentially higher diagnostic value than isolated individual muscle recordings.

Overall, the single repetition tasks provided the most differentiable network degree and weighted clustering coefficient, which both decrease monotonically from pitch glide to singing to speech with very high effect size (degree: $|r_{rb}| = 0.83$, weighted clustering coefficient: $|r_{rb}| = 0.85$) (Fig. 5). The effect is even greater for the single repetition than varied phonation tasks. Moreover, it should be highlighted that all subjects followed the group trend for network global efficiency of the single repetition tasks, emphasizing robust separability. This suggests that vocal tasks of different nature (pitch glide, singing, and speech) provide the greatest diversity and objectivity of muscle network performance which can be easily differentiated by the

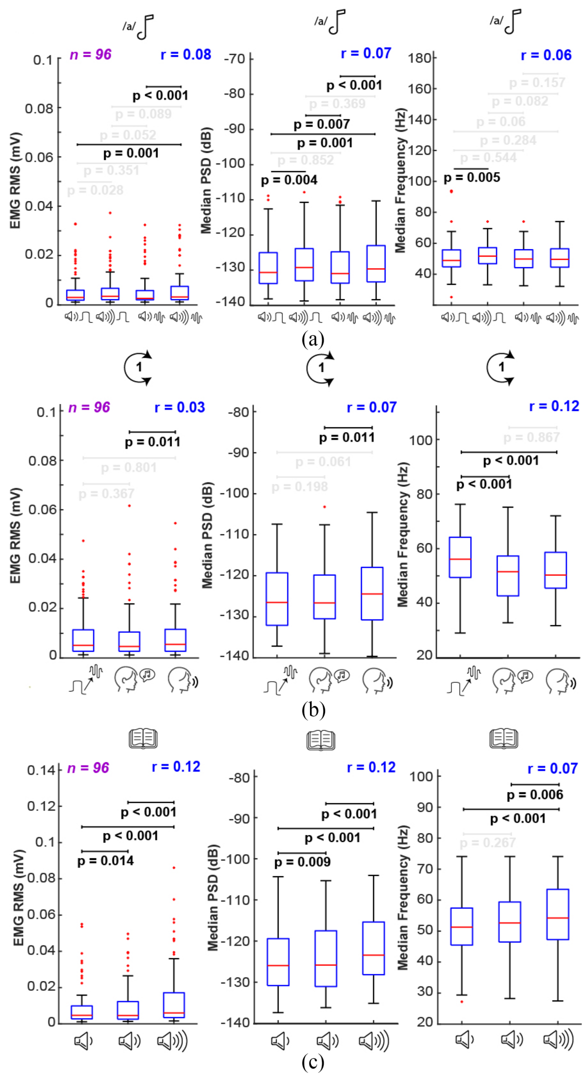


Fig. 7. **RMS, PSD** and **median frequency** of the three groups of tasks. Quantifying the muscle activity with RMS, median PSD, and median frequency produces 12 node values for each metric. All node values for all subjects are included in each distribution (12 nodes $\times$ 8 subjects $=$ 96 data points). (a) For **varied phonation** tasks, median PSD shows increases in response to raised loudness. Louder tasks have higher median PSD ($p < 0.008$) than habitual loudness tasks. RMS increases with elevated loudness for the high pitch task ($p < 0.002$). The adjusted significance level was $\alpha' = 0.05/6 = 0.0833$. However, overall effect size of varied phonation tasks is still low (RMS: $|r_{rb}| = 0.08$, PSD: $|r_{rb}| = 0.07$). No consistent patterns were observed for median frequency. (b) For **single repetition** tasks, pitch glide had a significantly higher median frequency than the two other tasks ($p < 0.001$). The adjusted significance level was $\alpha' = 0.05/3 = 0.0167$. No substantial trends were observed for RMS or PSD. (c) For **reading** tasks, RMS and median PSD show increases in response to elevated reading loudness. The adjusted significance level was $\alpha' = 0.05/3 = 0.0167$. However, the effect size is weak (RMS and PSD: $|r_{rb}| = 0.12$).

network metrics. The fact that pitch glide has the highest degree and weighted clustering coefficient among all tasks (Fig. 6) suggests that the smooth transition of the voice through octaves is assisted by very synchronous perilaryngeal-cranial muscle activity. Indeed, it is notable that singing had higher network coherence than regular speech. Using the results that (i) higher pitch led to increased network coherence for /a/ tasks and (ii) pitch glide has the maximum network coherence of all tasks, both the higher average pitch and the larger number of pitch changes for singing versus speech are consistent with singing having higher network coherence. This result is in contrast to a previous result with beta-band coherence between two muscles [23], which found that speech had higher beta-band coherence than singing. This difference might be due to benefiting from an intermuscular coherence network with 12 nodes and 66 edges in this study versus only two nodes and one edge coherence in the previous study. This work also considers a much wider frequency range (20–100 Hz) than the previous beta-band coherence [23]. Despite the higher expected pitch for singing vs. speech, neither median frequency nor PSD succeeded in detecting a difference, highlighting the superiority of muscle network coherence over node-wise methods in responding to quantifiably small physiological changes of external muscles related to vocal output.

Tasks with a higher pitch component produced a more synchronous network, and varying the pitch led to clearer network responses. Both the varied phonation (/a/) high pitch and pitch glide tasks were shown to have at least 3 $\times$ the network degree or weighted clustering coefficient of a reading task (Fig. 6), highlighting the ability of tasks with a high pitch component to produce the most pronounced network output. Moreover, the pitch-varying task sets showed clear responses to vocal parameter changes. The varied phonation tasks showed a gradually increasing response (Fig. 5(a)) from a median network degree ~ 0.25 for habitual loudness, habitual pitch to ~ 0.55 for elevated loudness, high pitch. The single repetition tasks showed a sharper decrease from pitch glide to singing than from singing to speech. For both varied phonation and single repetition tasks, the network metrics showed a distinctive response to each task; the effect size was large (weighted clustering coefficient: $|r_{rb}| > 0.5$) and the null hypothesis was rejected with the highest significance level ($p < 0.001$) for all task-wise comparisons. Tasks that produce the most responsive network behavior would be the most appropriate candidates in vocal therapy assessment, as the responsiveness of the perilaryngeal-cranial intermuscular coherence network should be high to maximize the sensitivity of detecting signs of dysphonia or improvements made by the therapy. Since the muscle network was most responsive to tasks with a higher pitch component and pitch-varying task sets, such tasks promise the best chance of success when monitoring patient progress during vocal therapy.

Our results demonstrate the efficacy and sensitivity of the intermuscular coherence network analysis in reflecting quantifiably small modulations in vocal output (Table I), such as detecting changes in vocal parameters and discriminating single repetition tasks, with a robustly high effect size.

Taking inspiration from brain connectivity networks that can detect functional changes, this is the first work that shows topographical features of the intermuscular coherence network can detect changes to indicate functional vocal characteristics. Since vocal impairments and aging also create significant changes in vocal parameters [30], [31], the outcomes of this research can be potentially translated into characterizing vocal disorders. This will be investigated in our future work. Our study showed that the high functional synchronicity of the perilaryngeal-cranial muscles produced a strong network response during pitch glide, suggesting that the laryngeal performance can be measured by the perilaryngeal-cranial network, which needs to function in synchrony to conduct the corresponding tasks. With such a high sensitivity, other vocal disorders could be potentially detected and monitored by our suggested configuration. Given the high range of muscles recorded, in addition to the wide frequency range covered, the perilaryngeal-cranial muscle network is a great candidate for an objective, digital method of detecting and monitoring a wide range of vocal disorders, using smart wearable clinical technologies, such as a smart EMG necklace. A further clinical application of the strong correlation between perilaryngeal-cranial muscle network features and vocal output lies in an EMG-based electrolarynx device [32], [33]. Regarding patients who have lost their vocal cords due to cancer, the muscle network features of healthy cervical muscle activity could be used to drive a laryngeal prosthesis, capable of distinct levels of output pitch and loudness.

Some limitations of this study include that the effect of vocal fatigue was not controlled and the order of the tasks was not randomized. It should be noted that the vocal experiment procedure for each subject required a total of $\sim$ 9 minutes of vocal effort, which is significantly lower than the effort needed for fatiguing a subject (which is $\sim$ 60 minutes for vocal fatigue with comfortable reading loudness or intermittent loud reading tasks for several hours [34], [35]). Even though the duration of each task was very limited in this study and would not trigger vocal fatigue or any major change in the neurophysiological status of the muscles, this study does not randomize the tasks to maximize protocol adherence, in terms of production of the targeted vocal tasks. The authors would like to highlight that this study was limited in terms of inclusion of older adults and patients with voice disabilities and degradation. These topics will form our future work. It should be noted that cross-talk can always affect sEMG recordings with densely-placed electrodes, including the one used in this paper. However, the robust effect size of the features of the intermuscular coherence network reported in this paper, the use of Delsys Mini sensors, and bipolar sEMG suggest that cross-talk has had minimum-to-no effects on the main aim of this study. The resiliency of the network to potential cross-talk can be specifically investigated in our future work.

V. Conclusion

For the first time, this work shows that the perilaryngeal-cranial functional muscle network can detect subtle changes in distinguishing vocal tasks. The network-based metrics showed a robust effect size for changes in loudness and pitch in the set of varied phonation tasks and an even more robust effect size amongst the single repetition tasks, which included a pitch glide, singing, and a short speech. The network outperformed conventional spectrotemporal node-wise metrics (RMS, PSD, and median frequency) regarding sensitivity to changes in vocal output. This robustness suggests that the perilaryngeal-cranial functional muscle network is a promising method to differentiate physiological abnormalities and may be used in the future to assess voice disorders and optimize vocal rehabilitation. The future direction of this study includes data collection from a more diverse control population and inclusion of the patient population to (a) enhance the robustness of the made observations and (b) evaluate the power of the proposed method as a potential diagnostic biomarker.

References

[1] N. Roy *et al.*, "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, Nov. 2005.

[2] N. Bhattacharyya, "The prevalence of voice problems among adults in the United States," *Laryngoscope*, vol. 124, no. 10, pp. 2359–2362, Oct. 2014.

[3] E. Van Houtte, K. Van Lierde, and S. Claeys, "Pathophysiology and treatment of muscle tension dysphonia: A review of the current knowledge," *J. Voice*, vol. 25, no. 2, pp. 202–207, Mar. 2011.

[4] I. Hocevar-Boltezar, M. Janko, and M. Zargi, "Role of surface EMG in diagnostics and treatment of muscle tension dysphonia," *Acta Otolaryngol.*, vol. 118, no. 5, pp. 739–743, Sep. 1998.

[5] N. Roy *et al.*, "Evidence-based clinical voice assessment: A systematic review," *Amer. J. Speech- Lang. Pathol.*, 2013.

[6] M. L. Andreassen, J. K. Litts, and D. R. Randall, "Emerging techniques in assessment and treatment of muscle tension dysphonia," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 25, no. 6, pp. 447–452, Dec. 2017.

[7] N. Jafari *et al.*, "A novel laryngeal palpatory scale (LPS) in patients with muscle tension dysphonia," *J. Voice*, vol. 34, no. 3, May 2020, Art, no. 488.

[8] S. M. Khoddami, N. N. Ansari, and S. Jalaie, "Review on laryngeal palpation methods in muscle tension dysphonia: Validity and reliability issues," *J. Voice*, vol. 29, no. 4, pp. 459–468, Jul. 2015.

[9] S. M. Khoddami *et al.*, "Validity and reliability of surface electromyography in the assessment of primary muscle tension dysphonia," *J. Voice*, vol. 31, no. 3, May 2017, Art. no. 386.

[10] R. T. Sataloff *et al.*, "Practice parameter: Laryngeal electromyography (an evidence-based review)," *Otolaryngol. Head Neck Surg.*, vol. 130, no. 6, pp. 770–779, Jun. 2004.

[11] C. E. Stepp *et al.*, "Neck surface electromyography as a measure of vocal hyperfunction before and after injection laryngoplasty," *Ann. Otology Rhinology Laryngology*, vol. 119, no. 9, pp. 594–601, Sep. 2010.

[12] M. A. Redenbaugh and A. R. Reich, "Surface EMG and related measures in normal and vocally hyperfunctional speakers," *J. Speech Hear. Disord.*, vol. 54, no. 1, pp. 68–73, Feb. 1989.

[13] E. Van Houtte *et al.*, "An examination of surface EMG for the assessment of muscle tension dysphonia," *J. Voice*, vol. 27, no. 2, pp. 177–186, Mar. 2013.

[14] J. Sarnthein and D. Jeanmonod, "High thalamocortical theta coherence in patients with Parkinson's disease," *J. Neurosci.*, vol. 27, no. 1, pp. 124–131, 2007.

[15] S. W. Tung *et al.*, "Motor imagery BCI for upper limb stroke rehabilitation: An evaluation of the EEG recordings using coherence analysis," in *Proc. IEEE 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 261–264.

[16] H. E. Rossiter *et al.*, "Changes in the location of cortico-muscular coherence following stroke," *NeuroImage: Clin.*, vol. 2, pp. 50–55, 2013.

[17] T. Mima and M. Hallett, "Corticomuscular coherence: A review," *J. Clin. Neuriophysiol.*, vol. 16, no. 6, pp. 501–511, 1999.

[18] T. W. Boonstra *et al.*, "Muscle networks: Connectivity analysis of EMG activity during postural control," *Sci. Rep.*, vol. 5, no. 1, pp. 1–14, 2015.

[19] T. W. Boonstra *et al.*, "Information decomposition of multichannel EMG to map functional interactions in the distributed motor system," *Neuroimage*, vol. 202, Nov. 2019, Art. no. 116093.

[20] R. O'Keeffe *et al.*, "InfoMuNet: Information-theory-based functional muscle network tracks sensorimotor integration post-stroke," *bioRxiv*, Feb. 2022, doi: 10.1101/2022.02.10.479324.

[21] J. N. Kerkman *et al.*, "Network structure of the human musculoskeletal system shapes neural interactions on multiple time scales," *Sci. Adv.*, vol. 4, no. 6, Jun. 2018, Art. no. eaat0497.

[22] J. N. Kerkman *et al.*, "Muscle synergies and coherence networks reflect different modes of coordination during walking," *Front. Physiol.*, vol. 11, Jul. 2020, Art. no. 751.

[23] C. E. Stepp, R. E. Hillman, and J. T. Heaton, "Modulation of neck intermuscular beta coherence during voice and speech production," *J. Speech Lang. Hear. Res.*, vol. 54, no. 3, pp. 836–844, Jun. 2011.

[24] C. E. Stepp, R. E. Hillman, and J. T. Heaton, "Use of neck strap muscle intermuscular coherence as an indicator of vocal hyperfunction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 3, pp. 329–335, Jun. 2010.

[25] R. L. Milbrath and N. P. Solomon, "Do vocal warm-up exercises alleviate vocal fatigue?," *J. Speech Lang. Hear. Res.*, vol. 46, no. 2, pp. 422–436, Apr. 2003.
[26] G. Fairbanks, *Voice and Articulation Drillbook*. Reading, MA, USA: Addison-Wesley Educational Publishers, 1960.
[27] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
[28] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010.
[29] E. E. Cureton, "Rank-biserial correlation," *Psychometrika*, vol. 21, no. 3, pp. 287–290, Sep. 1956.
[30] R. Shrivastav, D. A. Eddins, and S. Anand, "Pitch strength of normal and dysphonic voices," *J. Acoust. Soc. Amer.*, vol. 131, no. 3, pp. 2261–2269, Mar. 2012.
[31] M. Brückl, "Women's vocal aging: A longitudinal approach," in *Proc. Interspeech*, 2007, pp. 1170–1173, doi: 10.21437/Interspeech.2007-379.
[32] K. Oe, "An electrolarynx control method using myoelectric signals from the neck," *J. Robot. Mechatronics*, vol. 33, no. 4, pp. 804–813, 2009, doi: 10.20965/jrm.2021.p0804.
[33] H. L. Kubert *et al.*, "Electromyographic control of a hands-free electrolarynx using neck strap muscles," *J. Commun. Disord.*, vol. 42, no. 3, pp. 211–225, 2009.
[34] N. P. Solomon, "Vocal fatigue and its relation to vocal hyperfunction," *Int. J. Speech Lang. Pathol.*, vol. 10, no. 4, pp. 254–266, Jan. 2008.
[35] N. V. Welham and M. A. Maclagan, "Vocal fatigue: Current knowledge and future directions," *J. Voice*, vol. 17, no. 1, pp. 21–30, Mar. 2003.