Comparison of Neck Tension Palpation Rating Systems With Surface Electromyographic and Acoustic Measures in Vocal Hyperfunction

*,†Cara E. Stepp, †,‡James T. Heaton, †,[a]Maia N. Braden, †Marie E. Jetté, †,[c]Tara K. Stadelman-Cohen, and *,†,‡,[b]Robert E. Hillman, *Cambridge, Massachusetts, †‡Boston, Massachusetts

Summary: The purpose of this study was to evaluate current neck tension palpation rating systems to determine interrater reliability and possible correlation with necksurface electromyography (sEMG, collected from three electrode recording locations) and to measure the third formant for /a/ during various vocal behaviors. This prospective study examined the neck muscle tension of 16 participants before and after a single session of voice therapy. Interrater reliability and relationships between palpation ratings and objective measures of sEMG (anterior neck) and the third formant for /a/ were assessed using Pearson's correlations (r). Interrater reliability was relatively low as measured by Pearson's correlations, although Wilcoxon signed-rank test results were similar as those in a previous study. Correlations between palpation ratings and sEMG and between ratings of laryngeal height and the third formant for /a/ were generally low. Correlations increased between anterior neck sEMG and ratings of suprahyoid muscle tension when examined in a reduced set of individuals with higher interrater reliability. Palpation rating scales do not reliably capture changes that may occur in neck muscle tension of typical voice therapy patients over one session. Consequently, little can be concluded from correlations between sEMG and palpation ratings.

Key Words: Voice–Muscle tension–Laryngeal palpation–Surface electromyography–Acoustic measures.

INTRODUCTION

When individuals demonstrate increased intrinsic laryngeal muscle contraction (vocal hyperfunction), it is thought that they often simultaneously contract the extrinsic laryngeal muscles and other superficial neck muscles in a similar hyperfunctional manner.[1] Strap muscle tension can be assessed through both visual and tactile inputs. A study by Altman et al reported on 150 patients who had been diagnosed with muscle tension dysphonia (MTD; a voice disorder with symptoms of vocal hyperfunction and no known structural change to the vocal fold or neurogenic disease of the larynx).[2] Based on a speech pathology evaluation of these patients, 70% were found to have "obvious cervical neck tension visible."[2] Practitioners have previously reported that observation of the inferior bellies of the omohyoid muscle crossing the supraclavicular fossae may show them to be tense and prominent during speech,[3] whereas further information about the extent of muscle tension can be gained by palpation of the larynx at rest and during voicing.[3] Excessive tension in disordered individuals has been noted via palpation over the major horns of the hyoid bone, over the superior cornu of the thyroid cartilage, along the anterior border of the sternocleidomastoid muscle, and throughout the suprahyoid musculature.[4]

Although palpation of neck musculature is a routine clinical procedure in the assessment and management of vocal hyperfunction,[1,5–8] only a few standardized rating scales have been developed. As part of a surface electromyography (sEMG) study, one speech-language pathologist rated "laryngeal-area tonicity" on a 1–5 linear scale, finding a high correlation between a single clinician's scores and mean sEMG during vowel production.[9] Angsuwarangsee and Morrison (2002) developed a linear 0–3 grading system of neck muscle tension based on the experiences and work of Lieberman[5] for research use in which each muscle group is graded based on specific text descriptors (Table 1).[10] Kooijman et al (2005) modified the system proposed by Angsuwarangsee and Morrison to include more muscle categories, as well as documentation about body posture.[11] Mathieson et al recently proposed a new rating system in which the muscle resistance of four categories is rated on a linear scale of 1–5, and the laryngeal position is noted as being one of the following: high held, neutral, lowered, or forced lowered (Figure 1).[12]

Angsuwarangsee and Morrison[10] assessed their rating system on 57 successive voice patients, with two independent investigators (otolaryngologists) examining each patient. Interrater reliability numbers based on Wilcoxon signed-rank tests were presented, with the reliabilities presented in the form of P values. Only one category, pharyngolaryngeal, exhibited statistically significant scores (less than 0.05), which was interpreted by the authors as having low interrater reliability. Mathieson et al[12] used palpatory evaluations in 10 individuals with MTD pre- and post-laryngeal manual therapy. Interrater reliability was not noted, as more than one clinician

Accepted for publication August 4, 2009.

From the *Division of Health Sciences & Technology, Harvard-MIT Cambridge, Massachusetts; †MGH Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts; and the ‡Department of Surgery, Harvard Medical School, Boston, Massachusetts.

[a]Currently belongs to the Department of Surgery, Voice and Swallowing Clinic, University of Wisconsin, Madison, Wisconsin. [b]Currently belongs to the Departments of Computer Science & Engineering and Rehabilitation Medicine, University of Washington, Seattle, Washington. [c]Currently belongs to the Department of Communicative Disorders, University of Wisconsin, Madison, Wisconsin.

Address correspondence and reprint requests to Cara E. Stepp, Computer Science & Engineering, Box 352350, University of Washington, Seattle, WA 98195. E-mail: cstepp@alum.mit.edu

Journal of Voice, Vol. 25, No. 1, pp. 67-75
0892-1997/$36.00

doi:10.1016/j.jvoice.2009.08.001

TABLE 1.
Angsuwarangsee and Morrison Palpation System

Rating	Description
Suprahyoid muscles	
0	Soft at rest but may slightly contract on phonation
1	Soft at rest but mild low pitch and moderate high pitch on contraction
2	Some tension at rest and tense with jaw protrusion on phonation
3	Tense all the time and maximally tight on phonation
Thyrohyoid muscles	
0	No muscular contraction at rest but mild on phonation
1	Soft thyrohyoid space at rest and some contraction on phonation
2	Tense, narrow thyrohyoid space at rest and moderate contraction on phonation
3	Very tense with closed thyrohyoid space all the time
Cricothyroid muscles	
0	Normal cricothyroid space and phonatory movement
1	Narrowing of cricothyroid space at rest and some movement on phonation
2	Anterior displacement of cricoid cartilage with narrowing of cricothyroid space at rest and closing of the space on phonation
3	Closed cricothyroid space all the time
Pharyngolaryngeal muscles	
0	Soft, easy to rotate the larynx to 90° and palpate PCA muscle and arytenoid movement on sniffing
1	Slightly tense and cannot palpate PCA muscle movement on sniffing
2	Moderately tense and difficult to rotate the larynx but still can palpate the posterior edge of thyroid cartilage
3	Very tense and cannot rotate the larynx at all

Source: Reprinted with permission from Ref. 10.
Abbreviation: PCA, posterior cricoarytenoid.

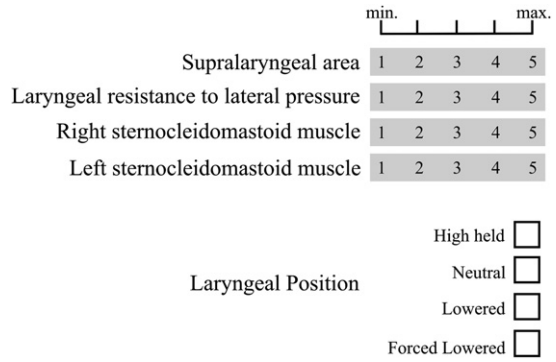

FIGURE 1. Mathieson et al palpation rating scale. Adapted with permission from Ref. 12.

was not used for evaluation. Further, it appears that the evaluator (a speech-language pathologist) was the same individual providing therapy.

The work of Redenbaugh and Reich[9] is the only study to explore the relationship between sEMG and neck palpation ratings. In their study, laryngeal-area tonicity was evaluated by a single speech-language pathologist during tidal breathing, production of the vowel /a/ for 15 seconds, and reading aloud. Laryngeal-area tonicity was rated using a 5-point, equal-appearing interval scale. The Pearson's correlations between the palpation score and the sEMG during the vowel and speech tasks were found to be 0.86 and 0.9, respectively. No interrater reliability measures were attempted given that there was only one rater. The study examined seven individuals with MTD and seven individuals with healthy normal voice. Because of the bimodal nature of this sample and with the participants likely to represent alternate ends of the spectrum of neck muscle tension related to voice disorder, correlation values may possibly be inflated. Even though the palpation procedure and scale used by Redenbaugh and Reich[9] was published in 1989, it has not been the subject of further published research, and to our knowledge, is not widely used in the clinic. Furthermore, this previous study used only one electrode position overlying the thyrohyoid membrane. To understand the relationships between sEMG and clinical ratings of palpation, it is necessary to determine among multiple electrode recording locations and vocal behaviors, those that may correlate most accurately with more widely used clinical ratings.

Roy and Ferguson[13] examined changes in formant frequencies pretherapy versus posttherapy in 75 participants with functional dysphonia, finding significant decreases in the first, second, and third formants after therapy. The authors interpreted this finding as evidence of laryngeal lowering as a result of therapeutic intervention. The neck palpation rating system proposed by Mathieson et al[12] requires the evaluator to note laryngeal position of the participant as being high held, neutral, lowered, or forced lowered using a nominal scale. Their study applied this system to 10 participants and found insignificant changes posttherapy in average laryngeal height. Acoustic analysis at both time points found a trend of increased second formant during vowel production posttherapy, which would be inconsistent with laryngeal lowering. Their study did not attempt to correlate changes in perceived larynx height with formant changes.

The third formant shows less variation across different vowel productions than the second formant, which is more likely to be affected by changes in vowel articulation. Moreover, recent

work has shown that treatment for MTD can also affect articulation, leading to increased vowel space.[14] The third formant should be more correlated with vocal tract length, such that changes to the third formant may offer objective confirmation of changes in laryngeal position, especially because of short-term therapy effects. Assessing the relationship between judgments of laryngeal height and corresponding changes in the third formant may offer more useful information about the utility of this clinical scale for assessing laryngeal position.

Clinically, the presence of excessive neck tension is noted as a sign of vocal hyperfunction, informing both diagnosis and treatment.[4–8,10–12] However, current methods of assessment of neck muscle tension[10,12] depend on tactile measures, which are subjective and lack a large dynamic range of measurement. The use of sEMG and objective acoustic methods to monitor changes in neck tension and/or laryngeal position in patients with voice disorders could lead to more standardized care and improved information about patient progress. It is currently still unknown whether neck sEMG recordings or formant changes correlate well with clinical palpation-based ratings. Also, to use sEMG optimally, it is necessary to determine the electrode recording locations and vocal behaviors that correlate most accurately with clinical ratings. The purpose of this study was to evaluate the neck tension palpation tension rating systems of Angsuwarangsee and Morrison[10] and Mathieson et al[12] to determine whether reproducible results could be obtained, as measured by interrater reliability measures (Pearson's correlations), when administered by speech-language pathologists previously unfamiliar with these systems. A further goal of this study was to ascertain whether the systems were correlated with objective measures of neck tension (sEMG) and laryngeal height (third formant for /a/) of individuals receiving therapy for voice disorders related to vocal hyperfunction. These two scales were used as a comparison with acoustic changes in the third formant and neck sEMG collected from multiple electrode recording locations during various vocal behaviors to understand how differences in scale structure may affect correlations with objective measures.

METHODS

Participants

Participants were 16 adult volunteers (13 females and three males) with mean age of 24.9 years ($R = 18–41$ years) receiving voice therapy because of a voice disorder related to vocal hyperfunction (eg, MTD, vocal nodules). Table 2 lists the diagnoses of the participants as well as age and sex. Participants were varied in their progress in voice therapy, with their research participation taking place during one of multiple visits in the course of their therapy.

TABLE 2.
Participant Diagnosis and Demographic Information

Participants	Age	Sex	Diagnosis
P1	31	M	Muscle tension dysphonia
P2	32	F	Vocal fold nodules
P3	22	F	Muscle tension dysphonia
P4	22	F	Vocal fold nodules
P5	18	F	Muscle tension dysphonia
P6	26	M	Muscle tension dysphonia
P7	27	F	Muscle tension dysphonia
P8	19	F	Muscle tension dysphonia
P9	22	F	Muscle tension dysphonia
P10	24	F	Muscle tension dysphonia
P11	21	F	Vocal fold nodules
P12	41	M	Muscle tension dysphonia
P13	20	F	Muscle tension dysphonia
P14	22	M	Vocal fold nodules
P15	26	F	Vocal fold nodules
P16	25	F	Muscle tension dysphonia

Abbreviations: M, male; F, female.

Clinical palpation methodology

Two of three total independent certified speech-language pathologists assessed each participant before and after therapy using the two clinical palpation ratings of Angsuwarangsee and Morrison[10] and Mathieson et al.[12] The "primary" rater was the same clinician providing therapy to the participant. A second rater was another speech-language pathologist who was unfamiliar with the patient.

A total of three certified speech-language pathologists who specialize in voice participated in the clinical assessment portion of this study. All three of the speech-language pathologists who participated in this study completed their clinical fellowship training in a specialized voice clinic, and all had at least 1 year of experience working full time in a specialized voice clinic, with case loads consisting exclusively of patients with voice disorders. All of them had extensive experience with laryngeal palpation and manipulation as a part of clinical practice before the initiation of this study. Participation among the three speech-language pathologists was approximately equal, with each completing pretherapy and posttherapy assessments for 8–12 participants. The speech-language pathologists were trained internally by reading the primary literature behind the rating systems[10,12] and a chapter on techniques of manual therapy,[5] and then, each applying the two neck tension rating systems to the same individual, comparing rating decisions, and discussing scoring issues at length. This internal training lasted approximately 1.5 hours. After official recruitment and recording of participants had been initiated, no feedback was given to participating clinicians regarding their agreement with one another.

Surface electromyography and acoustic recording methodology

The sEMG and acoustic recordings consisted of a brief vocal assessment of the participant, which included three trials of the vowel /a/, read speech (The Rainbow Passage[15]), and spontaneous running speech. Spontaneous speech was elicited in response to the investigator asking the participant a probing question (eg, "Can you tell me what you do in a typical therapy session?"). After completion of these speech tasks, maximal

voluntary contraction (MVC) maneuvers were performed. These consisted of asking the participants to perform neck contraction against manual resistance for the purpose of normalizing sEMG data (see the following Data analysis section). To ensure that systematic differences did not exist in the MVC force production in the pretherapy and posttherapy recordings, a dynamometer (Chatillon DPP-50; Ametek, Inc., Paoli, PA) was used during neck muscle contraction against manual resistance for all but three participants, and the maximal force was recorded. The MVC forces ranged from 14 to 42 lb_f by participant, but there was no statistically significant difference between pretherapy and posttherapy MVC forces (paired Student's t test, df = 12, $P = 0.85$).

Simultaneous neck sEMG and acoustic signals from a lavalier microphone (Sennheiser MKE2-P-K, Wedemark, Germany) were recorded digitally with Delsys (Delsys Inc, Boston, MA) hardware (Bagnoli Desktop System) and software (*EMG-works 3.6*) at 20 kHz. The sEMG signals in this study were recorded and analyzed in view of current European standards.[16] Participants' necks were prepared for electrode placement by cleaning the neck surface with an alcohol pad and "peeling" with tape to reduce electrode-skin impedance, noise, DC voltages, and motion artifacts. The neck sEMG was recorded using two two-channel Bagnoli systems (Delsys Inc) with three Delsys 3.1 double differential surface electrodes placed parallel to the underlying muscle fibers of the following: (1) thyrohyoid, omohyoid, and sternohyoid muscles; (2) cricothyroid and sternohyoid muscles; and (3) sternocleidomastoid muscle (Figure 2). The Delsys 3.1 double differential surface electrodes consisted of three 10-mm silver bars with interelectrode distances of 10 mm. Double differential electrodes were chosen to increase spatial specificity of the sEMG recordings and to eliminate the possibility of electrical crosstalk, a risk given the electrode proximity.

Electrode 1 was centered about 1 cm lateral to the neck midline, as far superior as was possible without impeding jaw opening of the participant. Electrode 2 was centered on the gap between the cricoid and thyroid cartilages of the larynx, and centered at 1 cm lateral to the midline, contralateral to electrode 1. Electrode 3 was centered one-third of the distance from the sternal notch of each participant to his or her mastoid process following the recommendations of Falla et al.[17] A ground electrode was placed on the superior aspect of the participant's left shoulder. The sEMG signals were preamplified and filtered using Delsys Bagnoli systems set to a gain of 1000 with a bandpass filter (roll-off frequencies of 20 and 450 Hz).

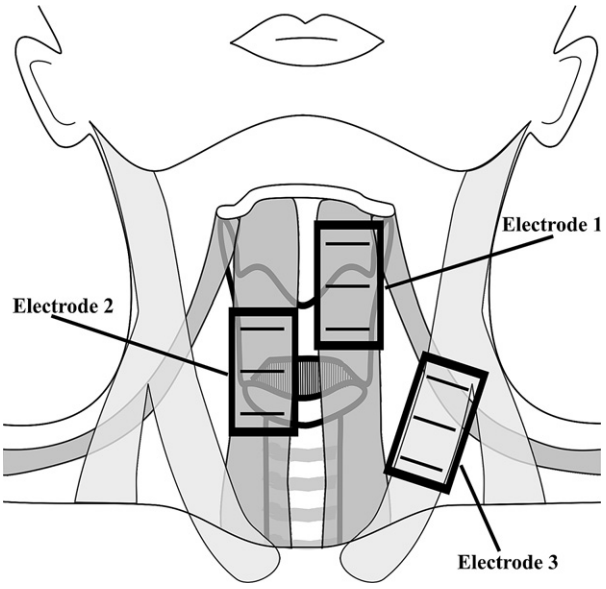

FIGURE 2. Schematic of sEMG electrode recording locations.

Data analysis

The variability associated with differences in neck-surface electrode contact and placement was minimized by normalizing the sEMG to a reference contraction at MVC so that sEMG data gathered could be compared between pretherapy and posttherapy recordings. All sEMG data were computed as the root mean square (RMS) and then normalized via MVC (in RMS) in windows of 1 second using custom software written in *MATLAB* (Mathworks Inc., Natick, MA). Although studies have shown that for simple, one-joint systems, submaximal contractions are more reliable for normalization,[18,19] it has been proven that the MVC references is more reliable for anterior neck musculature.[20] Consequently, all sEMG data were analyzed in terms of % MVC. The third formant during three trials of the vowel /a/ was estimated using the linear predictive coding analysis in *Praat* acoustic analysis software.[21] All of the formants found were consistent with the expected ranges specified in the literature (eg, Stevens, 2000[22]).

Correlations were calculated between the normalized RMS sEMG and clinical ratings of various muscle groups to ascertain the level of association between the assorted measures. Interrater reliability measures were calculated with Pearson's correlation for most elements of the two clinical rating systems using the assessment of the two speech-language pathologists. To compare these data with previous reports of interrater reliability, Wilcoxon signed-rank tests were also performed between raters. Interrater reliability of the larynx position measure of the Mathieson et al[12] palpation system was assessed using Cohen's kappa because of the nominal nature of the scale. A two-factor analysis of variance (ANOVA) was used to examine the effect of rater and perceived larynx height change (the larynx position measure of the Mathieson et al[12] palpation system) on the measured changes in the third formant for the /a/ vowel. Statistical analysis was performed using *Minitab* Statistical Software (Minitab Inc., State College, PA).

RESULTS

Interrater reliability

Interrater reliability between the two raters of neck tension using all pretherapy and posttherapy judgments was assessed using Pearson's correlations for all categories of the Angsuwarangsee and Morrison system, and for the first four categories of the Mathieson et al system. Pearson's correlations were generally poor but differed slightly as a function of muscle group. For comparison with the work of Angsuwarangsee and Morrison,[10]

Wilcoxon signed-rank tests were also performed on rater judgments. The Pearson's correlations and *P* values from the Wilcoxon signed-rank tests are shown in Figure 3.

None of the Pearson's correlations were greater than 0.6, with the lowest at 0.23. The Wilcoxon signed-rank tests showed no significant difference between raters for most categories except the cricothyroid and pharyngolaryngeal measures of the Angsuwarangsee and Morrison system.

When judging laryngeal position, no raters used the designations for "lowered" or "forced lowered," essentially creating a binary rating system of "high held" or "neutral." Of the 32 assessments of laryngeal position, a total of 22 matched perfectly (69%). Cohen's kappa was calculated for each response ("high held" and "neutral"), equaling 0.38 for both.

Interrater reliability between the two raters of neck tension using the change between pretherapy and posttherapy judgments was also assessed. The Pearson's correlations and *P* values from the Wilcoxon signed-rank tests are shown in Figure 4. Several of the Pearson's correlations were near zero or even negative, although the left sternocleidomastoid (SCM) of the Mathieson et al system had a Pearson's correlation greater than 0.6. The Wilcoxon signed ranks showed no significant difference between raters for any category. Of the 16 assessments of laryngeal position change, a total of 10 matched perfectly (63%). Cohen's kappa was calculated for each response, equaling 0.02 for both.

Correlation between palpation ratings and surface electromyography

The left panel of Figure 5 shows the Pearson's correlations between each palpation measure and sEMG from all possibly relevant electrode locations during rest, reading speech, and spontaneous speech. The pharyngolaryngeal measure was not included in the correlation analysis, because there were no appropriate electrode locations. The sEMG from electrode positions 1 and 2 were compared with suprahyoid, thyrohyoid, and cricothyroid ratings from the Angsuwarangsee and Morrison system and the supralaryngeal and lateral pressure ratings from the Mathieson et al system. The sEMG from electrode position 3 was compared with both left and right SCM ratings even though sEMG was collected only from the patient's left SCM.

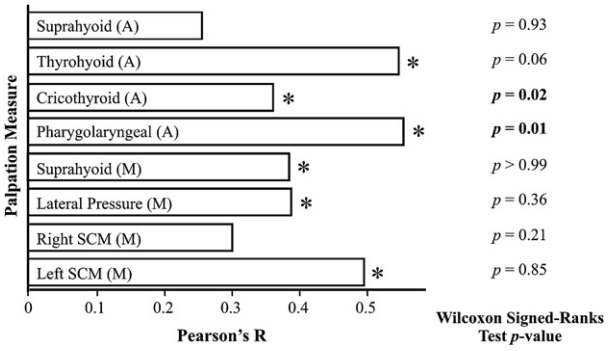

FIGURE 3. Interrater reliability for pretherapy and posttherapy palpation. Palpation measures marked with an (A) are part of the Angsuwarangsee and Morrison system[10] and those marked with an (M) are part of the Mathieson et al system.[12] *Asterisks* note those measures for which the Pearson's correlation was significantly ($P < 0.05$) greater than 0.

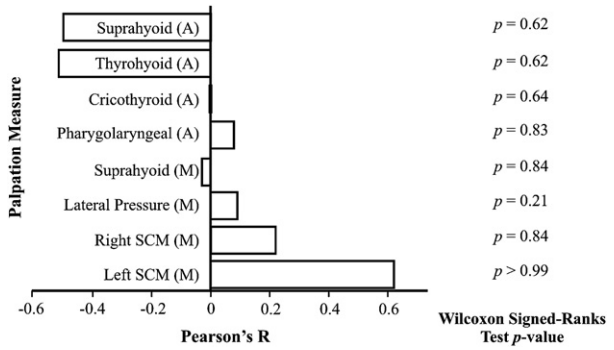

FIGURE 4. Interrater reliability for the change between pretherapy and posttherapy judgments. Palpation measures marked with an (A) are part of the Angsuwarangsee and Morrison system[10] and those marked with an (M) are part of the Mathieson et al system.[12]

To reduce the effects of poor interrater reliability on correlations between sEMG and palpation ratings, participants whose pre- and postchange ratings differed between raters by 2 or more scale points on any dimension were excluded, resulting in a reduced set of N = 8 "high-reliability" participants. The right panel of Figure 5 shows the Pearson's correlations for the reduced set.

Relationship between perceived laryngeal height and the third formant

Laryngeal height (the larynx position measure of the Mathieson et al[12] palpation system) was most frequently rated as the same in both pretherapy and posttherapy recordings. In some cases, one or both raters felt that a participant moved from "high held" to "neutral" during the course of therapy. Changes in the third formant averaged at 1 Hz, ranging from −164 Hz (indicating a lower larynx posttherapy) to 281 Hz (indicating a higher larynx posttherapy). These changes did not appear to be associated with perceived laryngeal height. A two-factor ANOVA assessing the effect of rater and perceived larynx height change on the measured changes in the third formant showed no effect of either variable ($P > 0.05$).

DISCUSSION

Interrater reliability

Interrater reliability based on single time-point assessments, as measured with Pearson's correlation, was generally low across all dimensions of both scales, and did not improve with the use of pretherapy and posttherapy differenced data. The highest reliabilities were seen for the thyrohyoid and pharyngolaryngeal assessments of the Angsuwarangsee and Morrison system and the left SCM assessment of the Mathieson et al system. No systematic differences in the interrater reliability emerged between the two systems. The difference in reliability between the right ($R = 0.30$) and left ($R = 0.49$) SCM assessments is puzzling given that clinicians tended to use both hands during both SCM assessments. One possibility is that patient asymmetries

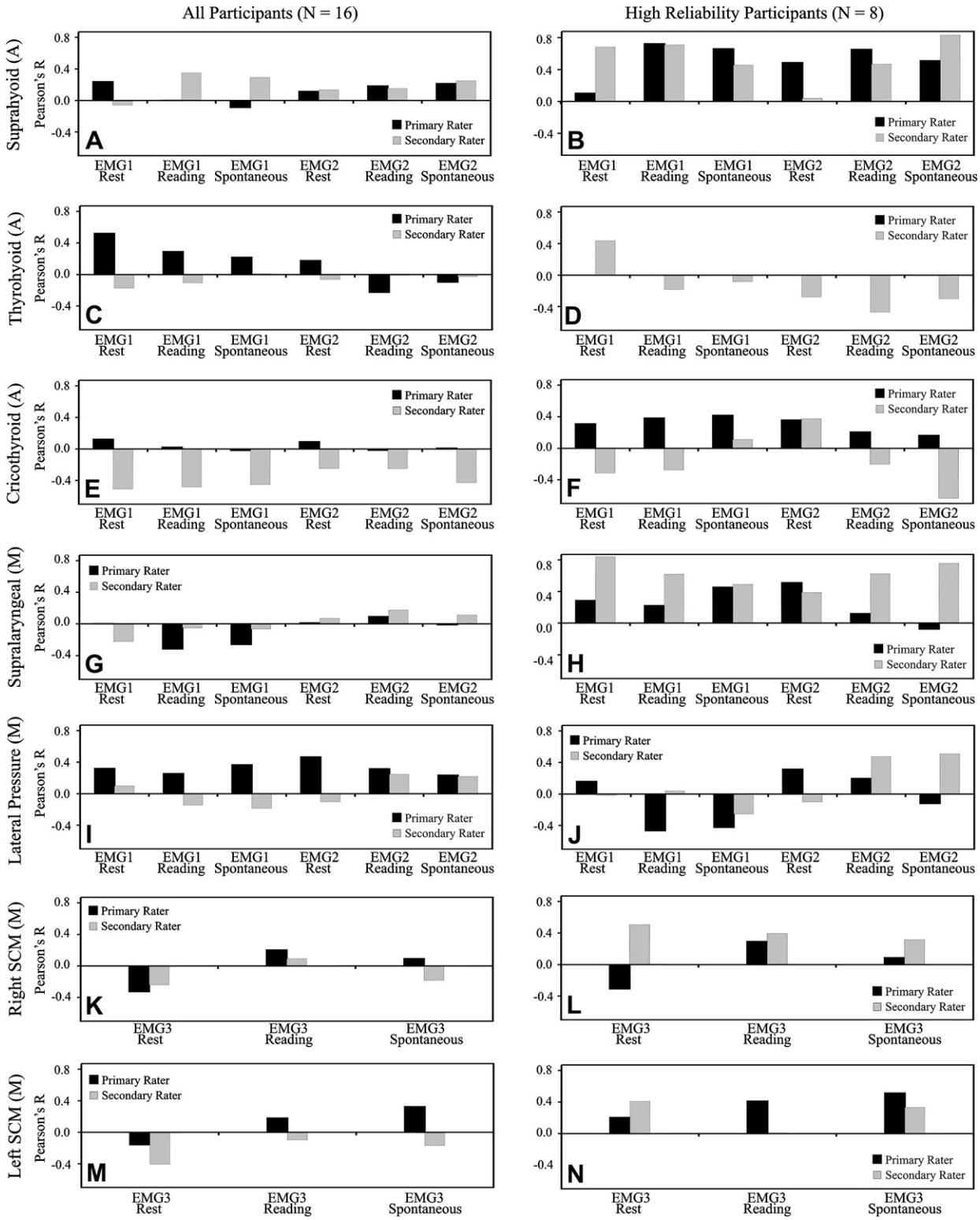

FIGURE 5. Pearson's correlations between palpation measures and sEMG. Palpation measures marked with an (A) are part of the Angsuwarangsee and Morrison system[10] and those marked with an (M) are part of the Mathieson et al system.[12] The left panel (**A**, **C**, **E**, **G**, **I**, **K**, and **M**) is for the entire set of participants (N = 16) and the right panel (**B**, **D**, **F**, **H**, **J**, **L**, and **N**) is for the reduced set of "high-reliability participants" (N = 8).

could, perhaps, have affected the variability of palpable muscle tension, leading to reduced interrater reliability, but there is no evidence here to support overall right-left asymmetry.

Angsuwarangsee and Morrison[10] used Wilcoxon signed-rank tests as a measure of interrater reliability, finding the only significant differences ($P < 0.05$) in judgments for the

pharyngolaryngeal assessment (interpreted by them as poor reliability). Similarly, we also found *P* values greater than 0.05 for all assessments except the cricothyroid and pharyngolaryngeal assessments of the Angsuwarangsee and Morrison system, indicating that the rater performance was not dissimilar to that from their study. However, the relationship between Pearson's correlation values and the *P* values resulting from the Wilcoxon signed-rank testing calls into question the appropriateness of using the Wilcoxon signed-rank test as a measure of interrater reliability. The Wilcoxon signed-rank test assesses the likelihood of mean differences between measures being nonzero, not reliability. Greater overall variance of two judges (lack of agreement) would, therefore, increase *P* values, whereas they could lead to lowered Pearson's correlations and vice versa. Raters who are highly unreliable but do not consistently agree in the direction of their disagreement would have a high *P* value because of the large variance in their differences but a low Pearson's correlation. As an example, in the interrater reliability data shown in Figure 3, the pharyngolaryngeal measure has both the highest Pearson's correlations (indicative of high reliability) and the smallest *P* values (indicative of a nonzero difference between raters). This throws into doubt the high interrater reliability reported by Angsuwarangsee and Morrison,[10] because the interpretation was based on the use of Wilcoxon signed-rank tests, and no Pearson's correlations were reported.

The nominal scale used to assess larynx position in the Mathieson et al system showed moderately low values of Cohen's kappa, with nonsignificant *P* values to assess the likelihood of kappa > 0. Kappa values range from −1 to 1, where a kappa of 1 indicates perfect agreement between raters, and a kappa of 0 indicates agreement the same as that expected by chance. The results of kappa analysis indicate that the interrater agreement of laryngeal height is not significantly higher than that because of chance. One possible factor in this lack of agreement is the prevalence of different internal definitions of laryngeal height: some clinicians may associate a high larynx position with merely a high hyoid, whereas others might require the entire larynx to be raised. Regardless, the low values of kappa indicate that these scales do not provide reliable indications of laryngeal height.

Correlations between palpation ratings and objective measures

Using the full data set, correlations between sEMG and palpation ratings were generally low, with many near zero or even negative. This is not surprising given the low interrater reliability of the palpation ratings. There does not appear to be an effect of task on correlations, with resting sEMG resulting in correlations similar to those for sEMG collected during running speech. Repeating correlation analyses on the high-reliability participants resulted in much higher correlations overall. In particular, correlations between sEMG from electrode positions 1 and 2 and suprahyoid/supralaryngeal ratings of both systems increased. Also, correlations between sEMG from electrode position 3 and both left and right SCM ratings increased. One interpretation is that there is an underlying correlation between these sets of ratings and corresponding sEMG, which was made clearer with the elimination of some of the variance in the palpation scoring. However, we cannot rule out the possibility that these increases are mere artifacts produced by our manipulation of the data set.

No association was seen between mean changes in the third formant of the /a/ vowels and the larynx position palpation rating changes pretherapy and posttherapy. Mathieson et al also found a lack of changes in formant frequencies (first and second) pre- and post-manual therapy in 10 patients with MTD.[12]

Participants in this study were current therapy patients reporting for one of a number of recommended therapy sessions. Unlike so-called functional dysphonia patients for whom voice quality frequently changes drastically over the course of a single therapy session, it is more likely that these individuals displayed patterns of voice production and muscle tension that were more resistant to change. Further, therapy sessions were not necessarily directly targeting muscle tension (eg, laryngeal massage), but varied as a function of individual patient needs. The lack of association between palpation ratings and objective measures could, therefore, also be a result of a lack of effectively large tension changes in the pretherapy and posttherapy conditions, given that the study was only conducted over a single session. However, these types of patients who report for multiple therapy sessions over time are those for whom a reliable palpation scale and/or objective assessment protocol would be most useful as a way of marking therapeutic progress.

Issues with respect to clinical adoption of palpation rating scales

Although neck muscle palpation for assessment and management of vocal hyperfunction is commonplace in specialized voice clinics,[1,5–8] formal documentation of neck tension is not widely practiced. The reliable recording of neck tension through palpation ratings or objective measures could lead to more standardized and well-informed patient care. Obstacles to the advised use of the two scales evaluated here stem from poor interrater reliability. The clinicians who participated as raters in this study described several major flaws that they perceived with these scales, which included overly broad distinctions, general lack of bilateral (left vs right) discriminations, neglect of essential categories, and inappropriate guiding text. None of the raters in this study felt that either system was a valuable addition to his or her current (qualitative) protocol for monitoring neck muscle tension across the course of therapy.

A major criticism of both systems was the lack of discrimination possible. The 4- and 5-point scales were often insensitive to within-therapy changes, even when the clinician believed that they could palpate a change in muscle tension. It is possible that a scale with more divisions or a visual analog scale, such as the one used by the Consensus Auditory-Perceptual Evaluation of Voice[23], could result in more reliable within-therapy results. However, more studies should be carried out on this matter given that increasing sensitivity from a 4-point scale to a visual analog scale can, in some cases, result in decreased interrater reliability.[24]

Vocal hyperfunction often causes patients to present with imbalanced muscular patterns.[5] These patterns cause asymmetry that may be evident during laryngoscopy and through palpation. However, with the exception of the right and left SCM categories of the Mathieson et al scale, no other categories distinguish between right and left muscle behaviors. The raters of the present study often felt significant differences bilaterally, leading to rater confusion given the limited options for ratings. Likewise, lack of discrimination between anterior and posterior stiffness for the thyrohyoid category of Angsuwarangsee and Morrison's system also lead to rater confusion, because the accompanying text referred to both muscular contraction and differences in the thyrohyoid space. Further, in palpation of the SCM (right and left), often differences were felt between the superior and inferior ends of the SCM. Ratings based on the "average" muscle tension in cases like these might mask clinically relevant changes in muscle tension that speech-language pathologists have the ability to palpate.

The accompanying text of the Angsuwarangsee and Morrison system often caused frustration for the raters of this study. The text descriptions of this system have multiple parts, and raters frequently identified parts of multiple text descriptors that spanned different numerical ratings within the same patient. One example of this was seen more than once for the thyrohyoid measure: agreement with "some contraction on phonation" for a rating of "1," and agreement with "tense, narrow thyrohyoid space at rest" for a rating of "2" (see Table 1 for reference to this system). In some cases, raters even identified with text descriptors of nonadjacent ratings (eg, agreeing with text for a rating of 0 and 2). In the particular case of the CT text descriptors, raters in this study felt that the emphasis on the size of the cricothyroid space rather than the tension felt in the cricothyroid muscle was misplaced. Likewise, raters felt that the description for pharyngolaryngeal category that asks the rater to attempt to rotate the larynx a full 90° was, in most cases, inappropriate. The general consensus of the raters of this study, all of whom had several years of experience working exclusively in voice, was that the text descriptions were a distraction. It is possible, however, that the text descriptors in this system may be of more use for clinicians having less experience with voice therapy, in which case, specific text descriptors may serve as a much-needed guide.

CONCLUSIONS

This study examined two recently published clinical neck tension palpation tension rating systems in individuals receiving a single session of voice therapy for hyperfunction-related disorders to determine whether the systems could produce reliable results when administered by speech-language pathologists previously unfamiliar to them. The study further attempted to determine whether either of the systems was correlated with objective measures of neck tension (sEMG and change in the third formant for the vowel /a/). For the 16 individuals studied, Pearson's correlations between raters were generally low, and little correspondence was found between ratings and objective measures. However, a smaller set of subjects with greater interrater agreement showed a stronger relationship between palpation ratings of the supralaryngeal area and sEMG measured on the anterior neck. These scales may be helpful in providing guidance for voice practitioners who are beginners and to mark long-term progress from a disordered to fully rehabilitated state, but the current results indicate that they may not be sensitive enough for use as monitoring tools across individual sessions in the course of therapy, and their clinical use is not recommended for this purpose.

REFERENCES

1. Aronson AE. *Clinical Voice Disorders: An Interdisciplinary Approach*. 1st ed.). New York: Thieme-Stratton, Inc.; 1980.
2. Altman KW, Atkinson C, Lazarus C. Current and emerging concepts in muscle tension dysphonia: a 30-month review. *J Voice*. 2005;19:261–267.
3. Morrison M. Pattern recognition in muscle misuse voice disorders: how I do it. *J Voice*. 1997;11:108–114.
4. Roy N, Ford CN, Bless DM. Muscle tension dysphonia and spasmodic dysphonia: the role of manual laryngeal tension reduction in diagnosis and management. *Ann Otol Rhinol Laryngol*. 1996;105:851–856.
5. Lieberman J. Principles and techniques of manual therapy: applications in the management of dysphonia. In: Harris T, Harris S, Rubin J, Howard DM, eds. *The Voice Clinic Handbook*. London: Whurr Publishers Ltd; 1998: 91–138.
6. Roy N, Bless DM. Manual circumlaryngeal techniques in the assessment and treatment of voice disorders. *Curr Opin Otolaryngol Head Neck Surg*. 1998;6:151–155.
7. Roy N, Bless DM, Heisey D, Ford CN. Manual circumlaryngeal therapy for functional dysphonia: an evaluation of short- and long-term treatment outcomes. *J Voice*. 1997;11:321–331.
8. Roy N, Leeper HA. Effects of the manual laryngeal musculoskeletal tension reduction technique as a treatment for functional voice disorders: perceptual and acoustic measures. *J Voice*. 1993;7:242–249.
9. Redenbaugh MA, Reich AR. Surface EMG and related measures in normal and vocally hyperfunctional speakers. *J Speech Hear Disord*. 1989;54: 68–73.
10. Angsuwarangsee T, Morrison M. Extrinsic laryngeal muscular tension in patients with voice disorders. *J Voice*. 2002;16:333–343.
11. Kooijman PG, de Jong FI, Oudes MJ, Huinck W, van Acht H, Graamans K. Muscular tension and body posture in relation to voice handicap and voice quality in teachers with persistent voice complaints. *Folia Phoniatr Logop*. 2005;57:134–147.
12. Mathieson L, Hirani SP, Epstein R, Baken RJ, Wood G, Rubin JS. Laryngeal manual therapy: a preliminary study to examine its treatment effects in the management of muscle tension dysphonia. *J Voice*. 2009;23: 353–366.
13. Roy N, Ferguson N. Formant frequency changes following manual circumlaryngeal therapy for functional dysphonia: evidence of laryngeal lowering? *J Med Speech Lang Pathol*. 2001;9:169–175.
14. Roy N, Nissen SL, Dromey C, Sapir S. Articulatory changes in muscle tension dysphonia: evidence of vowel space expansion following manual circumlaryngeal therapy. *J Commun Disord*. 2009;42:124–135.
15. Fairbanks G. *Voice and Articulation Drillbook*. 2nd ed.). New York: Harper and Row; 1960.
16. Hermens HJ, Freriks B, Merletti R, et al. *European Recommendations for Surface ElectroMyoGraphy: Results of the SENIAM Project*. Enschede, The Netherlands: Roessingh Research and Development; 1999.
17. Falla D, Dall'Alba P, Rainoldi A, Merletti R, Jull G. Location of innervation zones of sternocleidomastoid and scalene muscles—a basis for clinical and research electromyography applications. *Clin Neurophysiol*. 2002;113: 57–63.
18. Allison GT, Marshall RN, Singer KP. EMG signal amplitude normalization technique in stretch-shortening cycle movements. *J Electromyogr Kinesiol*. 1993;3:236–244.

19. Yang JF, Winter DA. Electromyography reliability in maximal and submaximal isometric contractions. *Arch Phys Med Rehabil*. 1983;64:417–420.
20. Netto KJ, Burnett AF. Reliability of normalisation methods for EMG analysis of neck muscles. *Work*. 2006;26:123–130.
21. Praat: doing phonetics by computer [computer program]. Version 5.0.20: Available at http://www.praat.org/; 2008. Last accessed August 25, 2009.
22. Stevens KN. *Acoustic Phonetics*. Cambridge, MA: MIT Press; 2000.
23. Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124–132.
24. Wuyts FL, De Bodt MS, Van de Heyning PH. Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice*. 1999;13:508–517.