7-29-2021

Measuring the Association Between Vocal Effort and Intensity with Emotion Picture Viewing using the Borg Centimax

Hayleigh Wilson

MEASURING THE ASSOCIATION BETWEEN VOCAL EFFORT AND INTENSITY WITH EMOTION PICTURE VIEWING USING THE BORG CENTIMAX

by

Hayleigh Wilson

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

Major: Speech Language Pathology

The University of Memphis

August, 2021

ABSTRACT

MEASURING THE ASSOCIATION BETWEEN VOCAL EFFORT AND INTENSITY WITH EMOTION PICTURE VIEWING USING THE BORG CENTIMAX

Objective. This research investigated the relative values of intensity when participants were given a target vocal effort level (VEL) and if transient mood impacted these values.

Method. Twenty-seven participants completed three experimental blocks. 36 trials were completed in Block 1 consisting of rote speech tasks, reading tasks, and 3 unique map tasks all executed at four target VELs. Block 2 consisted of an additional 36 rote speech task trials with the addition of emotional stimuli. Finally, 12 additional automatic speech trials were executed in Block 3.

Results. Results revealed vocal intensity was significantly distinct for all elicited VELs within a given experimental block and repeatable for each VEL when comparing the initial and final blocks of the experiment. Additionally, transient mood did not have a noticeable effect on the produced intensity levels.

Conclusion. The current study provided insight and evidence to the significant association between vocal effort and intensity output.

Table of Contents

Chapter		Page
List of Figures		v
List of Tables		vi
1.	Introduction	1
	Literature Review	3
	Definition of Vocal Effort	3
	Measurement of Vocal Effort	7
	Borg CR Scales and Measuring of Vocal Effort	9
	Borg's Relation to Other Scales	10
	External Influences on Effort	12
	Internal Influences on Effort	14
	Cognitive Influences on the Perception of Effort	15
	Emotional Influences on the Perception of Effort	16
2.	Purpose and Reason for the Current Research	19
3.	Methods	21
	Participants	21
	Measures	21
	Independent Measures (Stimuli)	21
	Dependent Variables	23
	Primary Variable of Interest	23
	Manipulation Variables	24
4.	Training and Procedures	25
	Instrumentation	27
	Analysis	28
5.	Results	29
	Self-Assessment Manakin	29
	Positive Affect Negative Affect Scale	30

Repeatability Across Participants and Setting 31
Repeatability within Participants and Setting 32
Effects of Mood on Vocal Effort Levels 34
Differences in Intensity Levels 34
Differences between Vocal Effort Levels 35
6. Discussion 36
Main Findings 37
Implications for Clinical Practice 40
Supplementary Findings 40
Limitations 41
Future Directions 41
Conclusion 43
References 44

List of Figures

Figure	Page
1. Example of the presentation of the vocal effort scale	23
2. Schematic for the instrumentation	27
3. SAM-V & SAM-A Verification	30
4. Significance between PANAS scores	31
5. Repeatability within Females	33
6. Repeatability within Males	34
7. Female VELs Compared Across Conditions	36
8. Male VELs Compared Across Conditions	36

List of Tables

Figure **Page**

1. Females: Means, standard deviations, and t statistics for each VEL between Block 1 and Block 3 33

2. Males: Means, standard deviations, and t statistics for each VEL between Block 1 and Block 3 34

Introduction

Vocal effort is the perceived exertion of a speaker's vocal output (Hunter et. al 2020). Regardless of the status of vocal health and hygiene, an individual can experience effort depending on the difficulty of the vocal task or the barriers to communication from the environment (Anand, Bottalico, & Gray, 2019; Hunter et. al, 2020). The degree to which the speaker's perception of effort relates to the demands of voice use, or physiological vocal output, has been investigated by many researchers in the field. One of the primary challenges in this line of investigation is the conflation of related terms such as vocal fatigue, vocal demand, and vocal load observed in the literature with many terms used interchangeably (Hunter et. al, 2020). A recent consensus paper addressed this issue making headway in clarifying these terms by creating definitions for vocal load (or demand), vocal demand response, vocal effort, and vocal fatigue that can be implemented universally in future studies to begin to disambiguate these constructs (Hunter et. al, 2020). These recommendations were put forth to allow for more precise construct definition that employs measures commensurate with their characterization. For example, the term vocal effort, which reflects a perceptual phenomenon, psychophysical scaling measurement techniques seem reasonable. Psychophysical scaling quantifies any perceptual experiences, making it possible to compare them to the physical phenomena that trigger these perceptions (Gescheider, 1988).

Historically, measuring vocal effort presented challenges in establishing a reliable and valid measure lies in the nature of voicing, which contains multiple physiological processes. Since multiple physical factors are involved in voicing, identifying which physical factor contributes to effort requires extensive study. Thus, a circular pattern exists. In order to study the physiological contributions of vocal effort, a measure of vocal effort level is necessary. However, to establish a measure of effort, physiological contributions must be known.

One line of research has relied on using the Borg CR10 as a means of capturing vocal effort (Anand, Bottalico, & Gray, 2019; Ford Baldner, Doll, & van Mersbergen, 2015; van Leer & van Mersbergen, 2017; van Mersbergen, Lyons, Riegler, 2017; van Mersbergen, Vinney, & Payne, 2019). This method employed scales that were validated on other populations but not necessarily on vocal behavior. Throughout the studies, vocal effort ratings were statistically different between negative and positive moods, (van Mersbergen & Delaney, 2014; van Mersbergen, Lyons, Riegler, 2017; van Mersbergen, Patrick, & Glaze, 2008), high and low cognitive load (van Mersbergen, Vinney, & Payne, 2019) and normal and disordered voices (Ford Baldner, Doll, & van Mersbergen, 2015; van Leer & van Mersbergen, 2017). However, differences fell within one scale level and created difficulty in using the scale in individual situations such as the clinic (Ford Baldner, Doll, & van Mersbergen, 2015; van Leer & van Mersbergen, 2017). The primary purpose of this research is to assess whether or not a precise effort scale with more delineation, the Borg Centimax scale, can reliably and validly capture the experience of vocal effort.

Considerations for any self-reported measure accounting for unrelated factors that might impact the output is necessary. Perceptual phenomena such as vocal effort is susceptible to change depending on external and internal effects from the individuals' environment (Bottalico, Graetzer, & Hunter, 2016; Solomon, 2008; van Mersbergen, Beckham, & Hunter, 2020; van Mersbergen & Delaney, 2014; van Mersbergen, Lyons, & Riegler, 2017; van Mersbergen & van Leer, 2016; van Mersbergen, Vinney, & Payne, 2019). However, the degree to which those factors change perceived effort is still unclear. The secondary purpose of this research is to determine the extent to which different vocal effort levels, the distance between those levels, or both are affected by one factor, transient mood.

Literature Review

Definition of Vocal Effort

Vocal effort, previously used interchangeably with other terms, suggests that a more specific definition be employed. In order to obtain an accurate definition of vocal effort and the potential ways in which to measure it, a consensus paper by Hunter et al (2020) reviewed the literature with the purpose of defining "vocal effort," "vocal fatigue," "vocal load," and "vocal loading." While analyzing the previous literature, the authors identified instances and intended uses for each term, discovering that many of these terms were substituted with one another in the absence of clear definitions. To address this issue, they proposed two new terms to replace vocal load and loading and standardized definitions for all four concepts (vocal demand, vocal demand response, vocal effort, and vocal fatigue) and suggested methods to quantify each term.

According to Hunter and colleagues (Hunter et. al, 2020), the proposed definition for vocal demand was the necessary output needed to accomplish a given communicative interaction. *Vocal demand* refers to the characteristics of the environment or length of time for a given communication event. Vocal demand is person independent, meaning there are no individual differences that factor into its measurement, such as personal traits. Specifically, the change in room acoustics or time requirements for speaking (teaching for one hour vs. teaching for three hours) is independent of the person talking and is a pure physical phenomenon. The authors suggested that this term be measured in the vocal pitch, loudness, quality, and duration requirements to satisfy the environmental demands. Conversely, *vocal demand response* is person dependent, meaning that individual traits and perceptions determine the physiological response to the vocal demand. One person may respond to a change in demand by increasing sub-glottic pressure to accomplish increased loudness while another person my respond by using increased medial compression (Titze, 2000). These accommodations depend on the person's individual characteristics and habits relating to verbal communication. While vocal demand response is measured using objective, physiological measures, it is important to note that these measures are influenced by the demand, the perception of that demand, and the person-specific differences in response to the demand.

Vocal fatigue is another definition in this consensus paper (Hunter et. al, 2020) which was described as the individual's specific perception of symptoms that influence vocal task performance over time. Vocal fatigue influences vocal performance after overloading the system over a period of time and can be observed as changes in performance ability and the perception of the performance. As fatigue increases, it is conceivable that an individual's sense of effort would simultaneously increase due to a reduction of central activation to the lower motor neuron pools of the peripheral nervous system (Solomon, 2008).

Vocal fatigue was then broken down into two different types of fatigue, *performance fatigue* and *perceived fatigue*. Performance fatigue was defined as the measurable outcome of changes in performance ability (as determined by the muscular contraction and adequate activation of the nervous system for a given task). Perceived fatigue was defined as the perception of fatigue the vocalist's experienced while regulating homeostasis and psychological states. In addition to the broad definition of fatigue, this consensus paper suggested that there are two types of fatigue observed in individuals: *state fatigue,* the change in perceived fatigue during an ongoing activity, and *trait fatigue*, the amount of fatigue one perceives as a result of physiological or psychological makeup. In either state or trait fatigue, the notion of perceived work over a period of time is the predominant distinction between fatigue and effort. Effort, on the other hand, is perceived work that is linked to a specific activity in a distinct, short, and immediately recent period of time.

Of the reviewed terms from the consensus paper (Hunter et. al, 2020), vocal effort appears to be the least perspicuous due the previous conflation of its definition. This consensus defined effort as perceived exertion of a speaker's vocal output. Before 1990, vocal effort had a fairly stable definition but has since undergone an explosion of interpretation. The two phrases most often used to define vocal effort had been "vocal loudness change" and "rise in fundamental frequency", both of which were used to describe the effects of increased voice use. To further obfuscate matters, the interchangeable use of vocal effort and vocal fatigue is ubiquitous. As fatigue increases, it is conceivable that an individual's sense of effort would simultaneously increase due to a reduction of central activation to the lower motor neuron pools of the peripheral nervous system (Solomon, 2008). Furthermore, in order to determine a change in perceived vocal fatigue over time, one must evaluate vocal effort. So, the measure of vocal effort must be used in order to track vocal fatigue over time, particularly the perceptual aspect of fatigue.

The lack of clarity in defining and measuring vocal effort seems to derive from previous conceptualizations that both physiological and perceptual phenomena are involved in effort (Hunter et. al, 2020). However, when vocal effort has been measured physiologically, personal factors such as height, weight, biological sex, respiratory fitness, and muscle composition of the vocal folds all contribute to vocal output. If an individual were larger and presented with greater respiratory fitness, then their perception of effort might be less for the same acoustic output. These personal factors make seemingly objective measures, less stable. Likewise, external factors such as ambient room noise, communication tasks (speaking to a group as opposed to one individual), communication content (good news vs. criticism), and distance would influence the acoustic demand of the speaker and therefore affect vocal effort. Previous perspectives for measuring vocal effort in the reviewed literature linked vocal effort with increased subglottal (tracheal) pressure (Eriksson & Traunmüller, 2002; Lagier, et al., 2010), a higher cervical muscle tension (Lagier, et al., 2010), increased fundamental frequency (Bottalico, 2017; Cheyne, Kalgaonkar, Clements, & Zurek, 2009; Eriksson & Traunmüller, 2002; Hazan, Tuomainen, & Pettinato, 2016; Primov-Fever, Lidor, Meckel, & Amir, 2013), first formant (Cheyne, Kalgaonkar, Clements, & Zurek, 2009; Eriksson & Traunmüller, 2002), and sound pressure levels (Bottalico, 2017; Eriksson & Traunmüller, 2002). However, while these factors influence an individual's perception of effort, they cannot be used to define or to measure of vocal effort.

Reflecting on the previous uses of the term "vocal effort", the authors of this consensus paper suggested that this term, be narrowed to a measure of the speaker's perceived exertion given a vocal task at a single point in time (Hunter et. al, 2020). Authors then conclude that effort should be treated as a perceptual phenomenon and measured solely using psychophysical procedures. In other words, measurement techniques should reflect the definition of perception and not physical or physiological factors. In addition, measurement practices should reflect this perception and account for parameters unrelated to voicing that may affect this perception.

Measurement of Vocal Effort

Historically, measuring vocal effort has been a difficult task given the concept had been fluid and poorly defined. Due to a lack of a clear definition, previous research has relied on a collection of tools within the field of voice as well as measures for the study of physical exertion to capture the construct of vocal effort. Self-reporting measures to capture perceived physical exertion in contexts outside of vocal effort, such as direct magnitude estimation and visual analog scales have been considered. Direct magnitude estimation is a form of ratio scaling in which subjects make direct numerical estimations of the perceived intensity, or magnitude, of the stimuli. (Baldner, Doll, & van Mersbergen, 2015). The ratio scale in particular is beneficial for measuring effort because it includes an absolute zero, which means you can construct a fraction with a ratio variable. Having the ability to compare the relative values of two numbers against a basepoint allows for comparisons within an individual's ratings of effort. For example, if someone experiences a rating of 10 in the first task and 20 in the second task than we can say they experienced twice as much effort in the second task. However, even though we obtain a ratio, we cannot determine if the scaling is linear or logarithmic. This means that we cannot guarantee that a rating of 15 would be a half-way point between 20 and 10. Since perceptions tend to be linear while their physical correlates are logarithmic, it is necessary for a scale to account for this relationship.

Another quality of direct magnitude estimation is that it requires a measurable physical quantity with which to manipulate and compare to the perception of effort.

While other perceptual constructs have clear physical correlates, such as sound pressure level for loudness, the physical parameters of vocal effort are unclear (e.g., muscular exertion, cognitive exertion, or both). Furthermore, effort is multifactorial in nature and without a clear understanding of the physical correlates of effort, we cannot obtain a measure of effort using magnitude estimation.

Visual analog scales (VAS) are used to capture characteristic values that span across a continuum rather than discrete levels (Crichton, 2001; Baldner, Doll, & van Mersbergen, 2015). These scales are often anchored, meaning they have word descriptions on each end to help define the endpoints which allows individuals to easily understand the scale. While there are recognizable benefits to visual analog scales, such as its ability to allow individuals to respond without constraint to an imposed structure of divided responses or reference any principal or physiological value, there are also drawbacks to using this tool for vocal effort measurement. First, since this scale is ordinal, meaning the variables are classified by order of quantity with no definite understanding of the magnitude of responses, leaving assumptions about the discrete levels of vocal effort tenuous. In addition, VAS do not presume that there are equal values between equal distances on the scale. This lack of measurable values causes poor inter- and intra-rater reliability because it only captures whether one rating is higher than another within an individual at that time point. Second, VAS do not possess a meaningful zero for ratio comparisons. A meaningful zero is when the zero point on a scale represents the absence of a construct and is necessary when measuring effort in order to allow for a person to express the perceived absence of effort. Third, VAS lack assumptions about the polarity of responses. For example, some VASs use one end of the scale to represent a low value of a construct and the other end to represent a high value (e.g. low effort vs. high effort). Other VASs use one end of the scale to represent a high level of a construct and the other end to represent a high level of an opposite construct with the middle of the scale representing a neutral construct (e.g. high unpleasant vs. high pleasant).

Recognizing the need for a perceptual scaling measure to capture complex activities, Gunnar Borg (Borg, 1982) created a category scale that had the positive attributes of both the direct magnitude estimation and visual analog scales. This scale was also structured with non-linear organization to reflect the non-linear relation between physical stimuli and perception. The scale values of this measurement corresponded with anchors, or expressions that could be understood by the layperson similar to VAS. Additionally, this scale was designed for phenomena with multiple physical factors such as physical work (i.e. when lifting a heavy box). This category ratio scale was later adapted by multiple voice researchers to fit the needs of vocal effort research (Anand, Bottalico, & Gray, 2019; Ford Baldner, Doll, & van Mersbergen, 2015; Steinhauer, Grayhack, Smiley-Oyen, Shaiman, & Mcneil, 2004; van Leer & van Mersbergen, 2017; van Mersbergen & Delany, 2014; van Mersbergen, Vinney, & Payne, 2019, van Mersbergen, Patrick, & Glaze, 2008).

Borg CR Scales and Measuring Vocal Effort. In most cases, perceptions tend to be linear while their physical correlates are logarithmic (Borg, 1982). To account for this, the Borg Category Ratio (CR) scale incorporates findings that direct magnitude estimation scale would capture by reflecting a logarithmic relationship with perception. The Borg CR scales also possess aspects of the visual analogue scale’s ease of use.

Another characteristic of this CR scale, borrowed from the VAS, is the inclusion of predetermined anchor levels with expressions such as “minimal”, “slight”, and “maximum” vocal effort. These anchors are an important characteristic of the scale to ensure that the individual has a clearer idea as to how to rate their perceived effort.

With Borg’s method, participants are asked to rate presented stimuli by assigning numbers to each. These numbers correspond with how they perceived a given stimulus in proportion to the physical properties of the stimuli. For example, Borg (1990) discussed that when a subject who is driving 100 km/h is asked to decrease their speed by half, they tend to decrease to about 70 km/h. This process is repeated with different speed levels to understand the relation of the physical output compared to the perceived output and plotted them in Stimulus/Response curves. According to Borg, perceptual responses are key indicators of the degree of physical strain. As discussed by Hunter and colleagues (2020), Borg suggests that other aspects of activity are necessary for a complete picture of physical strain including performance conditions and accuracy, physiological output, and perceptual responses (Borg, 1990). These additional factors are included in the Hunter paper in the form of vocal demand, demand response, and effort. Borg’s outlined effort variables are comparable to the terms presented in the consensus article. Though Borg did not define a clear parallel to vocal demand, it was often referred as the construct of work requirements. The presented construct included examples of physical work such as lifting and carrying heavy weights and running (Borg, 1982; Borg, 1990).

Borg’s Relation to Other Scales. Although the Borg CR 10 scales capture differences in vocal effort between voice patients and healthy controls (Baldner, Doll, &

van Mersbergen, 2015), pre- and post-therapy (van Leer & van Mersbergen, 2017), and negative and positive mood, (van Mersbergen & Delany, 2014; van Mersbergen, Lyons, & Rielger, 2017), differences in these populations all fell within one scaling point. Although participants were free to split each scale level into sub-levels (Borg, 2007) participants and patients often gravitated to main scale points. For example, in a study conducted by Ford Baldner, Doll, and van Mersbergen (2015), an attempt was made to see if the Borg CR10 could accurately separate those with voice disorders from those without voice disorders. Although the scale showed a significant difference between the two groups, the distance between them fell within one scaling factor (1.3 compared with 1.7). Since the nature of the Borg CR10 makes it less likely for an individual to answer between scaling points, for example 1.3, using the Borg CR 10 for individuals may be impractical.

The Borg CR10 has been shown to correlate moderately well with other measures to capture voice disorder severity within an individual. In past research, convergent validity for the Borg CR 10 has correlated this tool with the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) (Baldner, Doll, & van Mersbergen, 2015), Vocal Fatigue Index (Anand, Bottalico, & Gray, 2019), and Voice Handicap Index (van Leer & van Mersbergen, 2017). While the CAPE-V can be used to obtain the severity of perceptual aspects of a voice-related issue, it cannot be exclusively used when determining the nature of a voice disorder since data is collected through the clinician's perspective and is not a self-reported measure.

Another measure, that correlates well with the Borg CR10 is the Voice Handicap Index (VHI) which is a self-reported measure. van Leer and van Mersbergen (2017) found that item 14 in the VHI correlated well with the Borg CR 10. This instrument was designed to measure the impact a voice disorder has on an individual's quality of life and researchers have determined the results of Item 14 ("I feel as though I have strain to produce voice,") to be a decent marker of an individual's perception of effort. However, there are differences between perceived handicap of a voice problem and a perceived effort of a vocalization. A handicap would indicate a circumstance that creates difficulty completing or advancing in a task, while a perception of effort would mean that a task in that moment in time was introspectively judged to be taxing.

The Vocal Fatigue Index has been determined to be a valid and reliable resource measure of vocal fatigue in individuals and has been used in conjunction with the Borg CR10 (Nanjundeswaran, Jacobson, Gartner-Schmidt, & Abbott, 2015; Anand, Bottalico, & Gray, 2019). These studies found a relationship between Borg CR 10 and VFI score. As stated earlier vocal effort and vocal fatigue are related because perceived vocal fatigue is measured by tracking individual instances of vocal effort over time. It stands to reason that a self-report measure of vocal fatigue such as the VFI would correlate well with the Borg CR 10.

External Influences on Effort

Because effort is inherently multifactorial there are many influences that will change the way it is perceived. Some influences are external, such as the communication environment, whereas others are internal, such as psychological or physiological processes. External influences on effort can include vocal demands such as the length of time one vocalizes or the noise level of the speaking environment. Since vocal effort is a speaker's perceived exertion of a response to a vocal demand, when the vocal demand becomes more taxing and the speaker changes vocal technique accordingly, the perceived effort should reflect such changes. Thus, it is essential to consider the potential vocal demands that may contribute to one's perception of effort.

The duration of time spent talking within the day is highly variable as observed among different professions. These durations can contribute to the experience of effort. For example, a person who is required to engage in lengthy voice use, such as a teacher, will vocally respond to this greater demand than someone who only needs to speak minimally throughout the workday. A study by Titze and colleagues (2007) found that the voicing time spent by teachers accounted for 23% of their workday with approximately 1800 occurrences of voicing per hour at work. The increase of time required to speak then increases the physiological actions taken to meet that demand, which could in turn lead to an increase in vocal effort if these vocalizations are perceived as strenuous or fatiguing.

The additional noise with which one projects over in any given environment can also contribute the perception of effort. For example, a teacher speaking over noise as observed in a talkative classroom, may respond to the noise by increasing their intensity using medial compression and increased subglottal pressure, which results in higher collision forces of the vocal folds. These collision forces may result in edema that inhibits efficient cover mechanics (Titze, 2000), which might be perceived as more work. In contrast, a teacher with amplification would not need to work as hard to deliver a message (Jónsdottir, Rantala, Laukkanen, & Vilkman, 2001; Sapienza, Crandell, & Curtis, 1999). Conversely, librarians who usually speak in quiet environments require less sound pressure than those who have to communicate in a construction zone and therefore may not have to recruit the same physiological responses and expose the vocal folds to unwarranted impairment.

In addition to physical environments that influence vocal demand response and subsequent vocal effort, communication content also affects physiological responses. When an individual has to communicate subject matter that is emotional or requires high-level thinking, they might experience increased vocal effort because of the responses to these. For example, a healthcare worker in oncology, who may have to communicate bad news as compared to a healthcare worker in labor and delivery, who may offer congratulations. Additionally, some communicative content can also be more cognitively taxing. For example, an astronaut communicating to ensure safe space travel may experience increased vocal effort compared to a kindergarten teacher assisting a child in a classroom activity. Hence, our perceptions of effort are affected by both physical parameters (e.g. noise level) and contextual parameters (e.g. communication content).

Furthermore, embedded in communication context is also our communication partners. Less vocal effort may be experienced by an individual speaking with a peer or family member compared to a boss or hiring manager. So, not only where we speak and what we say, but also with whom we speak to will affect our vocalizations and perceptions of effort.

Internal Influences on Effort

Internal effects are the individual differences in physical, physiological, and psychological characteristics. Physical traits such as height, weight, and biological sex contribute to relatively stable internal influences on perceived vocal effort. For instance, a male may experience less vocal effort than a female due to the increased cushioning effect of the hyaluronic acid in the superficial layer of lamina propria. Conversely, intermittent factors such as affect and cognitive abilities can also influence an individual's perception. Among the many internal factors that have the potential to affect physiological response and the subsequent perception of those responses, the two factors of interest for this study are emotional and cognitive responses.

Cognitive Influences on the Perception of Effort. Cognitive vocal tasks can be an external factor in that in some situations one must convey difficult ideas independent of an individual's cognitive state or ability. However, cognitive ability can also be an internal factor, particularly when an individual has depleted their cognitive storage. Vinney, van Mersbergen, Connor, and Turkstra (2015) observed this when they investigated self-regulation's effect on vocal task performance in spontaneous speaking conditions. They employed various levels of self-regulatory vocal tasks (low, high, and high with an intervention) in order to determine if self-regulation affected the way someone responded to a task. The results of this study suggested self-regulatory depletion may have some effect on vocal behavior modification as exhibited by those who had little self-regulatory depletion suppressing the Lombard effect to a greater extent compared to the other groups whose self-regulation was depleted through complex tasks. In this study they also measured vocal effort, which reflected that those low in self-regulatory depletion experienced vocal effort less.

In a follow up study, van Mersbergen, Vinney, and Payne (2019) found an individual's ability to utilize cognitive resources to tackle challenging mental tasks determines the level of vocal effort one experienced. Mental effort, similar to vocal effort, is a perceptual concept – the perception of how hard someone thought about something. Unfortunately, it can be difficult to differentiate the two. From their findings, it appeared that vocal effort ratings were higher during the cognitively taxing condition even when engaged in relatively soft phonation that should not be vocally effortful. Conversely, vocal effort was rated significantly lower during easier cognitive tasks even during loud vocalizations. This suggests that vocal effort may be influenced by mental effort. The authors state that perceptions of vocal effort appear to mirror the ratings of mental effort during tasks for which vocal activity is relatively stable but cognitive demands fluctuate. As cognitive effort increases the strengths of the association between vocal effort and mental effort also increases. The authors suggest that decreasing cognitive effort in voice treatment may aid to a client's success by providing the opportunity to better conceptualize differences in vocal sensations and cognitive load when demonstrating a new vocal technique.

Emotional Influences on the Perception of Effort. Emotional content can be an external, independent aspect to communication as in when one must deliver sad news. However, emotional states are also an intrinsic factor influencing effort perception. Emotion has had a noticeably significant, yet still unclear, impact on the perception of vocal effort. Emotional content affects the way we voice (van Mersbergen & Delaney, 2014; van Mersbergen, Lyons, & Reigler, 2017) by increasing the contact quotient, or amount of time the vocal folds stay closed, which will in turn cause greater collision forces and edema resulting in more work and perception of effort.

At its root, emotions are a perceptual, observer-dependent characteristic in the mental life of a human. Emotions are defined from a collectively agreed upon complex of interoceptive sensations associated with the meaning and representation of emotional states. Thus, emotional states represent a learned physiological state associated with specific events that motivate cognitive and behavioral responses (Feldman Barret, 2017; Davidson, 2000). Across individuals, behavioral outputs come from a wide range of emotional inputs that are formed by three sources of stimulation: sensory stimulation from the body, sensory stimulation from the world, and background knowledge/prior experience (Barrett, 2009).

In order to address emotional responses to the perception of effort, temperament, or our unique way to responding to emotional stimuli, must be considered. Temperament is our hardwired tendency to experience and express emotion. The theoretical framework of temperament captured in the Big Three model of personality (Clark & Watson, 1999) include constructs of Extroversion/Positive Emotionality (E/PE), Neuroticism/Negative Emotionality (N/NE), and Psychoticism/Behavioral Constraint (P/BC). In summary, Clark and Watson state that all three factors are orthogonal, and one can have high or low levels of each trait simultaneously. E/PE represents the tendency for an individual to move towards reward and is generated from neurobiological systems underpinning appetitive action. N/NE represents the tendency for an individual to move away from punishment and is generated from systems of aversion. P/BC is the tendency to behaviorally act upon E/PE or N/NE systems.

Temperament has been investigated in the voice literature because of the clinical experience that emotions can affect vocalizations and, in some cases, lead to voice disorders. The Trait Theory of Voice Disorders (Roy, Bless, & Heisey, 2000) has drawn upon a rich history of research suggesting that certain personality disorders are temperamentally motivated. For example, those diagnosed with functional dysphonia have increased N/NE and decreased E/PE and P/BC compared to healthy controls (Roy & Bless, 2000; Roy, Bless, & Heisey, 2000a; Roy, Bless, & Heisey, 2000b; van Mersbergen, Patrick, & Glaze, 2008) Conversely, those with vocal fold nodules presented with increased N/NE and as well as increased E/PE compared to healthy controls (Roy & Bless, 2000; Roy, Bless, & Heisey, 2000a; Roy, Bless, & Heisey, 2000b). Additionally, there is a strong link between the Big Three and vocal demand. Those who are extroverted are more likely to go into professions that involve more talking and partake in activities that may require the use of vocal projection in a noisy environment, creating a larger vocal demand in this group of individuals (van Mersbergen, 2011).

In a follow-up study empirically testing the Trait Theory of Voice Disorders, van Mersbergen and colleagues (2008) investigated physiological measures sensitive to mood, and subsequently, temperament. They found evidence of how temperament may lead to voice disorders. They also found that these transient moods also effected vocal effort using the Borg CR 10. However, they did not employ physiological measures of voicing behavior and so, understanding how emotions and temperament affect vocalization was still unclear.

In a series of follow up studies, van Mersbergen and colleagues (2013; 2015; 2018), found that the transient states of mood, which are temperamentally motivated, affected certain physiological measures. Emotional stimuli in previous research on emotions effect on the voice (van Mersbergen & Delaney, 2014; van Mersbergen, Lyons, & Riegler, 2017; van Mersbergen & Lanza, 2018) contained pictures that varied in valence and arousal. Valence is the positive, negative, or neutral affectivity experienced by an individual, whereas arousal measures how calming or exciting the information presented is. The delivery of affective stimuli came from the International Affective Picture System, a corpus of over 1,200 photographs that contain pictures of negative, positive, and neutral content that vary in levels of arousal. van Mersbergen employed these pictures as stimuli to elicit transient emotions while participants performed various vocal tasks.

The first study set out to determine if Electroglottography (EEG) contact quotient is a practical measure to discern vocal changes in response to emotional stimuli (van Mersbergen & Delaney, 2014). Results indicated while there was an increase in EEG contact quotient when the participants were exposed to negative stimuli, the associated effect size was quite small. This led to a follow-up study that increased arousal along with affect stimuli to see if there would be a subsequent increase in separation between the negative and neutral EEG contact quotient ratings (van Mersbergen et al., 2015). Although EEG contact quotient increased with arousal, the results of this study still demonstrated a small effect size between conditions. Additional studies using picture stimuli also found relative fundamental frequency also varied with mood states. Positive conditions elicited more mitigated voicing patterns. From these experiments, it is clear that emotional states effect vocal output, which in turn, could contribute to a change in vocal effort.

Purpose and Reason for the Current Research

Based on previous research on vocal effort, Berardi and colleagues (2021) used a computer adaptation of the Borg Centimax to measure vocal effort in terms of intensity in a direct magnitude production paradigm where participants were asked to vocalize at a given vocal effort level (Berardi, van Mersbergen, & Hunter, 2021). They used the Borg Centimax scale to overcome the limitations of the Borg CR 10 where participants gravitated to anchor points. The Borg Centimax scale removes this barrier by introducing visible sub-levels that one can choose, reducing that bias toward main anchor points (Borg & Borg, 2002). In addition, the authors employed a direct magnitude scaling method to capture a Stimulus/Response curve as is typical in psychophysical research. This direct magnitude production method required participant to produce a vocalization at a given vocal effort level. By having participants produce a desired vocalization based on an anticipated vocal effort level the authors could assess various vocal parameters that might contribute to a sense of effort.

There are two aims of this current research. The first aim is to contribute to the reliability testing of this method of effort measurement by repeating the basic procedures across participants and settings as well as within a participant as observed in previous research (Berardi et al, 2021). The questions arising from these aims include whether intensity, as measured in decibels, is a linearization of a logarithmic property as related to produced vocal effort levels. Based on the results of previous research, it is hypothesized that vocal intensity patterns will be repeatable and distinct for each vocal effort level within a participant.

Since vocal effort has been shown to be susceptible to change from both external and internal factors, the second aim of this research was to capture changes to vocal effort levels when emotional stimuli are presented. Therefore, the secondary purpose of this study is twofold; 1) does transient mood change the intensities across all levels of vocal effort and 2) does transient mood affect the relative distance between each level? It is predicted that transient mood will have a noticeable and significant effect on both the intensity level, by increasing intensity level in negative moods, and the intensity distance between effort levels, by separating intensity levels for each vocal effort level elicited for all moods.

Methods

Participants

Participants for this study consisted of students from the University of Memphis CSD program and members from the surrounding Memphis community. Recruitment for this study was completed via word-of-mouth and flyers circulated through social media platforms and email distribution lists. A total of 27 participants (22 biologically determined females and 5 biologically determined males) between the ages of 21 and 69 years of age, (M = 30.7, SD = 12.5) were included in this study. Inclusionary criteria consisted of participants being mentally and vocally healthy as determined by scores greater than 24 on the Mini-Mental State Examination (MMSE) and below a 30 on the Voice Handicap Index (VHI), no prior voice training per self-report, and English proficiency in both speaking and reading per observation. Exclusionary criteria included a current or previous diagnosis of a neurological impairment, a history of head and neck cancer, a hearing impairment (as determined by a hearing screening,) and a history of moderately severe to profound mood disorders.

Measures

Independent Measures (Stimuli)

Spoken stimuli. To address the first question of this study and attempt to partially replicate a previous study of interest, a set of prompts elicited three separate voice task categories increasing in cognitive demands (automatic speech, reading task, and spontaneous speech) using the PsychoPy software employed in said previous study. Three variations of each task were used during this study. The automatic speech stimuli consisted of either stating the alphabet (English), counting to twenty-five, or saying the names for the days of the week and months of the year. The standard sentence reading task variations were all excerpts from standard speech acoustic reading passages which included the Marvin Williams Passage (Svec, Titze, & Popolo, 2005), the Rainbow Passage (Fairbanks, 1960), and the Stella Passage (Weinberger & Kunath, 2011). The final speech stimuli showed the participant a map of the Portland, Oregon subway system and requested a verbal description of one of three specific routes. The task used the carrier phrase "describe how to get from A to B via C" where "A", "B", and "C" were specific points on the map. Each task was completed at each of the four vocal effort levels (2-minimal, 13-slight, 25-moderate, and 50-severe vocal effort) in a randomized fashion for a total of 36 trials in the first portion of the experiment. To compare normal conditions with emotion conditions, only the rote tasks were analyzed for the current research. The output for the additional tasks (reading and map tasks) is beyond the scope of this paper and will be investigated at a later date.

Mood Induction Stimuli. The second portion of this experiment addressed the second purpose for this study, capturing the effects of emotion on vocal effort. The procedures for the automatic speech tasks from the first experiment were repeated with the addition of emotion stimuli. Emotion stimuli consisted of 60 images of negative, positive, and neutral content (20 each) from the International Affective Picture System (IAPS) (Lang, Bradley, & Cuthbert, 2008). Participants were instructed to complete the automatic tasks at a particular vocal effort level after progressing to the next slide which presented one of the IAPS pictures. All pictures were presented three times in a quasi-random order to obtain an additional 36 trials of emotion-driven tasks.

Vocal Effort Stimuli. The Borg Centimax, also known as the Borg CR100 (Borg &

Borg, 2002), is a self-reporting measure used to capture an individual's perception of effort. This scale ranges from 0 to 100 with verbal descriptors placed accordingly to provide ratio data comparable to the obtained dependent variable. A scale level of 0 indicates "no vocal effort", 50 indicates "severe effort", and 100 represents "maximum effort." For this study, participants were introduced to the scale using anecdotal anchors for the lowest and highest points using the following script: *"A vocal effort level of 1 would be quietly talking to someone next to me at home and A vocal effort level of 100 would be trying to shout at someone while standing on an airport runway."* To maintain consistency to partially replicate a previous study, we used a direct magnitude production research paradigm which requires participants to respond to a given stimulus level using the Borg Centimax. This paradigm had participants producing a vocal output when given a vocal effort level.

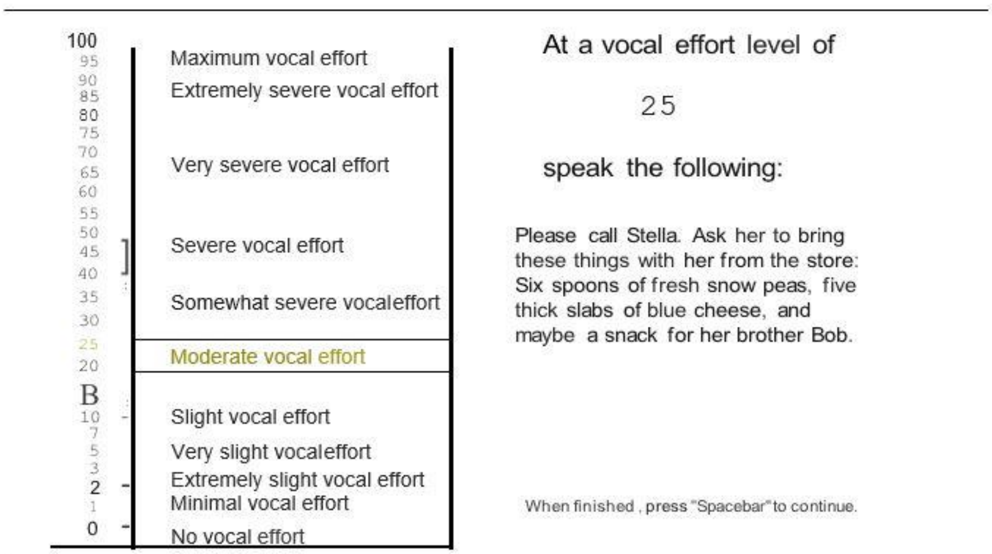

Figure 1 Example of the presentation of the vocal effort scale (left), target vocal effort level (upper- right) and speech stimulus (right) during an experimental trial.

Dependent Variables

Primary Variable of Interest

Intensity. The dependent variable of interest for this study was intensity measured in dB. Intensity measures were collected using a two-step calibration process. First, a Berhinger microphone (Behringer ECM8000 Omnidirectional Measurement Condenser Microphone, Stone Mountain, GA) was placed in a sound treated room 50cm away from participant's mouth and was calibrated to 94 dB SPL (relative to 20 µPa). Second, a head mounted microphone with a mic to mouth distance of 10 cm, 45 degrees from center of the mouth was placed and the participants was asked to produce a steady vowel on /a/ three separate times. Both the reference microphone and the head mounted microphone were connected to XLR inputs on a digital hand-held recorder (ZOOM H6, Hauppauge, NY) on tracks 3 and 4 respectively. The data collected was compared in post-hoc analysis to obtain the intensity data.

Manipulation Variables

Self-Assessment Manikin. The Self-Assessment Manikin (SAM) rating protocol measures levels of affect (excitement/calm, SAM-A) and valence (pleasantness/unpleasantness, SAM-V) along a nine-point Likert-type scale (Bradley & Lang, 1994). This measure provides the current study with the ratings for the International Affective Picture System (IAPS) in accordance with the participants' perspectives.

PANAS. The Positive Affect Negative Affective Scale (PANAS) is a widely used tool to assess mood and emotion. The brief scale was implemented for the current study consisting of 20 items divided equally for measuring between positive affect (ex. Excited, enthusiastic) and negative affect (ex. Nervous, upset). Each descriptive feeling is rated on a five-point Likert scale ranging from 1 (*very little or not at all*) to 5 (*extremely*) (Watson, Clark, & Tellegen, 1988). This scale measures the degree to which the affect has been experienced in a given time frame. For the current study, the PANAS was implemented 4 separate times; at the initial intake stage and after each of the three experimental blocks to assess for overall mood changes throughout the experimental process.

Training and Procedures

Following consent, participants received a hearing screening and cognitive screening (MMSE) (Tombaugh, McDowell, Kristjansson, & Hubley, 1996), then completed several questionnaires including an intake form, the voice handicap index (Jacobson et al., 1997), and the PANAS baseline (Watson, Clark, & Tellegen, 1988). After the initial intake process was complete, the participants were then brought into a sound-isolation booth to be set up with the head-mounted microphone and execute calibration procedures (Berardi, van Mersbergen, & Hunter, 2021). Upon completing the calibration process and setting up, participants underwent training through a video training script that employed artificial speech used in the previous research to eliminate bias and maintain consistency between participants and studies (Berardi, van Mersbergen, & Hunter, 2021). The training task was twofold; it trained the participants in the procedure and controlled for learning effects. This tutorial required the participant to practice each of the automatic speech tasks.

As part of the training, participants were introduced to the vocal effort scale. This scale was presented with anecdotal anchors for the extreme values of 1 and 100 as follows: “a vocal effort level of 1 would be quietly talking to someone next to me at home” and “a vocal effort level of 100 would be trying to shout at someone while standing on an airport runway.”

Each trial began by instructing the participants to speak a particular speech stimulus at specific vocal effort level. In order to replicate the previous study (Berardi, van Mersbergen, &

Hunter, 2021), the four vocal effort levels that were prompted included: (1) 2 or "Minimal vocal effort", (2) 13 or "slight vocal effort", (3) 25 or "moderate vocal effort", and (4) 50 or "Severe vocal effort." Following each prompt, the participants completed each speech stimulus at each vocal effort level (four) for a total of twelve trials. For each trial, the participant was shown the vocal effort scale and speech stimulus (Fig 1) to serve as a visual reminder. Stimuli variations were randomized between participants.

Block 1. The first part of the experiment included additional conditions for a total of 36 trials. The entire first portion consisted of three separate speech tasks with three stimuli per task (rote, reading, and spontaneous speech) executed at the four vocal effort levels. The reading and spontaneous speech tasks were included in the study to vary the tasks. However, the tasks of interest for this current study were the rote tasks. Following this experimental block, the participant completed another round of the PANAS (PANAS 1.)

Block 2. The second portion of the experiment was designed to assess responses to emotional stimuli. This was achieved by using the same rote speech tasks with the addition of the emotion stimuli. Participants went through an additional training phase to acquaint themselves to the procedure of the second experiment. This portion required the participants to say the phrases at the designated vocal effort level while looking at pictures on the screen. Participants were provided the task phrase and vocal effort level and instructed to proceed to the next screen before speaking the task. To facilitate a more automatic transition between screens and to prevent participants from forgetting what they had to say once they looked at the picture, they were instructed to rehearse which tasks and vocal effort level in their mind before advancing to the next screen. Following each trial, the participants were also directed to complete the Self-Assessment Manikin rating protocol (SAM) to verify the presence of an emotional response to the pictures and complete a third PANAS (PANAS 2.)

Block 3. The final portion of the experiment consisted of a replication of the first portion to serve as a condition that assessed for fatigue effects. This consisted of another set of twelve randomized trials and acted as a comparison to ensure reliability within the study. This block concluded with one final completion of the PANAS (PANAS 3.)

Instrumentation

As conducted in the previous study (Berardi, van Mersbergen, & Hunter, 2021), all participants were recorded using a head-mounted microphone (AKG, Model C520, Vienna, Austria) placed 10cm from the participant's mouth using a digital hand-held recorder (ZOOM H6, Hauppauge, NY) in a sound-isolation booth. Additionally, a calibrated reference microphone (Behringer ECM8000 Omnidirectional Measurement Condenser Microphone, Stone Mountain, GA) was placed 50 cm from the participant's mouth. A Dell laptop (Inspiron 7000 2in1; Dell, Inc. Round Rock, TX) using the open-source python-based program, PsychoPy (Peirce et al., 2019) was used to present experimental training and stimuli. A schematic of the instrumentation used can be seen in Fig. 2.

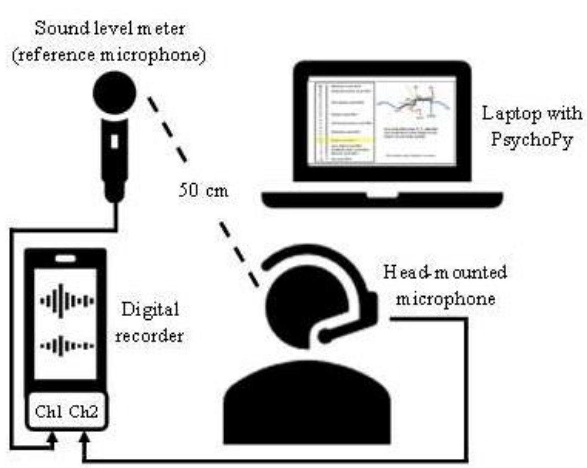

Figure 2 Schematic for the instrumentation (Berardi, van Mersbergen, & Hunter, 2021)

Analysis

A single factor ANOVA comparing reported mood on the SAM V between aversive, neutral, and positive conditions verified that emotion stimuli elicited the appropriate mood during emotion induction in Block 2. Additionally, a single factor ANOVA comparing reported arousal on the SAM A verified appropriate physiological engagement during mood induction. Finally, a single factor ANOVA comparing overall mood reported on the PANAS between each block tracked any changes in mood among Blocks 1-3.

To answer the first question investigating the repeatability of this method of capturing vocal effort, two analyses were performed. The first included comparisons of vocal effort levels (VELs) among levels at 2, 13, 25, and 50 during the first block using a single factor ANOVA to assess if the differences in VEL were significant from each other and evenly spaced between each level as was observed in previous literature (Berardi et al., 2021). Post hoc *t* tests comparing intensity differences between individual VELs described the nature of this association. This analysis confirmed repeatability across participants and settings. We hypothesized that they will not be different.

The second analysis for the first question included a two-way ANOVA with levels, Block (1, 3) and VEL (2, 13, 25, 50), to assess if there were differences between the first Block and the last Block. Post hoc *t* tests comparing specific VEL intensity differences between blocks determined at which VELs a difference occurs. We hypothesized that these will not be significantly different from each other and confirm repeatability within participants and setting.

To answer the second question that explored if there were differences in intensity levels of produced VELs during mood induction a two-way ANOVA with levels mood (baseline, aversive, neutral, positive) and VEL (2, 13, 25, 50) was performed. We hypothesized that there will be differences in intensity for produced VELs between aversive, neutral, and positive conditions given that past research found differences in perceived VEL (Baldner, Doll, & van Mersbergen, 2015, van Mersbergen et al., 2008; van Mersbergen & Delaney, 2014). Post hoc *t* tests assessed specific patterns of intensity differences among mood conditions.

Additionally, to answer the question if the distances between each VEL differed with mood induction, differences between adjacent VELs were compared across emotion condition in a two-way ANOVA with levels differences (VEL 50 - VEL 25; VEL 25 - VEL 13, VEL 13 - VEL 2) and mood (baseline, aversive, neutral, positive) was conducted. Post hoc *t* tests assessed specific patterns of distance between VEL among mood conditions.

Because of known intensity differences between biologically determined males and females (Bottalico et al., 2016) analysis addressing each hypothesis was performed on each group separately.

Results

Self-Assessment Manakin

Mood induction appeared to reflect intended directions. Participants reported significant differences in valence between aversive ($M = 1.97$, $SD = 1.30$), neutral ($M = 5.61$, $SD = 1.40$), and positive ($M = 6.78$, $SD = 1.54$) picture viewing, $F(2) = 861.7$, $p < .001$. Post hoc analysis revealed that mood was rated more unpleasant during viewing of aversive pictures than neutral, $t(550) = 1.96$, $p < .001$, or positive $t(550) = 1.96$, $p < .001$, pictures and that positive pictures induced a more pleasant mood than neutral, $t(550) = 1.96$, $p < .001$, pictures.

Participants also reported significant differences in arousal between aversive (*M* = 5.10, *SD* = 2.33), neutral (*M* = 2.48, *SD* = 1.81), and positive (*M* = 3.35, *SD* = 2.18) picture viewing, $F(2) = 109.2$, $p < .001$. Post hoc analysis revealed that arousal was rated more higher during viewing of aversive pictures than neutral, $t(550) = 1.96$, $p < .001$, or positive, $t(550) = 1.96$, $p < .001$, pictures and that positive pictures induced more arousal than neutral, t(550) = 1.96, $p < .001$, pictures. This confirmed that aversive pictures elicited greater arousal than positive pictures, suggesting that participants in this research found positive pictures less physiologically arousing overall. Figure 3 displays both mood and arousal for aversive, neutral, and positive picture viewing.

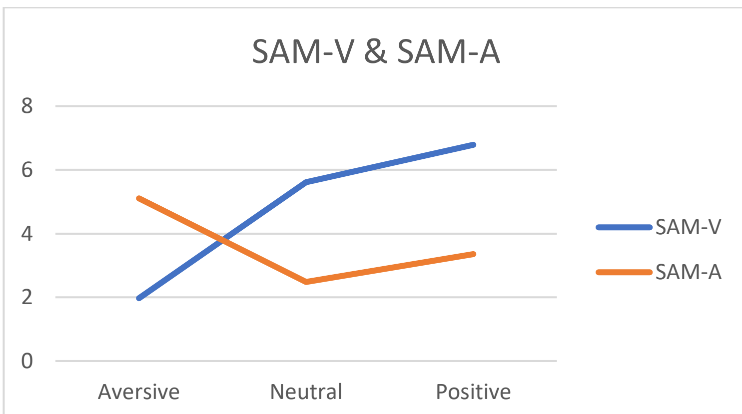

Figure 3 SAM-V & SAM-A Verification

Positive Affect Negative Affect Scale

Overall mood changed across blocks, $F(3) = 5.73$, $p = .001$. Overall mood appeared to be lower following experimental trails than at baseline (*M* = 3.63, *SD* = .25). Post hoc analysis revealed that Blocks 1 (*M* = 3.42, *SD* = .49) and 3 (*M* = 3.38, *SD* = .53) did not statistically differ, $t(42) = 2.01$, $p = .39$, in reported mood. However, participants reported more overall negative mood following Block 2 (*M* = 3.07, *SD* = .47) compared to Block 1, $t(42) = 2.01$, $p = .01$, and Block 3, $t(42) = 2.01$, $p = .02$. Figure 4 displays significance between PANAS scoring.

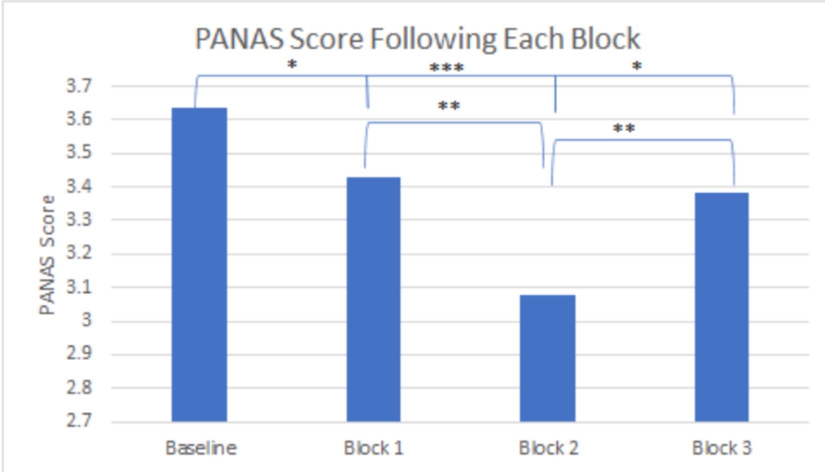

Figure 4 Significance between PANAS scores

Repeatability across Participants and Setting

Women. Intensity differences were observed across all VELs, *F*(68, 2) = 3.76, *p* = .01, which confirmed our first hypothesis that there would be difference in intensity between each VEL, which partially confirmed replicability across participants. Post hoc analysis revealed that women produced elicited VEL in a progression of increasing intensity at VEL 2 (*M* = 58.30, *SD* = 9.83), to VEL 13 (*M* = 63.19, *SD* = 10.06), to VEL 25 (*M* = 65.33, *SD* = 10.24), and finally VEL 50 (*M* = 69.49, *SD* = 10.60).

Post hoc *t* test revealed statistical differences between nonadjacent levels. So, for VEL 2, there were significant differences between VEL25 (*t*(34) = -2.10, *p* = .043) and VEL 50 (*t*(34) = -3.28, *p* = .002), but not VEL 13 (*t*(34) = -1.47, *p* = 149). Additionally, there were significant differences between VEL 13 and VEC 50 (*t*(34) = -1.83, *p* = .038) but not between VEL 13 and VEL and 25 (*t*(34) = -0.63, *p* = .266) or VEL 25 and VEL 50 (*t*(34) = -1.20, *p* = 120).

Average intensity difference between each level was 3.73 dB with the largest difference between VEL 2 and VEL 13 (4.89 dB) and the smallest difference between VEL 13 and VEL 25 (2.14). The difference between VEL 25 and VEL 50 (4.16 dB) was similar to the difference between VEL 2 and VEL 13.

Men. Similar findings occurred for men. Intensity differences across VELs, $F(16, 3) = 5.56$, $p = .008$, were significantly different and least significant differences also revealed a progression of increasing intensity at VEL 2 ($M = 61.69$, $SD = 5.95$), to VEL 13 ($M = 65.71$, $SD = 5.65$), to VEL 25 ($M = 68.96$, $SD = 3.91$), and finally VEL 50 ($M = 74.17$, $SD = 4.16$). Least significant differences were also observed between nonadjacent VEL levels.

Post hoc *t* test revealed statistical differences between nonadjacent levels. So, for VEL 2, there were significant differences between VEL25 ($t(7) = -2.28$, $p = .028$) and VEL 50 ($t(34) = -3.84$, $p = .003$), but not VEL 13 ($t(8) = -1.10$, $p = .152$). Additionally, there were significant differences between VEL 13 and VEC 50 ($t(7) = -2.70$, $p = .015$) and VEL 50 ($t(8) = -2.04$, $p = .038$), but not between VEL 13 and VEL and 25 ($t(7) = -1.05$, $p = .163$) or VEL 25.

Average difference between each level was 4.16 dB with the largest difference between VEL 25 and VEL 50 (5.22 dB) and the smallest difference between VEL 13 and VEL 25 (3.25). The difference between VEL 2 and VEL 13 (4.02 dB) was between the other two.

Repeatability within Participants and Setting

Women. To assess if there was repeatability within participants and setting a two-way ANOVA with levels Block (1, 3) and VEL, (2, 13, 25, 50) revealed a significant main affect, $F(4, 3) = 44.9$, $p = .001$. However, there were no significant differences observed during post hoc analysis between contrasts of interest, namely, specific VELs between Block 1 and Block 3. Results confirmed our hypothesis that this method of measuring vocal effort had repeatability (refer to table 1 for a summary of means, standard deviations, and *t* statistics.)

Figure 5 Repeatability within Females

Table 1 Females: Means, standard deviations, and t statistics for each VEL between Block 1 and Block 3

VEL	Block 1		Block 3		Comparison between Block 1 and Block 3	
	Average dB	*SD*	Average dB	*SD*	*t* statistic	*p* value
VEL2	58.30	*9.83*	61.53	*7.46*	*t*(27) = -0.06	*p* = .477
VEL13	63.19	*10.06*	66.82	*6.06*	*t*(28) = -0.08	*p* = .468
VEL25	65.33	*10.24*	70.06	*6.08*	*t*(28) = -0.40	*p* = .345
VEL50	69.49	*10.60*	75.41	*4.97*	*t*(29) = -0.67	*p* = .253

Men. Likewise, men produced similar results. A two-way ANOVA with levels Block (1,3) and VEL, (2, 13, 25, 50) revealed a significant main affect, $F(4, 3) = 6.66$, $p = .04$. There were no significant differences observed during post hoc analysis between contrasts of interest, namely, specific VELs between Block 1 and Block 3. A summary of means, standard deviations, and t statistics can be found in table 2.

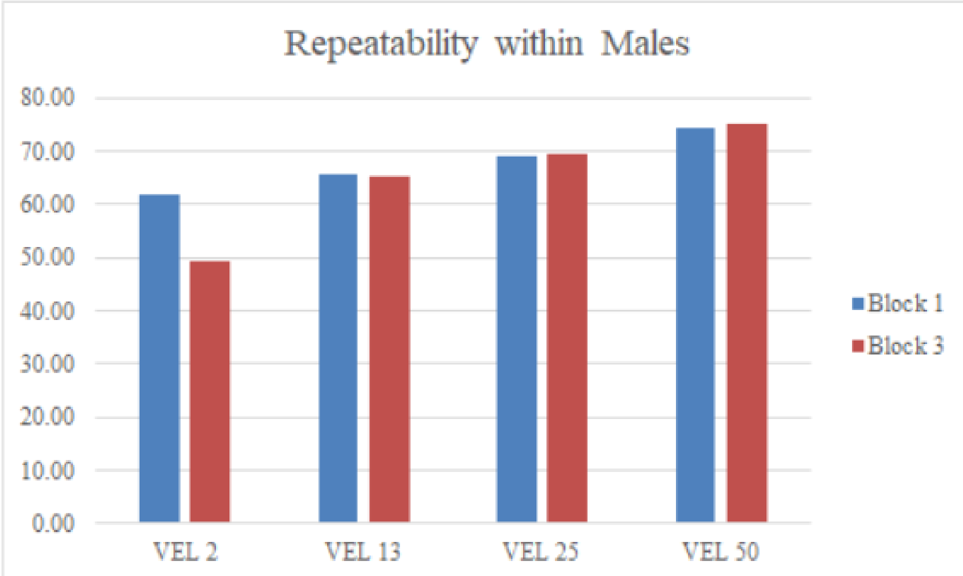

Figure 6 Repeatability within Males

Table 2 Males: Means, standard deviations, and t statistics for each VEL between Block 1 and Block 3

VEL	Block 1		Block 3		Comparison between Block 1 and Block 3	
	Average dB	*SD*	Average dB	*SD*	*t* statistic	*p* value
VEL2	61.691	*5.96*	61.53	*7.46*	$t(4) = 0.97$	$p = .195$
VEL13	65.7136	*5.65*	66.82	*6.06*	$t(8) = 0.08$	$p = .468$
VEL25	68.9634	*3.91*	70.06	*6.08*	$t(7) = -0.10$	$p = .462$
VEL50	74.1772	*4.16*	75.41	*4.97*	$t(8) = -0.35$	$p = .369$

Effects of Mood on Vocal Effort Levels

Differences in Intensity Levels

Females. To answer the question if mood effects intensity values for any given VEL, a two-way ANOVA with levels mood (baseline, aversive, neutral, positive, return-to-baseline) and VEL (2, 13, 25, 50) failed to reach significance, $F(4, 72) = 0.011$, $p = 1.00$. Post hoc analysis also showed no significance for comparisons of interest.

Males. Similar results occurred for males when answering the question if mood effects intensity values for any given VEL. A two-way ANOVA with levels mood (baseline, aversive, neutral, positive, return-to-baseline) and VEL (2, 13, 25, 50) failed to reach significance, $F(4, 21) = 0.644$, $p = .639$. Post hoc analysis also showed no significance for comparisons of interest.

Differences between Vocal Effort Levels

Females. To answer the question if the distances between each VEL differed with mood induction, differences between adjacent VELs were compared across emotion condition in a two-way ANOVA with levels differences (VEL 50 - VEL 25; VEL 25 - VEL 13, VEL 13 - VEL 2) and mood (baseline, aversive, neutral, positive) was conducted. Analysis failed to detect any significant differences between levels, $F(4, 72) = 0.609$, $p = .657$. Post hoc analysis also showed no significance for comparisons of interest.

Males. To answer the question if the distances between each VEL differed with mood induction, differences between adjacent VELs were compared across emotion condition in a two-way ANOVA with levels differences (VEL 50 - VEL 25; VEL 25 - VEL 13, VEL 13 - VEL 2) and mood (baseline, aversive, neutral, positive) was conducted. Analysis failed to detect any significant differences between levels, $F(4, 21) = 0.095$, $p = .923$. Post hoc analysis also showed no significance for comparisons of interest.

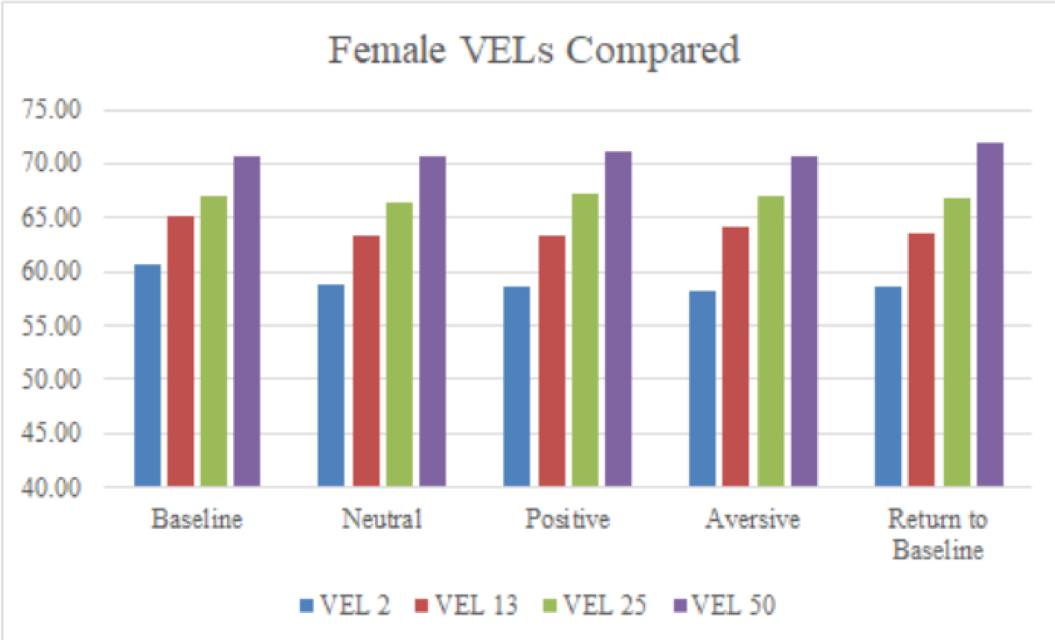

Figure 7 Female VELs Compared Across Conditions

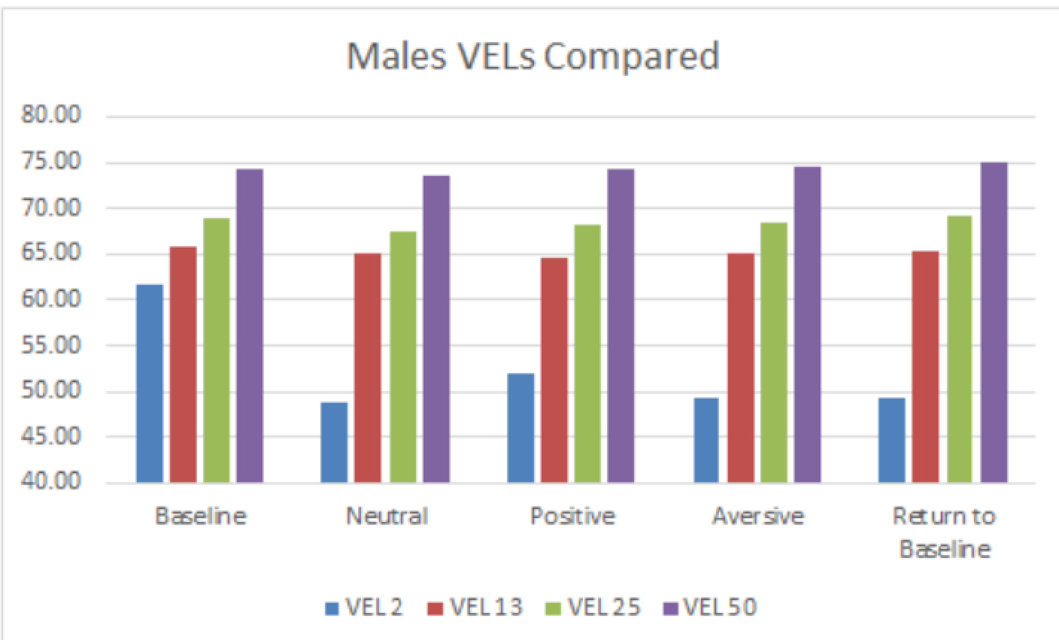

Figure 8 Male VELs Compared Across Conditions

Discussion

The current study sought to investigate the association between vocal effort and intensity in the initial and final experimental blocks of the experimental protocol. The analysis found that the four vocal effort levels (VEL 2, VEL 13, VEL 25, and VEL 50) were distinct and repeatable within the experimental blocks. Additionally, this study analyzed the differences in vocal effort levels with the introduction of emotional stimuli through aversive, neutral, and positive conditions in Block 2 of the experimental protocol. In the population investigated, no significant changes to vocal effort levels were noted between conditions for any of the VELs. However, results for VELs within the emotion conditions reflected similar results to Blocks 1 and 3. This emphasizes the first hypothesis of repeatability, but it does not support the second hypothesis of changes to VEL production with transient mood.

Main findings

Based on the average intensity output, as vocal effort levels increased, intensity output underwent a systematic increase that was different between each level. This suggests, within each experimental block, vocal intensity followed the proposed first hypothesis that vocal intensity was distinct for each level and repeatable between the initial and final conditions. Moreover, the association between those distinct levels were non-linear (having equal dB between each level) with the average differences equating to 3.73 dB SPL for females and 4.16 dB SPL for males between each VEL target. These findings reflect similar patterns to Berardi et al. (2021) who found that intensity differences were lowest for the VEL 13 to VEL 25 level and greatest between VEL 25 and VEL 50. However, the differences between VEL 25 and VEL 50 were somewhat greater in the Berardi et al study. These patterns were expected given that previous literature suggested physical correlates to perceptions tend to be logarithmic as opposed to linear.

The second purpose of this study did not produce results to support the secondary hypothesis. However, despite the consistent performance producing targeted VELs, there were changes in overall mood across each block as measured by the PANAS. The lack of significance in the data could be due to methodological considerations. In the emotional block of the experiment participants were required to hold the speaking task and targeted VEL in short term memory before advancing the screen to reveal the emotional picture. Only then did they speak. It is possible that the added cognitive effort needed to remember the speech task and VEL while waiting for the picture to begin could have added a level of complexity that impacted the intensity output.

There are other considerations that explain why the results of this study did not adequately capture changes to intensity output in a given VEL with the implementation of emotional stimuli. First, it is possible that the mood induction methods (i.e., picture viewing) were not appropriate to use for this research paradigm. Mood induction administered via picture viewing may not have elicited a substantial emotional response. Therefore, VEL would not be influenced by such small changes to mood.

Another consideration involves the paradigm used to capture the perceptual activity. Direct magnitude production may not be suitable for assessing intensity output changes with mood states and may require different perceptual experiences. Instead, mood may affect perceptions to a greater extent than production. A large body of literature suggests that mood effects cognition which would in turn affect perceptions more than productions (Derryberry, Rothbart, 1984). Due to this, our analyses of intensity would not demonstrate significant differences given that intensity output is physiological rather than perceptual. Instead, stronger results may be obtained by measuring how much effort the speaker perceived following a production such as with a direct magnitude estimation paradigm.

Previous research eliciting emotions through picture viewing has demonstrated subtle changes in contact quotient (van Mersbergen & Delaney, 2014; van Mersbergen, et al., 2015), relative fundamental frequency (van Mersbergen & Lanza, 2018), and cepstral peak prominence (van Mersbergen & Payne, 2020). However, these changes only varied by 1-3% difference. Given such a small change in vocal output observed in previous research, it is possible that for the current study, the sample size is too small to reveal such small, hypothesized effects across the emotion conditions. Thus, more trials to increase the power are necessary to obtain significant differences. If more trials are required, it is possible that any emotion effect would be eliminated due to physiological vocal fatigue because it would necessarily lengthen the duration of the experiment.

Furthermore, a different method to induce mood and emotion during this paradigm could have provided the necessary emotional experience to observe hypothesized changes in intensity. This study only captured transient moments of mood using a randomized presentation of three distinct mood states. It is possible that exposure to one specific emotion state for a greater duration (i.e. a longer period of positive picture viewing,) may have elicited a greater response in terms of intensity changes. However, there are caveats to this approach. One caveat is habituation. Participants might habituate to the mood-state and thus any effect of mood might dissipate after a number of trials. Another caveat is a task-shifting affect. If a participant spends an extended amount of time viewing positive pictures and then switches to aversive afterwards, a task shifting effect may occur and influence the first few points of data until the participant becomes fully engaged in the subsequent mood state. This would require an additional step that would facilitate a return to baseline before proceeding to the next emotion.

Finally, it is possible that the acoustic output variable assessed (intensity) may not have been the best measure to demonstrate clear differences. It could be that fundamental frequency, or another acoustic variable may be more appropriate to view clear changes following emotional stimuli.

Implications for Clinical Practice

Vocal effort is often experienced by many individuals facing vocal pathologies and other voice-related issues. Previously, it has been difficult to fully analyze someone's vocal effort perceptions due to the nature of the measurement tools at hand. Poor measurement systems inhibit our ability to capture patient perceptions, thus making it difficult to understand how these perceptions may impact physiological responses. By having the individual provide distinct and measurable vocal tasks based on their perceptions, both the patient experience and physiological output can be better analyzed in regard to how they may relate to one another. Through the results of the current study, we can reliably capture produced vocal effort in terms of intensity output through the Borg Centimax scale.

Supplementary findings

As discussed previously, two additional speech task categories were implemented in this study. The two additional task categories presented increasing cognitive demands on the participants and included three variations of standard sentence reading tasks and three different map tasks. While the reading task appeared to be relatively straightforward for the participants, the map task was significantly more challenging to complete for many in this sample. Per participant report post-experiment, many individuals struggled to understand and complete the task in a time-efficient and eloquent manner. In the Berardi et al. study (2021), the population involved had been from a metropolitan area where public transportation systems are used more frequently and therefore more familiar to the participants. Participants in this study, who were primarily from rural areas of the United States, may not have been as familiar with public transportation systems, and thus added an unintentional level of complexity to this study. Additionally, experience with public transportation may simply be a function of age and experience. The participants in this study may have been younger.

Limitations

A notable limitation of this study was the lack of male participants. With only 5 male participants, it is easy for one unusual data point within one participant to skew the rest of the analysis out of favor for determining if the hypotheses were correct. Running several more male participants through our protocol to obtain a sample size equal to or roughly the similar to the female participant numbers may have demonstrated results similar that of the female group in terms of significance between VELs in the various conditions.

Due to the lack of geographical relevance of the map task, the spontaneous speech trials were potentially too cognitively taxing for our given population, thus creating repercussions for the analysis of the non-emotional data. Furthermore, the map task might have unintentionally elicited emotional conditions in the form of frustration and worry at the lack of familiarity. Because our initial power analysis did not account for this unintended, extra condition, further analysis requires additional work beyond the scope of this paper. Therefore, the condition was disregarded.

Future Directions

To obtain better insight to the measurement of vocal effort, future studies should seek to analyze its association to other physiologic parameters and examine further practices regarding appropriate paradigms for this variety of research. From the acoustic data collected, other physiological features could have been selected to assess the association between perceived vocal effort and an individual’s physical response. Fundamental frequency, frequency range, and intensity range are all possible choices using simplistic acoustic data collection. Additionally, more elaborate physiological measures could have been taken, such as electroglottography, to assess non-perceivable physical output, such as contact quotient. This may be a reasonable measure to assess since contact quotient is often associated with greater collision forces of the vocal folds thus leading to edema causing an increase in work and perceived effort.

Future work may also consider implementing a different paradigm to measure the effect of emotional stimuli on vocal effort. For instance, a direct magnitude estimation paradigm that consists of participants estimating the vocal effort level they produced when given an intensity to speak at may be more appropriate. This paradigm more closely aligns with previous research on transient mood states effect on an individual’s perception of vocal effort.

While we had attempted to remain consistent with the methods of the previous study, additional experimental trials had been facilitated to address the secondary interests of the current study that had not been executed in the original study. Due to this, data analysis was different between studies.

Finally, if implementing a protocol similar to the one completed for the current study, future research must determine how to account for extraneous variables such as the impacts of increased cognitive effort. For example, future studies in non-metropolitan areas utilizing a similar program may benefit from using a different spontaneous speech task. While the map task in theory is great for providing a concise answer to a spontaneous prompt, other tasks could elicit equally concise responses without the additional cognitive effort necessary to answer. Utilizing tasks such as “tell me what type of car you drive” or providing a more simplistic map task should be investigated to provoke the spontaneous speech responses.

Conclusions

The current study provided insight and evidence to the significant association between vocal effort and intensity output. Future work is necessary to address how this association changes with the implementation of emotional stimuli as well as examine best practices in measuring these changes. Evidence found within this line of research will provide greater knowledge of how the perception of effort impacts the individualistic, physiological responses, which in turn poses implications for clinicians, researchers, and voice-users alike.

References

Anand, S., Bottalico, P., & Gray, C. (2019). Vocal Fatigue in Prospective Vocal Professionals. *Journal of Voice*. doi: 10.1016/j.jvoice.2019.08.015

Barrett, L. F. (2009). The Future of Psychology: Connecting Mind to Brain. *Perspectives on Psychological Science, 4*(4), 326–339.

Barrett, L. F. (2012). Emotions are real. *Emotion, 12*(3), 413–429.

Belsky, M. A., Rothenberger, S. D., Gillespie, A. I., & Gartner-Schmidt, J. L. (2020). Do Phonatory Aerodynamic and Acoustic Measures in Connected Speech Differ Between Vocally Healthy Adults and Patients Diagnosed with Muscle Tension Dysphonia? *Journal of Voice*, 1–6. doi: 10.1016/j.jvoice.2019.12.01

Berardi, M., van Mersbergen, M., & Hunter E.J. (2021). Relationship Between Tasked Vocal Effort Levels and Measures of Vocal Intensity. *Journal of Speech Language and Hearing Research, 64(*6):1-12

Borg, E., & Borg, G. (2002). A comparison of AME and CR100 for scaling perceived exertion. Acta Psychologica, 109(2), 157–175. doi: 10.1016/s0001-6918(01)00055-5

Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. Scandinavian Journal of Work, Environment & Health, 16, 55–58. doi: 10.5271/sjweh.1815

Borg, G. A. (1982). Psychophysical bases of perceived exertion. Medicine & Science in Sports & Exercise, 14(5), 377–380. doi: 10.1249/00005768-198205000-00012

Bottalico, P. (2017). Speech Adjustments for Room Acoustics and Their Effects on Vocal Effort. *J. Voice 31*(3), e1–392. doi:10.1016/j.jvoice.2016.10.001

Bottalico, P., Graetzer, S., & Hunter, E. J. (2016). Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *Journal of the Acoustical Society of America, 139*(5), 2870-2879. https://doi.org/10.1121/1.4950812

Bradley M.M. & Lang P.J. (1995). Measuring emotion: the self assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry*, 25, 49-50

Cheyne, H., Kalgaonkar, K., Clements, M., & Zurek, P. (2009). Talker-to-listener distance effects on speech production and perception. *The Journal of the Acoustical Society of America*, 126. 2052-60. doi: 10.1121/1.3205400.

Clark, L. A., & Watson, D. (1999). Temperament: A new paradigm for trait psychology. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research*, 399–423.

Crichton, N. (2001). Visual analogue scale (VAS). *Journal of Clinical Nursing*. 10. doi: 10.1007/978-3-211-89836-9_1450.

Derryberry, D., & Rothbart, M. K. (1984). Emotion, attention, and temperament. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognitions, and behavior.* New York: Cambridge University Press. (excerpt: pp. 132-153).

Dietrich, M. (2008). The effects of stress reactivity on extralaryngeal muscle tension in vocally normal participants as a function of personality (Doctoral dissertation). *University of Pittsburgh, Pittsburgh, PA.*

Eriksson, A., Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Perception & Psychophysics* 64, 131–139. https://doi.org/10.3758/BF03194562

Fairbanks, G. (1960). *Voice and articulation drillbook*, Harper & Row, New York.

Ford Baldner, E. F., Doll, E., & van Mersbergen, M. R. (2015). A Review of Measures of Vocal Effort With a Preliminary Study on the Establishment of a Vocal Effort Measure.

Journal of Voice, *29*(5), 530–541. doi: 10.1016/j.jvoice.2014.08.017

Gescheider, G. (1988). Psychophysical Scaling. *Annual review of psychology, 39*(1), 169-200. doi: 10.1146/annurev.ps.39.020188.001125

Hazan, V., Tuomainen, O., & Pettinato, M. (2016). Suprasegmental Characteristics of Spontaneous Speech Produced in Good and Challenging Communicative Conditions by Talkers Aged 9-14 Years. *Journal of speech, language, and hearing research: JSLHR, 59*(6), 1596–1607. https://doi.org/10.1044/2016_JSLHR-S-15-0046

Hunter, E.J. et al. (2020). Towards a consensus description of Vocal Effort, Vocal Load, Vocal Loading, and Vocal Fatigue. *Journal of Speech Language Hearing Research, 63*(2), 509-532. https://doi.org/10.1044/2019_JSLHR-19-00057

Jacobson, B. et al. (1997). The voice handicap index (VHI): development and validation. *Am J Speech Lang Pathol. 6*(3). 66-70. https://doi.org/10.1044/1058-0360.0603.66

Jónsdottir, V., Rantala, L., Laukkanen, A., & Vilkman, E. (2001). Effects of sound amplification on teachers' speech while teaching. *Logopedics, phoniatrics, vocology*. *26*, 118-23. 10.1080/14015430152728025.

Lagier A, Vaugoyeau M, Ghio A, Legou T, Giovanni A, Assaiante C. (2010). Coordination between posture and phonation in vocal effort behavior. *Folia Phoniatr Logop, 62*(4):195-202. doi: 10.1159/000314264

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). *International affective picture system (IAPS): affective ratings of pictures and instruction manual Technical Report A-8.* Gainesville: The Center for Research in Psychophysiology, University of Florida.

Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Verdolini Abbott, K. (2015). Vocal Fatigue Index (VFI): Development and Validation. *Journal of voice : official journal of* *the Voice Foundation, 29*(4), 433–440. https://doi.org/10.1016/j.jvoice.2014.09.012

Ohlsson, A. C., Andersson, E. M., Södersten, M., Simberg, S., & Barregård, L. (2012). Prevalence of Voice Symptoms and Risk Factors in Teacher Students. *Journal of Voice, 26*(5), 629–634. doi: 10.1016/j.jvoice.2011.11.002

Primov-Fever, A., Lidor, R., Meckel, Y., & Amir, O. (2013). The effect of physical effort on voice characteristics. *Folia phoniatrica et logopaedica: official organ of the International Association of Logopedics and Phoniatrics (IALP), 65*(6), 288–293. https://doi.org/10.1159/000361047

Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: a foundation for future research. *Journal of speech, language, and hearing research : JSLHR*, *43*(3), 737–748. https://doi.org/10.1044/jslhr.4303.737

Roy, N., Bless, D. M., & Heisey, D. (2000a). Personality and voice disorders: A multitrait-multidisorder analysis. *Journal of Voice, 14*(4), 521-548.

Roy, N., Bless, D. M., & Heisey, D. (2000b). Personality and voice disorders: a superfactor trait analysis. *Journal of speech, language, and hearing research : JSLHR, 43*(3), 749–768. https://doi.org/10.1044/jslhr.4303.749

Sapienza, C. M., Crandell, C. C., & Curtis, B. (1999). Effects of sound-field frequency modulation amplification on reducing teachers' sound pressure level in the classroom. *Journal of voice :official journal of the Voice Foundation, 13*(3), 375–381. https://doi.org/10.1016/s0892-1997(99)80042-3

Steinhauer, K., Grayhack, J. P., Smiley-Oyen, A. L., Shaiman, S., & Mcneil, M. R. (2004). The relationship among voice onset, voice quality, and fundamental frequency: A dynamical perspective. *Journal of Voice, 18*(4), 432-442. doi:10.1016/j.jvoice.2004.01.006

Svec, JG., Titze, IR., & Popolo, PS. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *Journal of the Acoustical Society of America, 117*(3), 1386-1394. https://doi.org/10.1121/1.1850074

Titze, I. R. (2000). Principles of Voice Production, Second Printing. *National Center for Voice and Speech*

Titze, I. R., Hunter, E. J., & Svec, J. G. (2007). Voicing and silence periods in daily and weekly vocalizations of teachers. *The Journal of the Acoustical Society of America, 121*(1), 469–478. https://doi.org/10.1121/1.2390676

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment*, *8*(1), 48–59. https://doi.org/10.1037/1040-3590.8.1.48

Van Leer, E., & van Mersbergen, M. (2017). Using the Borg CR10 Physical Exertion Scale to Measure Patient-perceived Vocal Effort Pre and Post Treatment. *Journal of Voice, 31*(3). doi: 10.1016/j.jvoice.2016.09.023

Van Mersbergen, M. (2011). Voice Disorders and Personality: Understanding their Interactions. *Perspectives on Voice and Voice Disorders, 21 (1),31-38*

Van Mersbergen, M., & Delany, M. (2014). Vocal responses to emotional picture viewing. *Logopedics Phoniatrics Vocology, 39*(3), 99–107. doi: 10.3109/14015439.2013.777108

Van Mersbergen, M., & Lanza, E. (2018). Modulation of Relative Fundamental Frequency During Transient Emotional States. *Journal of Voice,* 1-6

Van Merbsergen, M., Lyons, P., & Rielger, D. (2017). Vocal Responses in Heighted States of Arousal. *Journal of Voice, 13*(1) https://doi.org/10.1016/j.jvoice.2015.12.011

van Mersbergen M, Patrick C, & Glaze L. (2008). Functional dysphonia during mental imagery: testing the trait theory of voice disorders. *Journal of Speech, Language & Hearing Research, 51*(6), 1405–1423. https://doi-org.ezproxy.memphis.edu/10.1044/1092-4388(2008/06-0216)

Van Mersbergen, M., Vinney, L. A., & Payne, A. E. (2019). Cognitive influences on perceived phonatory exertion using the Borg CR10. *Logopedics Phoniatrics Vocology*, 1–11. doi: 10.1080/14015439.2019.1617895

Vinney, L. A., Van Mersbergen, M., Connor, N. P., & Turkstra, L. S. (2015). Vocal Control: Is It Susceptible to the Negative Effects of Self-regulatory Depletion. *Journal of Voice*, 1–11.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weinberger, S. H., and Kunath, S. A. (2011). The Speech Accent Archive: towards a typology of English accents. *Corpus-Based Stud. Lang. Use Lang. Learn. Lang. Doc.*, *73*, 265– 281.

Zadra, J. R., & Clore, G. L. (2011). Emotion and perception: the role of affective information. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 676–685. doi: 10.1002/wcs.147