Informed Consent Form

"MyoVoice to Restore Natural, Hands-free Communication to Individuals With Vocal Impairments"

NCT04762043

Approved 06-26-23

Study Title: "EMG Voice Restoration"
IRB Protocol: 20182089
Sponsor: Altec Inc., Natick, MA 01760
Funding: SBIR Phase II (NIH/NIDCD) ; Application # R44DC017097
PI: Gianluca De Luca

Background

This newly funded NIH Phase II SBIR research study will continue our development of subvocal speech recognition, which represents a cutting-edge technology with the potential to be used for a large variety of augmentative and alternative communication (AAC) applications. Unlike natural speech, subvocal (silently mouthed) speech is not dependent on acoustic interactions and thus is not susceptible to environmental noise corruption. More importantly, subvocal speech has the potential to provide a new and effective form of communication to those who have undergone laryngectomy and lost the ability to vocalize.

Recent studies of ours have demonstrated that surface electromyographic (sEMG) signals recorded from muscles of the face and neck can be reliably translated to text and synthesized into speech using a personalized, prosodic digital voice. When tested on a 2,500-word vocabulary of continuous phrases among participants with laryngectomy, we succeeded in translating subvocal speech-to-text with a 90%-word recognition rate. Taking these developments a step further in our Phase I SBIR, we demonstrated that it is possible to capture phrase-level differences in stress (realized via altered pitch, loudness, and/or duration) from the sEMG signals as well as synthesize natural, word-level variations in intonation and timing through a digital voice that is personalized to each individual. When tested among participants with laryngectomy, we observed clear improvements in speech intelligibility, acceptability, and stress discriminability over electrolaryngeal speech aid alternatives.

What Advancements are Provided by This Project?

Leveraging our prior advancements for sEMG-to-speech synthesis, we propose to develop the ***MyoVoice™***: a new sEMG voice restoration system that translates silently mouthed words and phrases into prosodic vocal output; i.e. digitized speech that includes intonation, emphasis, and rhythm to convey meaning that is personalized to the users original voice(see **Figure 1**). This innovation will also have broad impact among millions of AAC users by restoring not only the vocabulary, but also the prosodic attributes of speech within a personalized digitized voice.

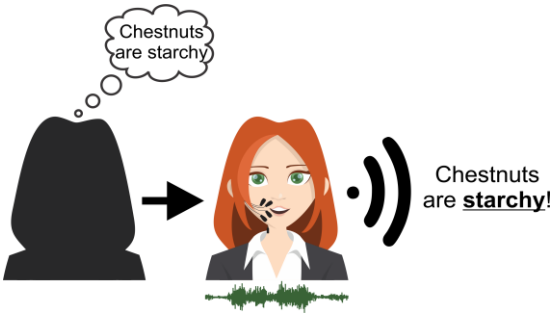

Figure 1. *Schematic of a proposed* ***MyoVoice™*** *system for restoring vocal communication by transcribing sEMG signals detected from speech muscles while silently mouthing words. The translation is re-interpreted as a personalized and prosodic digitized voice.*

Justification for Study

This project is prompted by the need for more effective, practical, and user-friendly AAC devices for persons unable to communicate effectively through vocalization. Because sEMG-based speech recognition does not rely on acoustic excitation of the vocal tract, it is readily applicable to recognizing subvocal (i.e., silently mouthed) speech. Subvocal speech is therefore an obvious alternative form of communication for people with laryngectomy who still retain use of their speech articulator muscles but lack the ability to vocalize.

Specific Aims of Study

In Aim 1, we translate our Phase I proof-of-concept phrase-level prosody recognition algorithms to ones that can transcribe within-phrase prosodic speech and track variations in prosody (pitch, loudness, and duration) to create linguistic boundaries, clarify lexical ambiguities, and convey emotional intent. To carry out this work, we acquired simultaneous acoustic and sEMG recordings as individuals with healthy voices speak a corpus of phrases comprising an extensive set of prosodic-phonemic combinations to elucidate the relationship between acoustic and sEMG representations of prosody. The data set was used to design new multi-level dynamic learning algorithms to identify person-specific variations in prosody based on a set of selected sEMG features. We also redesigned our Phase I subvocal speech algorithms that leverage the sEMG data to transcribe lexical content amidst the multitude of variations in pitch, loudness, and duration that exist in prosodic speech.

In Aim 2, we advanced our Phase I sensor and algorithm development efforts for conversational use by developing software to manage interactions between the sEMG sensors and our speech algorithms developed in Phase I. The Aim 1 algorithms were trained on a reduced sensor subset, and subsequently achieved comparable algorithmic performance with a smaller number of sensors that will be incorporated into a newly designed custom-fit, reusable sensor veneer to facilitate non-expert sensor placement for repeatable algorithmic performance for the individuals with total laryngectomy.

In Aim 3, the prototype MyoVoice™ system developed in Aims 1 and 2 was evaluated for social reception amongst persons with total laryngectomy and their primary speaking partner. Each participant evaluated the speech outputs from our sEMG-to-speech against the current gold-standard, electrolarynx. When evaluated for conversational efficacy, our MyoVoice™ system will be shown to yield greater preferences toward synthetic speech that integrates linguistic prosody based on fine-tuned estimates of pitch and loudness. The resulting pre-commercial AAC system will have broad-reaching impact for introducing a level of social interaction from an AAC device that is unprecedented among individuals living with articulatory-based vocal impairments.

Upon the successful completion of this Phase II SBIR, we will deliver a prototype ***MyoVoice™*** system complete with wearable, comfortable, user-accepted sensor and rapid mobile software that is cross-platform compatible and capable of translating sEMG signals recorded from muscles of the face and neck into personalized, prosodic digital voice for restoring the natural conversational capabilities for individuals with vocal impairments. The innovation of the ***MyoVoice™*** AAC device will have a broad-reaching impact for introducing a level of social interaction from an AAC device that is unprecedented among individuals whose articulator muscles remain intact yet lack the ability to vocalize effectively with a personalized, prosodic voice.

Involvement of Human Subjects

Human Subjects Involvement and Characteristics: Dr. Jennifer M. Vojtech, a Research Scientist at Altec, Inc., will coordinate the involvement of human subjects in the proposed project. She will work closely with the PI, Mr. Gianluca De Luca, and Secondary Study Coordinator, Dr. Serge H. Roy, to ensure that all aspects of human subject screening, enrollment, and testing complies with Federal regulations involving voluntary participation of human subjects in research, as determined by WCG IRB.

Human subject involvement will consist of the voluntary participation of adults for data collection experiments described in the Research Strategy. These experiments are of two types:

- *Study 1(a):* Collection of sensor and/or voice data during experiments in which subjects either read out loud and/or subvocally recite vowels by themselves, read text provided on a computer screen, and answer questions to elicit spontaneous speech; and

- *Study 1(b):* Listener experiments in which naïve control subjects listen to speech samples and make judgments about certain characteristics of the speech, such as how well they can correctly identify the vocabulary words (i.e., compute their *word error rate*) or distinguish between different speakers (i.e., evaluate *personalization* of digitized speech). The involvement of human subjects is described in greater detail for each experiment category:

Study 1(a): Data collection for sEMG-based recognition and synthesis of prosodic speech

General Description: This portion of the study is designed to create a data corpus to develop sEMG-based subvocal speech recognition and synthesis algorithms for prosodic speech in people with laryngectomy. The data collection will take place in the research facilities of Altec Inc. (23 Strathmore Rd, Natick, MA 01760) under the supervision of the PI and the assistance of Dr. Jennifer Vojtech, Dr. Serge Roy, and other research team members. Testing will also involve Dr. James Heaton, a consultant expert on speech disorders with extensive experience in recording EMG signals from facial and neck muscles in control subjects as well as people with laryngectomy.

The data collection will involve individuals with healthy voices. We will collect sEMG signals from key muscles of articulation from the face and neck (see **Figure 2**) while subjects are seated and asked to silently mouth vowels by themselves, read text that we provide to them on a computer screen, and answer questions we ask to elicit spontaneous speech. For experiments involving patients with laryngectomy, we will also acquire audio recordings of a subset of these phrases during either electrolaryngeal for comparison with subvocal speech. These alternative forms of speech are commonly used by people with laryngectomy and we therefore do not anticipate difficulty in finding subjects who are familiar with at least one, if not both, of these forms of speech production. Illustrations of these methods are provided in **Figure 3**. Electrolaryngeal speech is produced as an individual articulates their intended message while holding a hand-held electromechanical device (electrolarynx) against their neck to operate as an artificial larynx. Tracheoesophageal speech relies on a Tracheoesophageal Voice Prosthesis (TEP), which is a device made of medical grade silicone that is positioned between the trachea and the esophagus. The voice prosthesis itself does not produce a voice; instead it allows air to be delivered from the lungs into the esophagus where it is expelled through the mouth, resulting in a vibration of tissues in the lower pharynx, producing sound.

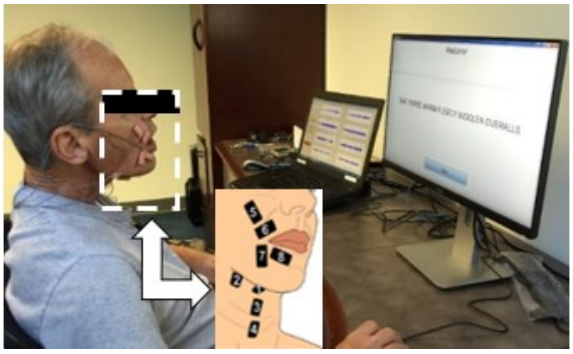

Figure 2. *Subject with laryngectomy operating the data acquisition system. One screen displays sEMG signals, the other displays sentence prompts.*

For volunteers with healthy voices, the same phrase tokens will be vocalized while acoustic speech and sensor signals are recorded.

All subjects will be video recorded during the experiments to help interpret the findings. These recordings will be destroyed at the end of the study.

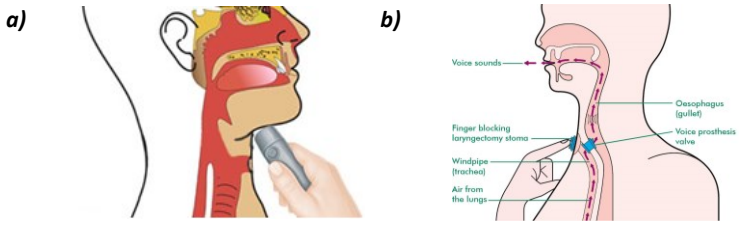

Figure 3. *Subjects will use their a) electrolarynx and/or b) TEP for comparative purposes with our* ***MyoVoice™*** *system.*

Study Site: The experimental data collections involving human subjects will take place in the research laboratory of Altec Inc. (23 Strathmore Rd, Natick, MA). The research laboratory and its equipment are adequate for implementing the

experimental protocols. When necessary, we will also exercise the option of conducting the data collection procedure in the prospective subject's home. This will be considered for instance when transportation is a problem. Our equipment is portable, and the protocol is of extremely low risk, so we do not see this accommodation as modifying the benefit/risk ratio.

Data Acquisition Experiment: During the experiments, sensor data will be acquired from 4–8 wireless sEMG sensors from the face and neck (see **Figures 2 and 4**), and exercise the option of adding 2 more sensors to the forehead and one wireless inertial (IMU) sensor to the chin to detect movement of the jaw or mouth. Sensor locations will be guided by previous EMG studies on pitch and volume modulation. A commercially available wireless data acquisition system designed for facial and other small-muscle applications (Delsys, Inc., Natick, MA) will be used to simultaneously record sEMG signals from the locations on the face and neck while the subject subvocalizes indicated stimuli (e.g., words or phrases; see **Figure 4**). A data acquisition graphical user interface (GUI) and infrastructure designed from previous studies will be used to implement the data corpus protocol in a manner that minimizes fatigue by allowing the subject to control the pacing of the sentences they read in a subvocal (silently mouthed) manner.

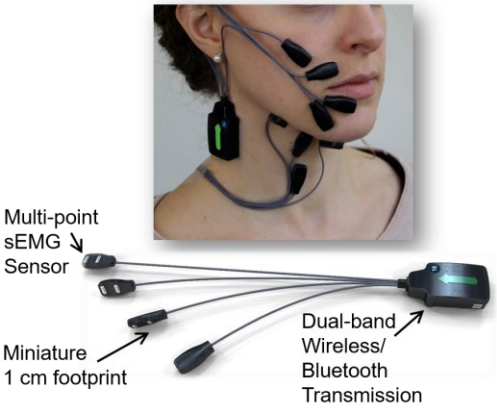

Figure 4. *Surface EMG data will be recorded using commercially available Delsys, Inc. sEMG sensors.*

Text prompts to the subject will be derived from a subset of a 2,500-word corpus that was previously used for building subvocal speech recognition models. Phrases will be selected to include a minimal set of training vocabulary with comprehensive phoneme-prosody combinations.

For the experiments involving individuals with healthy voices, the subjects will vocally speak the phrases while their voice is recorded by a microphone and sEMG signals are recorded from articulator muscles of the face and neck. These data will be used for developing sEMG-based algorithms for recognizing segmental and prosodic content of subvocal speech as well as for advancing sEMG voice synthesis engines for digitizing a personalized, intelligible, and expressive digital voice.

Sensor Placement: The skin area of each sensor location will first be prepared using a disposable shaver (if needed to remove facial hair), alcohol wipe, and several tape peels to exfoliate the site. The ventral neck sensor pairs will be placed in submental (#1, 2) and ventromedial (#3, 4) regions, and the face sensors will be placed over supralabial (#5, 6) and infralabial (#7, 8) regions (see **Figures 2 and 4**). Two additional sensors will be placed on the forehead to measure contraction of eyebrow elevating muscles (frontalis) and eyebrow furrowing muscles (corrugator supercilii), respectively. A double-sided hypoallergenic adhesive tape with cutouts for the electrode contacts with skin will be used to secure the sensors to the skin. This unilateral sensor placement allows for audio recordings using an electrolarynx on the contralateral side.

Subject Population: Although laryngectomy is much more prevalent in males than females (4:1; American Cancer Society, 2001), it affects both sexes. We will make every effort to include both males and females. Subjects must be proficient in English so that we can accurately associate muscle activation patterns to intended speech characteristics, as well as literate in English so that they can read speech task prompts from a computer monitor. Children will be excluded from participation as laryngectomy is extremely rare in this population.

Recruitment Process: Persons with total laryngectomy will be recruited through referrals from Dr. James Heaton, a consultant from the MGH-Center for Laryngeal Surgery & Voice Rehabilitation in Boston, MA. Our group has been collaborating with Dr. Heaton for the past 8 years during the early stages of developing subvocal speech capabilities. The Center maintains a sizeable database of clients with laryngectomy who have expressed their interest in participating in research studies and Dr. Heaton will be available to guide us in the selection and screening process. Dr. Heaton is also directly involved in local Laryngectomy support groups in New England and will have access to additional candidates. Control subjects will be recruited by word of mouth. The relatively small recruitment goal is not expected to necessitate postings, advertisements, or flyers.

Potential Risks/Minimizing Risks: Subjects will be exposed to no more than minimal risk as a result of their participation. Whenever electrical devices make contact with the body, there is always a risk of electric shock. This risk is minimized because the sEMG sensors and IMU sensors are wireless, low-voltage devices (i.e., no cabling connected to mains) and therefore are completely isolated from electrical sources. The sensors are also commercially available (Trigno™ Mini; Delsys, Inc.) and comply with U.S. and internationally recognized medical device standards as well as FDA oversight.

For persons with fragile skin, there is a risk of irritation or breakdown of the skin from the placement of sensors on the skin. The risk is equivalent to wearing a Band-Aid for several hours and peeling it off. This risk will be minimized by wiping the skin using an alcohol prep pad and securing the sensors using interface materials that are hypoallergenic or of medical grade (such as adhesive backings or hydrogels) and approved for medical use on the skin. We will also screen the subjects and exclude subjects with skin lesions or a history of frequent skin breakdown from bandages.

Areas near the stoma or approximating scarring or tissue sensitivity due to surgery or radiation will be avoided. Dr. James Heaton will be present during these data acquisition experiments and has extensive clinical experience in supervising such recordings. The skin will be inspected following removal of the sensors. We will have over-the-counter skin moisturizer available to alleviate any dryness to the skin that may result.

Because recording procedures are anticipated to last on average 4–5 hours, there is possibility of subjects feeling tired or bored during the recording session. Providing rest periods of at least 5 minutes per half hour will help alleviate such discomforts and maintain subject attention. Subjects will be free to stand up and walk around and we will provide light snacks, a lunch (if during mid-day), and beverages. If the subject appears to be getting tired despite these efforts and cannot attend to the task, we will ask them to return on another day to complete the study.

Due to the ongoing COVID-19 pandemic, subjects may increase their risk of exposure to COVID-19 by coming into Altec, Inc. and interacting with other people. COVID-19 is an infectious disease that is easily transmitted by air and can pose a serious health threat to some people. We will ensure that proper CDC and Massachusetts state regulations and guidelines are followed to minimize risk of COVID-19 transmission during the study.

Study 1(b): Listening experiments to evaluate recognition of synthesized voices

This portion of the study will be limited to n=10 subjects with healthy voices. Subjects must speak and be literate in English so that they can understand instructions, as well comprehend speech recordings that will be presented to them in English. We will seek recruits with no clinical experience with laryngectomees. Listeners must also be comfortable in typing phrases into a computer using a keyboard and mouse. Both males and females will be included in equal numbers so that possible sex-specific differences can be accommodated for men and women during the experiments. We will exclude candidates with conditions which might disqualify them for participation – such as poor hearing, speech disorders, or memory disorders that are identified through an initial interview screening.

Description of Test Procedures: The experiments will take place in the research facilities of Altec Inc. under the supervision of the PI and the assistance of Dr. Jennifer Vojtech and Dr. Serge Roy. Listening experiments will require 1–2 hours. Study participants will first undergo a brief hearing screening at 1000, 2000, and 4000 Hz randomly to each ear at 25dB through the headphones of an audiometer (Maico Diagnostics MA-27). Any participant who is unable to hear any of the three tones in both ears will have failed the screening test and will be compensated for their time/transportation and dismissed. After the hearing screening, subjects will be seated at a computer workstation in a quiet office setting. They will be instructed to listen to a series of short phrases through speakers or headphones that are like the phrases in Experiment 1(a). Subjects will be prompted at the completion of each phrase to make judgments about certain characteristics of the speech, including the acceptability of the speaker's voice, the identity of the speaker, and the emotion conveyed within the message. The criteria on which to judge acceptability will be based on acceptability instructions advocated by Bennett and Weinberg[44] which includes the speaker's fluency, rate, inflection, and "pleasant" quality. Subjects will also be instructed to orthographically transcribe the phrase they heard (i.e., to assess intelligibility) and answer a question about the phrase (i.e., to assess comprehensibility). Answers will be entered by the subject using a keyboard and mouse in a field visible on the screen below the prompt. The subject will be able to control when a new prompt appears and can interrupt the testing at any time to rest or ask a question. There will be no sensor recordings in this experiment or any instrumentation other than the computer and acoustic stimuli.
Recruitment: Same as described for control subjects in Experiment 1(a)

Potential Risks/Minimizing Risks: The primary risk/discomfort will be boredom or fatigue associated with the relatively long duration of the experiment. We will provide the same mitigating activities (rest, refreshments, etc.) as in Experiment 1(a) description. No risks associated with the use of sensors are a factor in this portion of the study.

GENERAL INFORMATION FOR BOTH STUDIES

Informed Consent: A single consent form will be used for Study 1(a) and 1(b). All persons will be properly informed of the purpose of the study by members of the research team and will have the opportunity to read the informed consent outlining the experimental procedures and review their rights as a human subject. Dr. Heaton will make the first contact with the prospective laryngectomy subjects to explain the involvement of human subjects in the study and answer any questions. Dr. Heaton will provide a copy of the informed consent form to the subject as a written description of the study which they will not sign until they have had the opportunity to visit Altec Inc. and meet with Drs. Vojtech and Roy, who will demonstrate the various equipment and review once more the nature and details of the study. Dr. Vojtech, Dr. Roy, or Mr. Gianluca De Luca (PI on the study) will be administering consent. Control subjects will be sent a copy of the informed consent form to review before deciding whether they wish to volunteer. The consent form will be administered upon arrival to Altec, Inc. after answering any questions and demonstrating the protocol. All subjects should be competent to give informed consent.

A signed and dated copy of the consent form will be provided to each subject, with the signature of the person obtaining consent. Identifiers linking the subject to their participation in the study will be kept in a locked cabinet accessible by the PI. Each of the researchers on this project is familiar with the appropriate Federal guidelines regarding the use of human subjects for research, and certificates of compliance required by the NIH are on file.

Confidentiality: It will be difficult to completely de-identify the data collected in Study 1(a) because there will be voice and videotaped recordings in addition to the sEMG recordings. For this reason, data will be stored on a password-protected PC computer with limited local accessibility and a high-security firewall preventing file access from external locations. All data will be backed-up, encrypted and password-

protected on external drives. Confidentiality will be assured by relying on a subject identification coding system. No references to the subject name or other identifiers (initials, address, license number, date of birth, etc.) will ever appear in publicly accessible documents or publications. The subjects' test results will only be made available to the subject and the researchers directly involved in the study. Since these experiments do not involve medical testing, these results will not become part of the subjects' medical record. The video/voice data will be kept in storage during the study and destroyed once completed.

Importance of Knowledge and Potential Benefits: The study is intended to provide an AAC technology that translates silent mouthed words and phrases directly into synthesized prosodic and personalized voice for those with vocal impairments. The potential benefits to society are high in terms of developing an EMG voice restoration system that restores not only vocabulary, but also the prosodic attributes of speech using an emerging personalized voice synthesis technology. The technology will be packaged into a wearable sensor system that provides natural embodiment, cosmetic appeal, and intuitive vocal capabilities not otherwise possible.

Potential Benefits to the Subjects: There are no anticipated benefits for subjects in this study aside from the potential benefits society may experience if our results lead to new EMG voice restoration technology.

Alternatives to Participation: This study does not involve medical diagnostics or treatments for which alternatives can be offered.

Costs to the Subject: Subjects will not incur costs for their participation.

Compensation for Participation: Subjects will receive compensation for travel expenses if they need to take public transportation or taxi. No parking expenses are anticipated as we have our own free parking lot. Individuals who come to our facility but fail to meet enrollment criteria will receive $25 remuneration for their travel time and effort. Subjects who complete the 4-5 hour recording protocol will receive $25/hour, and subjects completing only part of the protocol (due to their choice or technical difficulties on our part) will receive payment of $25/hour for their time spent. A private car transportation service door-to-door will be available at no cost to subjects that do not have other means of transportation.