A Qualitative Analysis of a Novel Face-Tracking Application and a Proposed Research Design to Study Glossectomy Speech using this Method

by

Roujan Khaledan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech-Language Pathology

Department of Communication Sciences and Disorders
University of Alberta

Abstract

Background: Following glossectomy, patients often utilize compensatory articulatory behaviors (CABs) to compensate for their acquired speech impairments. Incidental findings have indicated the presence of lower lip inversion CABs, but no study has investigated these inversions directly. A prototype face-tracking application was developed in-house at the University of Alberta that had the potential to be used for the proposed CAB study. However, due to its novelty for use in research, a qualitative analysis was needed to demonstrate how whether the prototype could capture the data necessary for the proposed CAB study.

Objective: The objectives of this thesis were to: 1) design a cross-sectional matched control study to determine the extent of lip inversion in glossectomy speakers when producing alveolar consonants; 2) determine the feasibility of the prototype face-tracking application for capturing speech kinematics; 3) investigate potential sources of error under different recording conditions.

Method: Objective one (designing a study) was addressed by designing a cross-sectional matched control trial to characterize lower lip CAB for alveolar sounds in glossectomy speech using the prototype application. This study was designed to be run with 20 participants in each of the experimental (glossectomy) and control (healthy) groups. Experimental words had voiced or voiceless alveolar sounds either word initially or finally. Objective two (determining the feasibility of the prototype for research) was evaluated by measuring various moveable and static portions of a single healthy participant's face and plotting movement trajectories during speech and non-speech tasks. Objective three (investigating potential sources of error) was investigated by assessing the standard deviations (SD) of select points on the face at rest under different recording conditions; the error during a speech task as also investigated by assessing the SD of a static point.

Results: The proposed research protocol was designed to compare lower lip movement of glossectomy speakers during alveolar sound production to healthy speakers by using a mixed-effect linear model (α=0.05, two-tailed). Then, the qualitative analysis showed that the prototype face-tracking application captured movement of articulators in speech and non-speech tasks. Also, the SD of measured points across all recording conditions were 'highly' accurate.

Conclusion: The work from this thesis provided a method for future researchers to study CAB in glossectomy speech. It also demonstrated the feasibility of the prototype for use in the proposed CAB and as well as other speech kinematic studies.

Preface

This thesis is an original work by Roujan Khaledan. The research project, of which this thesis is a part, received research ethics approval from the Health Research Ethics Board of Alberta Cancer Committee (HREBA-CC), “Compensatory Articulatory Behaviours of the Lips in Post-Glossectomy Speakers”, Ethics ID: HREBA.CC-19-0520, January 20, 2020.

Acknowledgments

I would first like to thank my supervisor, Dr. Daniel Aalto, for the wisdom and patience with which he approached this project. His expertise was invaluable in shaping this thesis. You encouraged me to organize my thinking and allowed me to critically evaluate my work at each step. Next, I would like to acknowledge the members of my committee, Dr. Caroline Jeffery and Dr. Gabi Constantinescu for their insight and support.

Finally, I would like to thank my rich support system of family and friends for the inspiration and motivation they provided.

Table of Contents

List of Tables and Figures

TABLES

FIGURES

Chapter 1. Compensatory Articulatory Behaviors in Glossectomy Speech: a proposed research protocol

The aim of this thesis project was to design a research study to investigate whether glossectomy speakers use lip inversions as compensations in their speech. To track the participants' face motion, the use of a prototype face tracking application was proposed. However, as this prototype face-tracking application was novel for use in research, a qualitative analysis was required to determine whether the prototype was capable of capturing speech motions and whether different recording parameters were potential sources of recording error.

Chapter 1 outlines the methods by which the research study was designed, and Chapter 2 discusses the qualitative analysis of the prototype face-tracking application.

The overall contribution of this thesis is to present a means for understanding face motion during speech. The first chapter provides a specific method for future researchers to investigate glossectomy speech. The second chapter demonstrates that the prototype face-tracking application can be used for face-tracking research, including the proposed glossectomy speech study.

1.1. Introduction

The intricate structures of the upper aerodigestive tract allow an individual to breathe, speak, and swallow. The presence and treatment of head and neck cancer (HNC) can impair critical functions (Kreeft et al., 2009; Martin-Harris et al., 2017). In particular, individuals with oral cavity HNC (o-HNC) (Kreeft et al., 2009) and oropharyngeal cancer (op-HNC) (Argiris et al., 2008) often undergo glossectomy (Hinni et al., 2013). This treatment can lead to changes in speech functioning and negative social (Bressmann et al., 2009), or self-perceptions (Constantinescu et al., 2017) of the individual's speech and reductions in quality of life (QoL) (Campbell et al., 2000; Dzioba et al., 2017; Petruson et al., 2005; Radford et al., 2004).

HNC is typically the result of a squamous cell carcinoma (Schwartz & Hayes, 2020; Scully & Bagan, 2009). One-third of these HNC cases present with early-stage I or II disease (Argiris et al., 2008; Duvvuri et al., 2004). Additionally, public awareness about the numerous environmental risk factors for HNC, including tobacco consumption (Do et al., 2003; Hecht, 2003; Kumar et al., 2016; Nair et al., 2004; Parkin et al., 2005), have shaped its incidence, leading to declines in HPV-unrelated HNC and increases in HPV-related HNC (Chow, 2020; Forte et al., 2012). The latter type has been linked to better treatment outcomes (Rhodus et al., 2014; Tsai et al., 2016) because it commonly affects younger individuals (Forte et al., 2012) with fewer other risk factors (Argiris et al., 2008; D'Souza et al., 2007). Epidemiological and histological differences of HPV-related HNC also lead to more positive treatment outcomes compared to HPV-unrelated cases (Chaturvedi et al., 2011; Nguyen-Tan et al., 2014).

Given the high survival rate of HNC patients, cancer treatments that increase QoL are important to develop (Shin et al., 2012). In their scoping-review of cancer survivorship, Richardson et al. (2011) recommend that future research be done on areas that impact patient wellbeing on a longer scale over 5 years. Due to the long-lasting effects that acquired speech impairments in HNC patients can have on QoL (Bressmann et al., 2009; Chuanjun et al., 2002; Crombie et al., 2014; Dwivedi et al., 2009; Dzioba et al., 2017;

Rieger et al., 2006), research investigating how HNC patients speak following intervention is critical (Rogers et al., 2007; Shiboski et al., 2000).

1.1.1. Treatment

The heterogenous presentation (Schwartz & Hayes, 2020) of HNC necessitates a variety of treatment options, each one can have differing effects on patients' functional outcomes. The treatment plans available to HNC patients most commonly included glossectomy (Hinni et al., 2013). However, treatment could also include a combination of neoadjuvant, adjuvant, or postoperative (chemo) radiation therapies (Shin et al., 2012). One possible outcome of radiotherapy that could affect speech was tissue fibrosis and loss of range of motion (Abendstein et al., 2005; Dirix et al., 2006; Garden et al., 2006; List et al., 1999; Rosenthal et al., 2006; Trotti et al., 2003).

Glossectomy

Glossectomy is a surgical resection of a cancerous portion of the tongue and a margin of healthy tissue (Hinni et al., 2013). Due to the variability in cancer presentation, and the complex anatomy of the tongue, a standardized nomenclature for glossectomy resection locations does not exist. However, Ansarin et al.(2019) proposed several terms to refer to glossectomy based on the location and amount of resection.

A glossectomy can be closed locally through primary closure, or the resected portion can be reconstructed from a flap of tissue harvested from a healthy part of the patient's body. Flaps may be either free, completely removed from a different body part, or pedicled, tissue that remains partially attached to a proximal anatomical location (de Vicente et al., 2008). Both flap types can be harvested from a variety of anatomical locations (Argiris et al., 2008; Frederick et al., 2013; Huang et al., 2012; Janik et al., 2019; Kini, 2015). The surgeon's choices in reconstruction reflect the specific characteristics of that particular case and will have a specific functional impact for the patient. There are many decisions when selecting a reconstruction method. First, the decision to choose a reconstruction over a primary closure is dependent on the size of the resection. Larger resections are typically reconstructed (Acher et al., 2014; Yadav, 2013). Next, the type of flap chosen for reconstruction depends on its characteristics and its similarity to the resected parts of the tongue. Reconstructing the anterior two-thirds of the tongue, the mobile portion in the oral cavity, requires less bulk than the posterior one-third, the immobile root in the oropharynx, as the tongue is leaner and more mobile anteriorly (Green & Wang, 2003; Yadav, 2013).

Tumors in the oropharynx are difficult to visualize, so invasive techniques were previously used for resection. However, newer techniques, such as transoral robotic surgery (TORS), allow for greater visualization with fewer surgical risks (Bhayani et al., 2010; Mercante et al., 2013; O'Malley et al., 2006). These techniques improve organ preservation (Lewin & Hutcheson, 2014; Nouraei et al., 2015; Yadav, 2013), and are related to more positive functional outcomes (Mercante et al., 2013; Moorhouse & Edwards, 2018).

Mobility and sensation can both be affected by the type of flap chosen, and both can have a functional impact on speech outcomes (Chuanjun et al., 2002; Loewen et al., 2010; Stone et al., 2012). Flap mobility is affected by flap volume (Kreeft et al., 2009), amount (Dios et al., 1994), location (Korpijaakko-Huuhka et al., 1999), and the flap's connections to other tissues (Buchaillard et al., 2007). The type of flap chosen also dictates whether sensation can be restored in the flap (Biglioli et al., 2006; Loewen et al., 2010; Yadav, 2013). Finally, the extent of healthy tissue resected is at the surgeon's discretion (Hinni et al., 2013). Resecting these margins is a way to decrease the chances of cancer recurrence (Grimm et al., 2017).

While the amount of tissue resected is dependent on the size and location of the tumor (Bressmann et al., 2009; Fagan, 2014), even a small resection can cause changes to tongue movement for speech (Neligan, 2013). Despite these severe impairments, speakers are highly intelligible following glossectomy (Dziegielewski et al., 2019). However, research as to how these speakers can attain these levels of intelligibility is lacking (Blyth et al., 2015). Therefore, given the variability that exists in cancer presentation, treatment, and the resulting functional impacts, it is important to collect demographic data from participants in order to fully characterize glossectomy speech.

1.2. Compensatory Articulatory Behaviors in Glossectomy Speech

Compensatory Articulatory Behaviors (CAB) are the non-standard use of articulators to produce a speech sound in order to compensate for an impairment. Standard articulations used in healthy speech are discussed in the following section (speech sound classification and terminology). Preliminary evidence has shown that glossectomy speakers use a variety of CAB during speech (Barry & Timmermann, 1985; Georgian et al., 1982; Greven et al., 1994; Hagedorn et al., 2014; Morrish, 1984). The research designed in this chapter served to provide a method to study lower lip CAB by comparing differences in lower lip movement during glossectomy speech to healthy speech. Preliminary evidence for lower lip CAB existed in the literature (Barry & Timmermann, 1985; Georgian et al., 1982; Greven et al., 1994; Hagedorn et al., 2014; Morrish, 1984). The study proposed in this thesis provided a method for replication of these incidental findings.

Lower lip CAB were observed in Hagedorn et al. (2014). They used real-time magnetic resonance imaging (rtMRI) to observe the articulation of five glossectomy speakers. The participants had undergone glossectomy with reconstruction and radiation therapy. Their cancer had either occurred on the oral tongue, base of tongue, or both. None of the participants had received speech intervention prior to the study. Interestingly, one participant who had a resection of the oral tongue only, and had difficulty producing finely controlled movements with his anterior tongue was observed to be using CAB for the alveolar plosive, /t/. This participant made a dorsal constriction (similar to what would be seen in the production of a velar plosive such as, /k/) while elevating his lower lip to the upper teeth. He demonstrated this CAB in the single word context when producing the alveolar plosive, /t/. This CAB was also observed when the participant attempted an alveolar-lateral, /nl/, consonant cluster and the alveolar continuant, /s/, in connected speech. This study provided preliminary evidence for CAB of the lips, but the study design was limited as characterizing CAB was not the researchers' central goal, therefore their research methodology, including the number of participants, was not sufficient to fully capture these behaviors.

In a case study of a total glossectomy speaker, Whiting (1965) reported that the speaker accurately produced /t/ using a CAB of the lips spontaneously. She also reported that two partial glossectomy speakers spontaneously adapted their articulations to produce normal sounding speech. Nonetheless, neither of these reports clarified the nature of the CABs beyond clinician observations. Though, Whiting reported that more posterior sounds, such as /k/, were more likely to be accompanied by pharyngeal CABs than labial ones.

Similarly, Georgian et al. (1982) described a patient whose tongue was sutured to the buccal mucosa inside of the cheek following a partial glossectomy (20%), and hemi-mandibulectomy, resection of the mandible. The patient had received a total of 8 months of speech intervention targeted at training CAB articulation. His production of alveolar plosives was observed through videofluoroscopy in a sagittal view. CABs were observed when he produced /t/ and /d/; he made a dorsal constriction followed by bringing the two lips together. He brought his lips together by protruding the upper lip and retracting the lower lip and chin. It is noteworthy to mention that these authors note a significant acoustic difference between his productions and typical ones. Typically, the articulatory occlusion produced in plosives allows a build-up of air, which is then released all at once. This release, a stop burst, was a critical feature of plosive sounds, and was not observed in this participant's production. The authors hypothesized that this difference was attributed to a lack of strength which kept the speaker from maintaining the occlusion. However, it was not clear whether strength was the only factor that led to this difference or whether it would be expected of other glossectomy speakers using the same CAB. What is clear, is that CABs have implications for the acoustic properties and therefore perceptual aspects of speech production in glossectomy speech as well. This connection with perceptibility is an important reason to characterize CABs more fully as their use can have an impact on glossectomy speakers' intelligibility.

The presence of other CAB in glossectomy speakers has been observed too. In their study of total glossectomy speakers, Greven et al. (1994) found that speakers substituted labiodental, bilabial, and laryngeal gestures for various lingual sounds in both word and sentence contexts. Moreover, Bressmann et al. (2004) found that speakers with poorer non-speech mobility made more speech sounds at an erroneous place of articulation. Despite the fact that they only classified one of their speakers' productions as a CAB, this relationship between mobility and place errors shows that glossectomy speakers use non-standard productions in response to their impairments. In another study, a speaker used his epiglottis and the reconstructed floor of the mouth to create CABs following a total glossectomy (Morrish, 1984). The author postulated that the large redundancies in normal speech facilitated the use of CABs to achieve intelligibility. However, a limitation of this study was in the method of capturing movements, as researchers only used a video camera with a sagittal view.

1.2.1. Speech Sound Classification and Terminology

Articulatory phonology theory is a branch of linguistics that characterizes speech sounds in normal speech based on how speakers produce them. This section provided a review of speech sound classification and terminology as an overview of natural healthy speech. The International Phonetic Association classifies speech sounds according to their place and manner of articulation. Manner of articulation refers to whether a sound is voiced or voiceless. To produce a voiced sound, the speaker occludes their vocal folds to which causes the vocal folds to vibrate. The place of articulation refers to the positioning of the articulators for each particular sound. There are several different places of articulation: bilabial, labiodental, dental, alveolar, postalveolar, retroflex, palatal, velar, uvular, pharyngeal, glottal. Bilabial and labiodental sounds are produced through lip movements, the lips touching each other and the lower lip touching upper teeth, respectively. Excluding retroflex and glottal places of articulation, the remaining places are named for the area of the palate or the teeth with, which the tongue makes contact. The positioning of the tongue shapes the acoustic energy emitted from the vocal folds, and these resonances create the acoustic patterns that listeners rely on to perceive sounds as speech sounds (Buchaillard et al., 2007; Story & Bunton, 2010).

Sometimes, the tongue does not make full contact with the palate or the teeth, as in semi-occluded sounds such as /s/. Other times, when producing fully occluded consonants, otherwise known as plosives, the speaker must make full contact with the palate above. This contact is required to halt the airstream momentarily, as in /k/ or /t/. In both cases, speed, precision, and compressive force are required for the speaker to accurately make the desired sound. However, while semi-occluded sounds also require the speaker to raise their tongue, a full range of motion or compressive force is not required, because the tongue does not need to be raised as high as in plosive sound production nor does the speaker have to exert effort to seal the occlusion.

In this proposed study, voiced and voiceless alveolar plosives (/t/ and /d/) will be compared to voiced and voiceless velar plosives (/k/ and /g/). Alveolar plosives were chosen to be the experimental sounds because of the preliminary evidence presented for CAB in Georgian et al. (1982) and Hagerdorn et al. (2014). Specific details about measuring these articulations were discussed in the Methods section of this chapter. The choice to use velar sounds as controls was made because no lip movement was expected for the production of these sounds since the tongue is making contact with the posterior portions of the palate, the velum. So, while there may be some impairment to velar plosive production, glossectomy speakers were not likely to use their lips to support these productions.

1.2.2. Speech Anatomy and Physiology

Tongue

In addition to glands, blood vessels, fatty tissue, mucosa and connective tissues (Perrier et al., 2003; Sanders & Mu, 2013), the tongue is composed of several muscles that vary in direction and origin (Sanders & Mu, 2013; Stone et al., 2018). It consists of extrinsic muscles, those originating from a surrounding bone, and intrinsic muscles that originate and insert within the tongue (Sanders & Mu, 2013). These intrinsic muscles are highly interdigitated and are situated between the extrinsic muscles of the tongue. Extrinsic muscles are interdigitated at their place of insertion but bundled elsewhere. Finally, the floor-of-mouth muscles are contiguous with the tongue and are not interdigitated (Stone et al., 2018). The hypoglossal nerve (Cranial Nerve XII) controls tongue movement and the trigeminal (CN V) and glossopharyngeal (CN IX) nerves predominantly control tongue sensation (Lowe, 1980; Palmer et al., 2008).

Historically, it was accepted that the extrinsic muscles of the tongue, were responsible for movement, since their stable bony origin leads to a predictable movement outcome, while the intrinsic muscles were thought to be solely responsible for shaping the tongue. However, evidence that external muscles are also interdigitated and have multiple insertion locations, and complex innervation shows that the extrinsic muscles may be involved in shaping the tongue too. While the functional effects of every muscle in the tongue is not known (Stone et al., 2018), the anatomical and physiological complexities of tongue movement can be simplified as general patterns. The tongue muscles responsible for protrusion are the genioglossus (Sanders & Mu, 2013; Stone et al., 2018), and the vertical and transverse muscles (Stone et al., 2018). Tongue elevation is mediated by the superior longitudinal, styloglossus, transverse, mylohyoid, geniohyoid, anterior belly of the digastric and the genioglossus muscles, though the effects of elevator tongue muscles are secondary to that of the jaw muscles. Both protrusion and elevation are important components of alveolar stop production as the tongue must elevate and protrude to reach the anterior edge of the hard palate, the alveolar ridge, and co-contraction of these muscles is likely (Stone et al., 2018).

Physiologically, the complex musculature comes together to form functional units that could be used to deform the tongue to almost any shape (Kier & Smith, 1985). There was disagreement over the number, location, and independence of functional regions within the tongue (Green & Wang, 2003). For example, Öhman (1967) and Stone (1990) have proposed three or four, respectively, functionally independent tongue regions, however, the delineation and exact degree of coupling between these units during connected speech tasks were not known.

Nevertheless, there was still evidence that the different regions of the tongue could be coupled or decoupled depending on the sound being produced (Green & Wang, 2003). For instance, the tongue could either protrude or retracted as a unit, or specific portions of the tongue could be moved while other parts were kept stable (Hiiemae & Palmer, 2003; Sanders & Mu, 2013). Producing the alveolar sound /t/ requires decoupling of the anterior and posterior tongue as the anterior tongue was elongated it became less massive and this mass was transferred to the posterior tongue (Green & Wang, 2003). Stabilizing the posterior tongue required active involvement from the musculature in the jaw (Hiiemae & Palmer, 2003). Similarly, decoupling of the tongue and jaw negatively affected maximum tongue strength (Solomon & Munson, 2004). Precise tongue movements during speech require fine control of the motor units within the tongue (MacNeilage, 1970; Ostry & Keller, 1983). However, speakers with primary closure underwent a loss of tissue (Grimm et al., 2017), and those with flap reconstructions lack active control over their flap's mass (Stone et al., 2014), in both cases, glossectomy could affect a speakers' control over their tongue movements.

Lips

The lips receive their motor innervation from the buccal branch of the facial nerve (CN VII). Sensory innervation for the lips comes from the trigeminal nerve (CN V); the maxillary branch (V2) innervates the upper lip while the mandibular branch (V3) innervates the lower lip. The orbicularis oris is a circular muscle that comprises the lip, and is predominantly involved in lip aperture. This muscle is held in place by the surrounding facial muscles (Piccinin & Zito, 2020). Visually, the lips are the red, pink, or brownish structure that is encapsulated within the vermilion lines of the upper and lower lips. The lips are independent of the tongue, as they do not share any muscles or nerves (Fischbein et al., 2001; Hoefflin, 1998). However, both structures are attached to the jaw (Stone et al., 2018; Sussman et al., 1973) and are coupled with it for movement (Westbury et al., 2002).

In speech, the lips are the primary articulators in bilabial and labiodental consonant sounds and certain vowel sounds (Santosh et al., 2017). They also have secondary involvement in other consonants through phonetic variation. The patterns of primary lip involvement for speech articulation have been organized into visemes, the general posture of the lips for different classes of sounds. Bilabial sounds all share a single viseme, as do labiodental sounds (Bozkurt et al., 2007; Popat et al., 2012). However, muscular activation for two different bilabial sounds does not arise from identical motor activity, despite their shared place of articulation (Gick et al., 2012; Mayer et al., 2011). Based on the variation in muscle activation, it is important to have an accurate, reliable, and objective face-tracking measure as is the case with the prototype face-tracking application. In vowel sounds, lips are protruded and rounded during the production of rounded sounds (Bozkurt et al., 2007). Acoustically, lip aperture is an important element in vowel production. Lip aperture, the opening of the lips, is achieved by spreading the upper and lower lips and by lowering the jaw (Browman & Goldstein, 1992). When healthy speakers separate their lips, the lower lip coupled with the jaw move more than the upper lip; but all three articulators have highly synchronous movement (Green et al., 2000). The lips may also be retracted or spread for certain vowels and consonants (Garnier et al., 2006).

In their investigation of lip movement for speech, Gracco and Abbs (1986) found that speakers used consistent movement patterns i.e., they ordered the movements of their upper and lower lips, and jaw temporally. Though there was greater within speaker variation of the lower lip and jaw than for the upper lip. Fowler & Salzmann (1993) found an additional movement pattern, that the jaw actively rises during alveolar sound production with the lips passively rising. This shows that the movement of the lips and the jaw are coupled with each other. It also shapes the expectation for the present study, that healthy speakers will not have active lip involvement during alveolar sound production. The lip and jaw can also be decoupled. In rounded sounds, Borghese et al. (1997) found maximal movement variation in the lips between participants, but very little in the jaw. They suggested that in these cases, the jaw acts as a stabilizer for the movements of the lips. It may be possible that glossectomy speakers who actively use their lips for CABs will follow a similar pattern of decoupling. The sensitivity of the prototype face-tracking application to capture these patterns, is advantageous for this proposed study because it will be possible to characterize the interaction of different articulators in glossectomy and healthy speech.

Connected Speech

Speech production relies on rapid and precise tongue movement to create the articulatory postures (Bennett et al., 2007). In healthy speech, the transition is fast and effective, such that the speaker's message is intelligible to the listener. Production of consecutive sounds does not occur sequentially, and each sound is produced in a slightly different way based on the sounds that occur before and after it. Coarticulation is the tendency of previous or anticipated sounds to shape the production of the articulated sound (MacNeilage, 1970, 1970). For example, in a context like "skit", the /s/ may shape the production of /k/ such that the tongue makes contact more anteriorly than it would have if the /k/ occurred word-initially in a word like "kit". While the articulation, and therefore the acoustic properties, of a coarticulated sound, are affected by surrounding sounds, perceptually there is no difference in the resulting sound produced (Browman & Goldstein, 1992). Therefore, coarticulation creates the smooth effect of running speech; without coarticulation speech sounds would be produced in halting succession.

Articulation is also mediated by listener factors. The speaker is aware of the listener's familiarity with a word, related to its frequency of use, and the predictability of an upcoming word based on the context. These factors mediate their productions. While, according to H&H theory, speakers' defaults are to hypoarticulate (underarticulate with smaller movements and effort) their awareness of these listener factors lead them to choose a production that is maximally distinguishable and therefore understandable (Lindblom, 1990).

Despite the acceptable variability arising from coarticulation and listener factors, there are constraints on speech production. According to the theory of categorical perception, listeners perceive the range of possible acoustic productions in categorical boundaries (Liberman et al., 1957). That is, a variety of productions can be perceived as a single speech sound so long as they are produced within the perceptual boundary for that sound. If a speaker produces a sound beyond that boundary, the listener will either perceive a different speech sound or it is possible that the intelligibility will be affected. In the case of plosive sounds, speakers must not only position their articulators so that airflow is occluded, but they need to maintain the closure for long enough that sufficient air pressure builds up behind the occlusion. The resulting release of air is called a stop burst. Therefore, based on the theory of categorical perception, while variability in speech production is abundant, correct productions are confined to some boundaries (Holmes, 2002; Story & Bunton, 2010) to be considered accurate.

Healthy Aging Speech

As people age, several factors lead them to produce speech with a set of compensatory gestures. These factors mainly have to do with the loss of strength and mobility of the tongue due to atrophy and neuromuscular factors. These results indicate that as people age they are at increased risk for disorder following disease or injury (Bennett et al., 2007).

1.2.3. Post-Glossectomy Speech Physiology

In general, speech outcomes are measured in a variety of ways, intelligibility, perception, and accuracy. Due to the complexities in tongue muscle orientation and the variability in size and location of resections, glossectomy speakers are a heterogeneous group. This variability, taken with the diverse outcomes from surgery and radiation, makes predicting the functional impacts of glossectomy difficult (Barry & Timmermann, 1985). However, numerous factors shape speech outcomes for glossectomy patients, despite the fact that there is not a consensus as to which factor is the critical one. It may be that the site of resection (Petruson et al., 2005), amount (Buchaillard et al., 2007; Rentschler & Mann, 1980), reconstruction type (McConnel et al., 1998; Michiwaki et al., 1993) or mobility (Bressmann et al., 2004; Hagedorn et al., 2014) is related to better speech outcomes.

Tongue mobility is important for good post glossectomy speech, as measured by intelligibility (Bressmann et al., 2004). Mobility may be affected by scar tissue formation or anchoring secondary to reconstruction techniques (Chuanjun et al., 2002; Wong et al., 2007). The location of a resection may also impact a speaker's range of motion; base of tongue resections often done in op-HNC have the effect of pulling the tongue posteriorly and limiting anterior range of motion (Logemann et al., 1997). Bressmann et al. (Bressmann et al., 2004) found that mobility scores for the patients whose genioglossus muscle, the one primarily responsible for protrusion, was significantly lower than those whose genioglossus muscle was not affected. In this study, glossectomy speakers performed the worst for simultaneous tongue elevation and protrusion postures, such as reaching the tongue to the nose. This non-speech task has a similar movement trajectory to alveolar sound production. These participants had the most mobility for the task of raising the dorsum of their tongue to the posterior aspect of the palate, a posture used for velar consonant production as well. The results from this study indicate that mobility may play a role in speech outcomes and provides support to observations that speakers use posterior constrictions in CABs. However, a lack of tongue mobility in a different study led to erroneous productions of velar plosives. However, as mobility is not the only, nor the most important factor in intelligibility (Bressmann et al., 2004), characterizing CAB in glossectomy speakers is valuable to understand glossectomy speech.

Tissue volume also played a role in speech outcomes. In postoperative radiation therapy (PORT) a loss of volume results in reduced coordination of muscle movement, which was needed for accurate speech production (Shin et al., 2012). In glossectomy, specifically, a hemi-glossectomy, the entirety of the styloglossus, and inferior and superior longitudinal muscles were removed. Significant portions of the genioglossus, and the upper part of the hyoglossus were also removed (Buchaillard et al., 2007). Through a computer simulation, Buchaillard et al. found that this loss of volume led to asymmetric movements and tongue shapes during speech production. This had implications for the production of alveolar consonants since the occlusion needed to produce these sounds requires sufficient volume in the palatal regions of the speaker's tongue. Following the physical changes resulting from glossectomy, speakers may compensate for their impairments by slowing their rate, stressing sounds (Engelbrecht et al., 2006) or CABs (Hagedorn et al., 2014).

1.1. Proposed Research Question

The primary goal of this section was to design a rigorous research method that can meaningfully answer if glossectomy speakers employ CAB of the lower lip for alveolar consonants was attained by designing a protocol that answered the following research questions. (1) Do glossectomy speakers employ different lip movements when producing alveolar plosive sounds compared to healthy speakers?; (2): Were there differences in lip movement in the alveolar sound condition than the velar sound condition? The first question will determine whether glossectomy speakers are using their articulators differently than healthy speakers. The second question will determine whether glossectomy speakers use their lips differently when producing alveolar plosives than when they produce velar plosives.

1.2. Methods

To test the research questions, a cross-sectional matched control trial method was proposed. Differences in lip movement between groups will be determined by comparing the Euclidean distance between the nose and bottom lip when participants were producing alveolar and velar plosive sounds. Differences between the glossectomy and control participants will be analyzed using a mixed-effect linear model (α=0.05, two-tailed) approach. If a statistically significant difference is noted in the lip movement between groups for alveolar plosive sounds, this will be considered a CAB of the lower lip. If no difference is found, or if a difference regardless of condition, or in the velar plosive condition only is found, that will not be considered a CAB. In both cases, speakers' articulations will be characterized using movement trajectories and simultaneous video and audio recordings. The Health Research Ethics Board of Alberta - Cancer Committee had approved this project *HREBA.CC-19-0520* (Appendix I).

1.2.4. Participants

HNC survivors, who have undergone partial glossectomy, will be compared to age and gender-matched healthy control speakers. The inclusion criteria are listed in Table 1.

Table 1
Inclusion Criteria

- HNC patients with a partial glossectomy, regardless of other treatments will be recruited for this study, however, given the differences in outcomes for various treatments, descriptive data about each participants' treatment will be collected through a chart review. - For HNC participants, they must not have any bone, mandible or maxilla, or lip resections that could interfere with their ability to produce sounds beyond the impairments that result from glossectomy alone. - Total glossectomy participants will be excluded. - All participants must be native speakers of English and not have any disorders of speech or language (this is to control for any differences in articulation attributed to accented speech or disordered speech/language).

Participants will report their age and gender at the time of sampling, two participants will be matched if their ages are within ±5 years of each other. The rationale for age and gender-matching participants were the age-related changes in speech production (Bennett et al., 2007; Schötz et al., 2013), and sociolinguistic variations of speech production based on age, region, and to a lesser degree, gender (Jacewicz et al., 2009). Potential regional differences will be partially controlled by recruiting only native English speakers who all live in the greater Edmonton area. Since the targets in this study are limited to lip movements during single sound production, the possibility of recruiting speakers of varying dialects does not introduce much error. When when the other controls were taken into account, taking the other controls into account, the chances that dialect-related rate and production differences will be reflected in the data is minimal. Information about the participants' resection size, location, and reconstruction type (if applicable) will be gathered through a chart review as well. This will provide insight into the characteristics of the speakers who utilize CABs.

With repeated measures, a total of 1600 observations (data points) are needed per statistical parameter for sufficient power based on the work done by Brysbaert and Stevens (2018) to observe an effect if there was one. Therefore 20 participants for each speaker group will be recruited and each participant will say 40 experimental and 40 control sounds. A mixed-effect linear model will be used to parametrically analyze the data (α=0.05, two-tailed).

1.2.5. Materials and Procedures

Participants will attend a single data collection session at a lab in the University of Alberta, they will sit 55 cm from an iPhone placed on a phone stand, so that the camera is at a consistent height. The recordings will be captured using the prototype face-tracking application. Participants will read one of two experimental phrase lists, presented on a large screen adjacent to the iPhone while hearing cocktail party noise through over-the-ear headphones. Participants will receive the instructions listed in Appendix II.

Cocktail-Party Noise

To ensure natural productions, participants will hear cocktail party noise through over-the-ear headphones playing 85dB. The use of this background noise has been used in other studies to increase naturalness and to induce the Lombard effect (Šimko et al., 2016). Increasing naturalness will facilitate generalizability of the results gathered in a lab setting to real-world situations. Thus, increasing external validity. The Lombard effect yields clearer and more extreme articulations (Garnier et al., 2006), which will decrease the risk of a type II error (a false negative) by making it easier to observe an effect if there is one.

Stimulus

The experimental sounds, the alveolar plosives /t/ and /d/, will be compared to control sounds, the velar plosives /k/ and /g/. The rationale behind choosing velar sounds as the control sounds is that the production of velars have very little, if anything, to do with the lips. For these sounds, the posterior aspect of the tongue must be raised to the soft palate, velum. Since this movement is not likely to recruit the lips, it is unlikely that impairments could be compensated with the lips.

These target sounds are embedded into words and then into short phrases for a variety of reasons. First, doing so, controls for the articulatory variability that arises through coarticulation while ensuring that the naturalness of the task is maintained. It also ensures that the participant is unaware of the sound of interest. Preserving naturalness increases the applicability of the results to real-life situations by accounting for the linguistic variability in language. Furthermore, by embedding alveolar plosives into linguistically variable contexts (adjacent to varying sounds), insights into the potential effects of phonetic context on CABs are possible. It is also possible that the speaker will use slightly different CABs depending on the sounds preceding or proceeding the alveolar sound. Future research should be targeted to understanding the extent and subtypes of CABs once their presence in a naturalistic context is established.

Following this methodology also has the advantage of extending the evidence that currently exists in the literature for CABs at the word-level, to the sentence-level. In contrast to the word-level, exact productions of each word are not necessary for maintaining intelligibility (Kashino, 2006). Therefore, it is possible that the hyperarticulation typical of word-level productions in laboratory settings encouraged CAB to a greater extent than would a naturalistic speaking task at the sentence level. Generalizability is critical when studying functional outcomes of glossectomy.

Finally, this methodology has the strength of specifically observing alveolar plosive sounds. The previous evidence of CABs was the result of incidental observations, which yielded a broad description of the types of sounds that might be aided by CABs. However, by directly observing alveolar sounds, this study will provide a clearer picture of CAB use in this context.

Proposed Experimental Phrase-Lists

The experimental phrase lists for the proposed study are generated following the procedure outlined in Vainio et al., (2010). First, minimal word pairs are generated for voiceless (/t/ and /k/) and voiced (/d/ and /g/) alveolar-velar pairs. A minimal pair consists of two words that are identical except for one sound. For instance, *bid* and *big*. Minimal pairs are identical except for the presence of /d/ or /g/. In the minimal pairs generated for this study, the target sounds are either word-initial or word-final. Target sounds are placed into a word-medial position due to the common phonological processes of alveolar sound reduction. For example, the ‘t’ in, *butter,* is rarely produced as /t/, instead of making full contact with the alveolar ridge, the speaker’s tongue quickly taps it. This phonological process is referred to as a reduction because a tap, /ɾ/, is easier and faster to produce; acoustically, this results in a very short stop burst. Since /ɾ/ is distinct from a /t/ and this process is so prevalent, articulation data from medial ‘t’ sounds would not be reliable or comparable to that of word-initial or final /t/.

These minimal word pairs are placed into identical, or near-identical, sentence frames. The minimal pair words are never clause initial or clause final. Each sentence pair is designed to be semantically and pragmatically felicitous with either minimal pair word. For example, the minimal pair ‘gust’ and ‘dust’ are embedded into the frame, ‘the dust/gust on the shore was particularly evident today, wasn’t it?’. This sentence is meaningful and natural when either ‘dust’ or ‘gust’ are used. Sometimes a preceding sentence provides additional context to support the pragmatic felicity of the sentence frame. For example, ‘tub’ and ‘cub’ are placed in the following sentence frame, ‘Sherry looked and looked all day. She finally found the tub/cub near the corner store’. For context, either ‘Sherry was renovating her house’ or ‘Sherry was a new zookeeper’ proceeds the sentences. Sentence frames are designed in a way where the target word is not the semantic focus of the sentence. In the previous example, the sentence "look at the tub/ cub” would not be appropriate because the sentence semantically directs the reader’s attention to the minimal pair words. Such a sentence could not only alert the participant of the task, but it could also affect the intonation or stress they place on the word. When it is not possible to have identical sentence frames for a minimal pair, the word immediately preceding the minimal pair words is always identical. For instance, the minimal pair words ‘call’ and ‘tall’ are embedded in the following frames, ‘I’d like her to call my order’/ ‘I’d like to, but it’s too tall an order’.

As part of this process, 86 minimal pair words are generated. Frequency effects of words are controlled by ensuring that words containing the alveolar target sounds occur as frequently as words containing the velar target sounds. Word frequencies are collected from the Corpus of Contemporary American English (COCA) database (Davies, 2008). The logarithm of usage per million words across 8 categories: blog, web, TV/ movies, spoken, fiction, magazine, newspaper and, academic, from the years 2015-2019 is collected for all minimal word pairs. Frequency of velar (M = 1.35, SD = 0.80) to alveolar (M = 1.22, SD = 0.88) words are compared using an independent samples t-test, t (172) = -1.042, p= 0.578, two-tailed. Therefore, word frequencies are not significantly different across sound types and no frequency effects are expected.

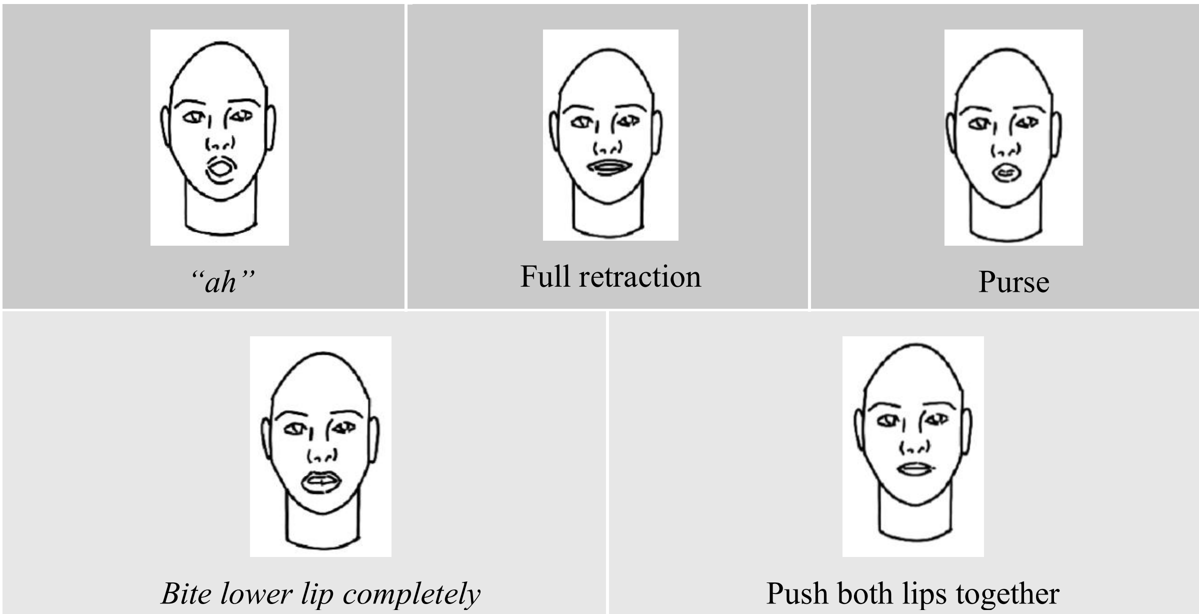
From these minimal pairs, 20 alveolar phrases and 20 velar phrases are pseudo-randomly selected to create one experimental list (Appendix III). The same process is repeated with the remaining phrases to create a second word list. The experimental word lists were pilot tested with a native speaker of English. The goal was to have the speaker read the sentences aloud and to report whether any sentences were difficult to say aloud or whether they were perceived as being semantically or pragmatically odd. The speaker did not give any such feedback. As the pilot participant was reading, the researcher was listening for any cases where target words were overly stressed or the focus of the sentence. Only one sentence was excluded due to overemphasis on the target word. In this case the next sentence in the list took its place.

Participants will be randomly assigned to one of the experimental lists. Additionally, as a measure of internal consistency, each participant will read 3 sentences from the other experimental phrase list and will re-read 3 sentences from their experimental list. These repeated sentences will provide a basis for direct comparisons to be made between participants' productions. This will also provide additional control for between-speaker variability or any potential sentence effects from either experimental list.

Additional Stimuli

In addition to the experimental sentence lists, participants are also instructed to produce five non-speech facial expressions, Figure 1. These facial expressions are chosen based on oral mechanism structure and function exams (OME). Speech-Language Pathologists use OME's to assess mechanisms related to speech and swallowing. These facial expressions discussed in detail in Chapter 2. The movement trajectories from the non-speech stimuli can also be used to comment on non-speech function. This information can support conclusions about the presence, absence and/ or degree of CAB noted in speech. Additionally, a secondary benefit of non-speech information are the potential insights into functional changes related to glossectomy size, location, or surgical procedure.

Figure 1
Non-Speech Tasks



1.3. Proposed Analysis

The descriptive statistics gathered from the participants will provide insight into CAB in terms of resection size. The insights gained from the descriptive statistics may contribute to scientific understanding of how resection size relates to speakers' use of CABs. Previously, Bressmann et al., (2009) conducted a study relating social perceptions of glossectomy speech to the amount of tongue tissue resected. These researchers found that a resection greater than 20.4% predicted a deterioration in speech acceptability of glossectomy speech. These findings may be better understood by relating resection size to CAB.

The methods for analyzing face motion are discussed broadly here, Chapter 2 provides a more detailed discussion. For the proposed study, the data from the prototype will be used both qualitatively and quantitatively. First, plotting movement trajectories and positional curves will provide detail into the participants' behaviors, refer to Table 2 for a definition of both.

Table 2

Movement Trajectories and Positional Curves

Term	Definition
Movement Trajectories	The distance between two articulators or the distance of certain positions on the face relative to an origin point (3-D position) plotted by frame
Positional Curves	3-D position of several points on the face plotted at one point in time (i.e., one frame)

Plotting movement trajectories will show how the participants' articulators move over time, and positional curves will provide insight into the extent and symmetry of movements at a given frame. Next, Euclidean distances between the chin and the nose at the time of articulation of the experimental or control sounds will be analyzed parametrically using a mixed-effects linear model (α 0.05, two-tailed). The decision to use the distance between the chin and the nose rather than a point on the lips is motivated by the notion that, if an inversion is present, the participants lip will be out of the camera's view. Therefore, tracking the lips increases the potential that the tracked point will be obscured, and the data will be affected by recording error. Likewise, with positional curves, recording error attributed to the obstruction of the lips is possible. Therefore, for a more thorough understanding of participants' articulations, positional curves in various positions should be plotted. Specifically, the points along the vermillion lines of the lips in a coronal view and points from the lower lips to the chin along a sagittal view should be plotted. A positional curve of the lower lip along the coronal view will provide insight into the symmetry of the participants' lip inversions, if they are present. Additionally, a positional curve along the sagittal view can provide insight into the extent of inversion.

In addition to the three-dimensional (3-D) data, observations from the audio and video files will provide insight into the participants' behaviors. Also, intelligibility estimates from the audio files can provide insight into the effectiveness of CABs if they are present.

1.4. Discussion

The information from this proposed study can provide ample evidence to characterize CABs in glossectomy speakers. First, the presence of CAB will be investigated qualitatively, then lip motions between the two groups will be compared parametrically. Also, the potential for incidental, but detailed, secondary findings is high due to the rich data that can be captured with the prototype face-tracking application. The results from this study could, therefore, contribute to the field by characterizing CAB and other traits of glossectomy speech in detail. These findings will have the potential to contribute to future treatment planning. Another, equally important addition that this project would bring to the field, is the that it would be the first research study to utilize the prototype face-tracking application for face-tracking in research.

1.4.1. Intervention Research

Interventions for glossectomy speakers have not been adequately researched, and often did not have high treatment fidelity. As such, standard speech intervention protocols or guidelines did not exist for this population (Logemann, 1999). In their meta-analysis of intervention studies for partial glossectomy, Blyth et al., (2015) reviewed seven articles, Skelly et al. (1971), Meyerson et al. (1980), Dworkin (1982), Bryant (1991), Denk et al. (1997), and Zhen et al. (2012). Each article described speech or swallowing. Of these, only four articles researched speech interventions. Nonetheless, all of these studies involved training CAB in response to patients' acquired impairments, that is, "matching target behavior of clear sound production using a new method of articulation" (Blyth et al., 2015, p. 407). Half of these studies incorporated tongue and jaw range of motion exercises as well. From their review, these researchers concluded that the research was low quality overall with poor experimental design and control. Also, all the studies lacked treatment fidelity and replicability. Therefore, it was not clear which technique was the effective one.

There was evidence to support that training compensatory articulations could improve intelligibility (Furia et al., 2001; Sunila et al., 2011). For example, in their study of a patient with cancer of the buccal mucosa, Sunila et al. (2011) found that in addition to increasing awareness of target sounds, it was effective to work on behavioral modifications where the speaker was expected to consistently substitute articulations and modify his rate. Despite the differences between participant characteristics in their study versus the proposed study, Sunila et al. provided evidence for the notion that CABs, when trained, could have a positive impact on intelligibility. LaBlance et al. (1991) also mentioned training CABs in their guide to rehabilitation of glossectomy speakers. In particular, they recommended training speakers to bring their lower lips to their upper incisors or alveolar ridge. However, it was not clear what evidence they used to make their recommendation. Furthermore, the expected effects on speech were also unclear.

These research gaps could possibly be attributed to the lack of understanding of spontaneous CAB in glossectomy speech. The findings from this project could, therefore, shape future intervention research.

Chapter 2. Qualitative Analysis of Prototype Face-Tracking Application for Research

2.1. Introduction

In the previous chapter, a research design aimed at investigating glossectomy speakers' articulation was proposed. The proposed method for tracking movements in this study was a custom face-tracking application (prototype face-tracking application), which was developed at the University of Alberta. This prototype had not been used previously in facial kinematic research. Therefore, the following qualitative analysis was needed in order to ensure that this prototype was capable of capturing the data necessary for the proposed glossectomy speech study as well as other future face-tracking studies. In this qualitative analysis, a range of speech and non-speech motions were tracked to demonstrate that the prototype was capable of tracking a broad range of motions. Then, potential sources of recording error were investigated. Insights regarding error were useful for providing recommendations for future research protocols.

2.2. Measures of Facial Kinematics

The kinematic data required for the proposed glossectomy speech study and the kinematic measures used in previously published face-tracking studies were discussed in the following two subsections. This was because, the central aim of this chapter was to ensure that the prototype face-tracking application was able to capture the data necessary for a range of future face-tracking research, as well as the study proposed in Chapter 1. This demonstrated that the system was feasible for a range of studies positioned the prototype face-tracking application as a viable option among existing face-tracking methodologies.

2.2.1. Proposed Study

As discussed in Chapter 1 (Stimulus), the glossectomy speech study the stimuli consisted of a natural speech task (phrases with alveolar and velar plosives) and a non-speech task (similar to an OME). In this study, lip inversion CABs were the behavior of interest. Therefore, the prototype needed the capability to track the movements of the lips during both speech, and non-speech tasks. The planned analysis for the glossectomy speech study had a qualitative and quantitative component. This meant that the prototype application needed to yield movement trajectories and positional curves to provide qualitative data. The prototype also needed to yield numerical measurements to provide the quantitative data needed for parametric analyses.

2.2.2. Other Studies

The measures used in previous facial kinematic studies provided insight into the range of behaviors that future researchers may use the prototype to capture. Therefore, previous face-tracking studies were discussed here. Speech kinematic researchers have investigated the degree, and timing of movement. For instance, in Löfqvist & Gracco (1997) the timing and degree of lip closure were used to investigate labial kinematics during consonant production. Similarly, Westbury, Lindstrom, and McClean (2002) studied the amount of coupling between the lips and jaw by comparing the movements of these articulators during a natural speech task. Other parameters studied by previous researchers were, articulator shape, movement trajectory, velocity, opening, and retraction (Gómez et al., 2019; Löfqvist & Gracco, 1997; Parrell & Narayanan, 2018).

Non-speech motions have also been studied. The non-speech motions that have been studied were similar to the non-speech task proposed in Chapter 1. These tasks provided researchers with a measure of participants' neurological function following a stroke (Bandini et al., 2018). In the future, these non-speech tasks may also serve as a pre- and post-treatment measure.

2.3. Prototype Face-Tracking Application

The prototype face-tracking application made use of Apple's Augmented Reality Kit (ARKit) to track the face. Originally, the ARKit was designed for consumer usage (discussed in the, 'Current Uses in the Consumer Market' subsection), but the ARKit had limited use in research. Previous studies have utilized a variation of this technology. For instance, Breitbarth et al.(2019) tested the accuracy of the Truedepth camera for distance measurements of a flat surface using a different Apple framework. Additionally, Scherr et al.(2019) used the ARKit to detect when a participant was making certain facial expressions. Other researchers have compared the reliability of static measurements made with the ARKit to other face-tracking systems (Borduas et al., 2020). The prototype was the first, to the author's knowledge, to extract the positional data obtained from the ARKit for facial kinematic research. Therefore, this thesis was the first to investigate how the data from the ARKit, through the prototype face-tracking application, could be used for facial kinematic research.

2.3.1. Underlying Technology

The prototype face-tracking application made use of Apple's Augmented Reality Kit (ARKit) to track the face. By projecting a numbered mesh onto the face, the system used a series of infrared lights to track the face. This static structure-from-motion reconstruction (Linowes & Babilinski, 2017) technology allowed the system to track 1220 points on the face. The movement of the points were tracked in 3-D as a displacement from an origin point along the x-y-z coordinate planes. Relative to the origin point, the x-axis indicated a left-right displacement; the y axis indicated up-down displacement; z axis indicated frontward-backward displacement. Using a single-shot technique, this system was designed to capture data at 60 frames per second (fps) (Appendix IV discussed the actual frame rates for a series of recordings).

While the ARKit was compatible for use with all iPhones and iPads with a Truedepth camera, the data for this thesis were captured with an iPhone X. Therefore, it was not possible to comment on any between device differences.

In addition to the ARKit, Apple's ReplayKit, allowed the prototype to capture audio (44.1kHz) and video (60 frames per second (fps) using a 7megapixel (f/2.2) camera). The face-tracking data and the audio/visual (A/V) data were not simultaneous, as the A/V recordings started after the face-tracking had begun. The figures in Appendix V illustrated the topics discussed in this section namely, the lag between the 3-D and A/V recordings, the numbered points on the face, and the orientation of the axes.

2.3.2. Current Uses in the Consumer Market

The usage of the ARKit in the consumer market demonstrated that this technology was valuable for consumer-based security, pleasure, and communication applications. Specifically, iPhones and iPads with a Truedepth front-facing camera made use of the ARKit for the FaceID. This allowed users to unlock their device by orienting their face towards the Truedepth camera. The Animoji feature, standard on these iPhones and iPads, allowed users to control a cartoon avatar with their facial expressions. Third-party developers have designed other applications that use the ARKit. For instance, Swiftable created an Augmentative and Alternative Communication application called Jabberwocky (Swiftable, 2021). This application allowed users to make selections by orienting their face to the portions of the screen they wished to select.

2.4. Problem Statement

Despite its merits in the consumer market, a qualitative analysis of the prototype was needed for two reasons. The first being, the novelty of the prototype in the research context. Since it was initially designed for consumer usage, the ability of the system to capture data for facial kinematic research was not known. In fact, observations about the alignment of the projected mesh and the face revealed that, contrary to expectations, the mesh did not align with the participant's face during extreme facial motions. Further, as the prototype relied on infrared light to track the face, it could be affected by obstructions to certain points on the face. For example, the lip inversion CAB discussed in Chapter 1, would lead to participants obscuring their lower lip from the camera. Therefore, it was critical to investigate whether the prototype was capable of capturing a range of speech and non-speech motions.

The second reason the qualitative analysis was warranted was that the level of error attributed to various recording conditions were not known. Understanding the effect of recording conditions on the data, provided insight for future use of the prototype for research. Prior to this thesis, there were no recommendations for research protocols with this prototype application.

2.4.1. Research Questions

The following research questions were investigated in this thesis. The rationale for both questions was discussed in the subsections below.

> Question 1. Was the prototype capable of capturing kinematic data necessary for the study designed in Chapter 1 and for other future face-tracking studies?
>
> Question 2. What level of error could be expected from the data given different recording conditions when the participant was at rest; and what was the level of error in a static position when the participant was completing a speech task?

Question 1: Kinematic Data

The aim of question one was to determine whether the prototype was able to capture kinematic data typically used in face-tracking research. Three criteria were developed in order to determine whether the prototype was successful in capturing kinematic data. These criteria were used for both the speech and non-speech tasks. The first criterion was that the movements of various points were temporally aligned and followed a consistent movement pattern. This criterion was developed because of the kinematic relationships between articulators (Solomon & Munson, 2004; Westbury et al., 2002). The second criterion was that the movement patterns of particular articulators for a given task were consistent across repeated measures. Descriptions of movement patterns in both criteria demonstrated the qualitative information that the prototype could yield. The final criterion was that kinematic measures such as: movement trajectory, positional curves, movement velocity and amplitude could be tracked and plotted with the prototype. Specific details regarding the tracked points were discussed in the Procedure section.

Question 2: Error

As this prototype had not yet been used for research, it was not known whether certain recording conditions (recording distance, angle, and participant blinks) were potential sources of error. Additionally, the recording error during a task where the participant was moving was not known.

2.5. Methods

2.5.1. Participants

The author of this thesis collected data on herself to test the prototype face-tracking application. She was 24 at the time of recording and was bilingual (since birth) in both English and Farsi. She had no underlying movement, speech, or language disorders.

2.5.2. Materials

The prototype face-tracking application was accessed through an iPhone X throughout this thesis. The iPhone was placed on a rigid stand. During the production of the speech tasks (discussed below) the participant listened to multi-talker babble through over-the-ear headphones.

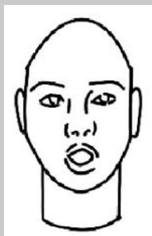
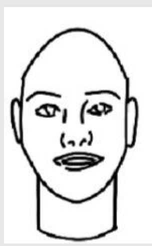
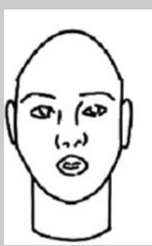
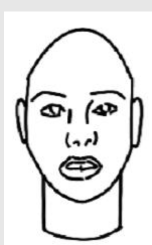
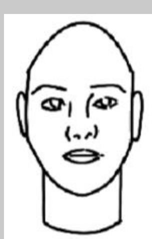
2.5.3. Question One

Stimulus

The stimuli in question one were designed to be representative of the range of possible speech and non-speech tasks. This decision allowed for a thorough examination of the prototype's capabilities in capturing facial kinematic data. In order to meet this objective, the speech task consisted of five words in which the lips were active articulators (rounded, open, bilabial, and labiodental speech sounds). The choice to incorporate these particular words was made according to Bozkurt et al.(2007). In their study, Bozkurt demonstrated that there is a particular face shape that corresponds to each speech sound; these face shapes are called visemes. From their work, the visemes where the lips were most actively involved were chosen (Figure 2). When two speech sounds led to the same viseme (for instance viseme e. Figure 2), a single speech sound was chosen. Each of the chosen vowel and consonant sounds were then placed in a minimal pair. Vowels were placed between two d's, yielding three minimal pair words: deed, dude and dyed. Likewise, consonants were placed at the beginning of words ending in 'in'. The words were: 'bin' and 'fin'.

Figure 2

Visemes and Corresponding Speech and Non-Speech Tasks

Viseme	Corresponding Speech Sound	Speech Task	Non-Speech Task
a. 	/a/	Dyed	”ah”
b. 	/i/	Deed	Full retraction
c. 	/u/	Dude	Purse
d. 	/f/	Fin	Bite lower lip completely
e. 	/b/	Bin	Push both lips together

Note. Visemes were modified from the figures presented in Bozkurt et. al. (2007).

The non-speech task consisted of maximal lip pursing, opening, retracting, compressing, and retracting postures. These tasks were designed to be the non-speech correlate to the visemes discussed above. For instance, viseme e. corresponded to bilabial sounds such as /b/. The non-speech correlate to this movement was pressing the lips together. This decision allowed for qualitative comparisons to be made between the speech and non-speech tasks.

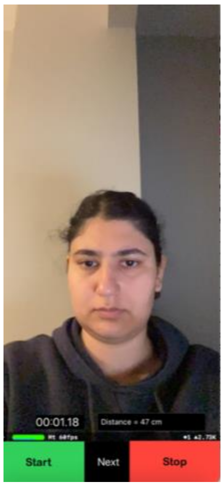
The participant produced each non-speech task five times then she completed the speech tasks five times, in the order listed above. As there was a break between each recording, due to the time taken to stop the recording and start the next recording, no order effects were expected. Furthermore, because the question in this study was whether the prototype could capture speech and non-speech kinematic movements, rather than commenting on the kinematics themselves, it was not necessary to randomize the order of the stimuli.

Procedure

The iPhone X was placed on a stand and was 113 cm off the ground for all recordings. The speech and non-speech tasks were recorded at a distance of 55 cm from the participant. This decision was motivated by the work done by Breitbarth et al. (2019), who found that stable measurements with a Truedepth camera were possible at 50 cm and above. The phone was angled such that the participant was facing the camera as seen in Figure 3. While the participant sat 55 cm from the phone's front facing camera, her head movement was not restricted. Therefore, throughout the course of recording, the participant moved up to 2 cm closer or farther from the screen.

Figure 3

Position of the Participant During Data Collection



Since the aim of this question was to demonstrate whether the prototype was capable of capturing kinematic data relevant to speech, points on the lips and the chin were tracked. The lips were tracked because they were the active articulators; the chin was tracked because of the contributions of the mentalis muscle and the mandible to lip movement (Hur et al., 2013; Sussman et al., 1973).

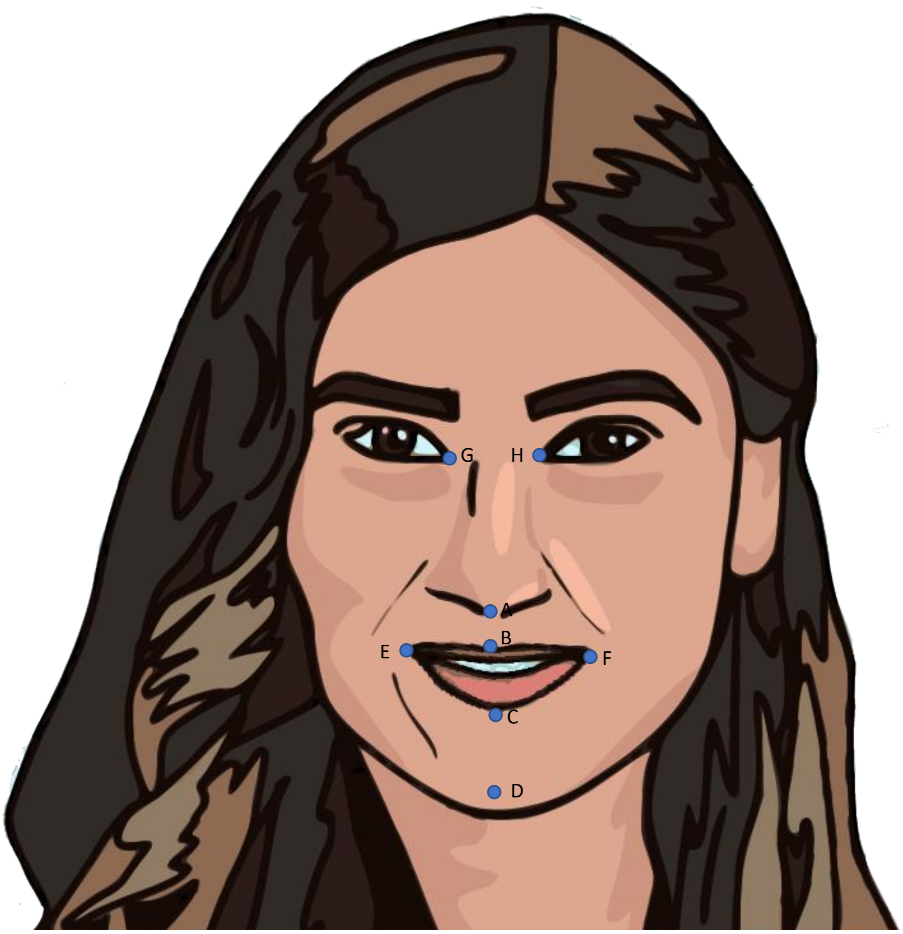
The 3-D Euclidean distances between articulators and a stable point (the nose) were calculated to ensure that the tracked points accounted for all three axes (refer to Figure 4 for the formula). Measuring the distance of active articulators from a stable point provided insight into the movement of the articulator only, as the nose was not expected to move.

Figure 4
3-D Euclidean Distance Formula

$$AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Note. The raw data from the prototype was converted to millimetres from metres prior to making any calculations—this was done by multiplying values by 1000.

Figure 5
Tracked Points for Question 1



Note. Tracked points are as follows: nose to upper lip (A-B); nose to lower lip (A-C); nose to chin (A-D); upper lip to lower lip (B-C); width of lips (E-F); between eyes (G-H). All points are at the midline except for E, F, G, and H.

Finally, the distance between the eyes were measured in order to demonstrate that the system was sensitive to static distances on the face. However, an unexpected secondary benefit to measuring the eyes was, that it demonstrated the sensitivity of the prototype to extremely small movements attributed to eye blinking. Points along the midline were chosen as this thesis aimed to demonstrate whether movements were captured generally rather than demonstrating a specific movement. In future studies, it is possible to measure the distances between any two points. Figure 5 demonstrated the portions of the face that were analyzed.

2.5.4. Question Two

Procedure

To answer question two, the iPhone was placed on the same stand used for question one according to the recording conditions listed in Table 3. Three factors, distance, angle, and eye opening, were compared. A recording according to the basic configuration, 55cm with the participant perpendicular to the camera, was compared to a recording where one of the factors had been modified.

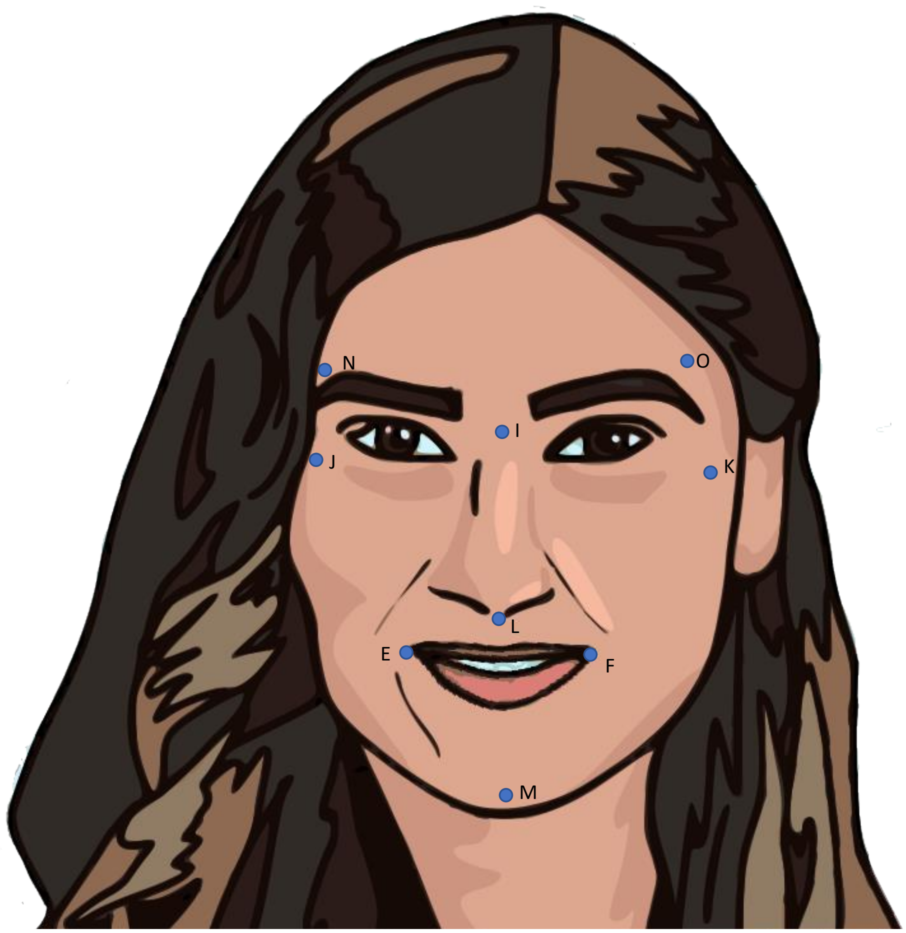
Each configuration was recorded once, all recordings were approximately 15 seconds long. As the prototype gathered positional data at every frame, the individual frames from each recording were taken as individual samples. Although, it is worth noting that these are not truly independent observations, because each frame is part of a single recording session. However, this procedure was chosen as it yielded between 715-923 frames for each condition. So, while statistical analyses were not conducted to determine the difference between points, standard deviations were used to analyze the differences in error between the groups.

Table 3
Recording Conditions and their Descriptions

Recording Condition	**Description**
Basic Configuration	- 55 cm - Participant facing camera - iPhone at participant's eye level (phone 109 cm off the ground) - Eyes open, blinking naturally
Distance Configuration	- 60 cm - Participant facing camera - iPhone at participant's eye level (phone 109 cm off the ground) - Eyes open, blinking naturally
Angle Configuration	- 55 cm - Participant facing camera - iPhone above participant's eye level (phone 113 cm off the ground) - Eyes open, blinking naturally
Eyes Closed Configuration	- 55 cm - Participant facing camera - iPhone at participant's eye level (phone 109 cm off the ground) - Eyes closed

The Euclidean distances between several anatomical landmarks on the face were measured from the recordings. Measurements were taken between the following anatomical points: Width of Lips (E-F); Face Length (I-M); Nose Length (I-L); Zygofrontale (G-H); Zygion (J-K), refer to Figure 12. The reason 3-D Euclidean distances were measured for these positions was twofold; first, at rest, distances between chosen landmarks were not expected to change, so measuring distances allowed the researcher to determine whether there had been a change. The second reason was for consistency with previous studies. The studies done by Zhuang et. al(2010) and Kovacs (2006) used the same anatomical landmarks, therefore the measurements in this thesis were consistent with previous literature.

Figure 6
Tracked Points for Question 2



Note. Tracked points are as follows: width of lips (E-F); face length (I-M); nose length (I-L); zygion (J-K); zygofrontale (N-O)

In order to determine the degree of recording error, the following benchmarks were used. In order to maintain methodological consistency between this thesis and previous literature, the benchmarks for recording error were taken from Aung et al. (1995) (accessed through Kovacs et al. (2006)). In their work, both Aung et al., and Kovacs et al. compared the accuracy of manual measurements to that of a face-tracker. They, therefore, defined error as the deviation between manual measurements and measures made from the face-tracking system. They defined a deviation of 2 mm or greater was deemed 'unreliable' while a deviation less than 1 mm was deemed 'highly accurate'. Deviations between 1 and 1.5 mm were deemed to be 'reliable', while 1.6 to 2 mm was termed a 'little reliable'. These decisions can also be understood by considering that the diameter of a human hair ranges from 0.017 mm to 0.181 mm, therefore, 1 mm error is the equivalent of roughly five human hairs (Ley, 1999).

The judgments about recording error were made in comparison to an external recording measure, the manual measurements, in Aung et al., (1995) and Kovacs et al. (2006). In this thesis, however, judgements about recording error were made using the standard deviation of a set of recordings made with the prototype face-tracking application. Since no external measurement was used, the benchmarks discussed above were applied to standard deviations. The reasoning behind using standard deviations was that the face was at rest, therefore the measured distances were not expected to change due to the participant's behavior. Therefore, variability in the measured distances tracked points on the face were thought to be attributable to the recording error in the system. Therefore, if there had been a small amount of error, a small standard deviation would be expected. Conversely, if there was a large amount of error in the system, a larger standard deviation would be expected.

Then, these benchmarks were applied to a recording where the participant was not at rest. In a recording of the phrase, 'buy Bobby a puppy', this standard deviation of the nasion point (point I in Figure 6) were judged for recording error. The nasion point was chosen because it was not expected to move as a result of the speech task, therefore the standard deviation of its position was used as a measure of the prototype's recording error. In addition to the nasion, the movement of the upper and lower lips were plotted during this task. Points B and C (Figure 6) were plotted separately (i.e., the 3-D Euclidean distance between the two was not measured). The vertical displacement of the three points (the nasion and the two lip points) relative to the origin point along the y-axis were plotted. The decision to plot the data in this way was made in order to highlight the range of movement of the lips and nasion along a single axis. Further, as the y-axis represents an up-down movement relative to the origin, this was the most salient movement for the bilabial sounds present in the phrase, 'buy Bobby a puppy'. Finally, the choice to demonstrate recording error in this phrase was made in order to be consistent with previous literature (Walsh et al., 2015).

2.6. Results and Discussion

2.6.1. Question One: Results

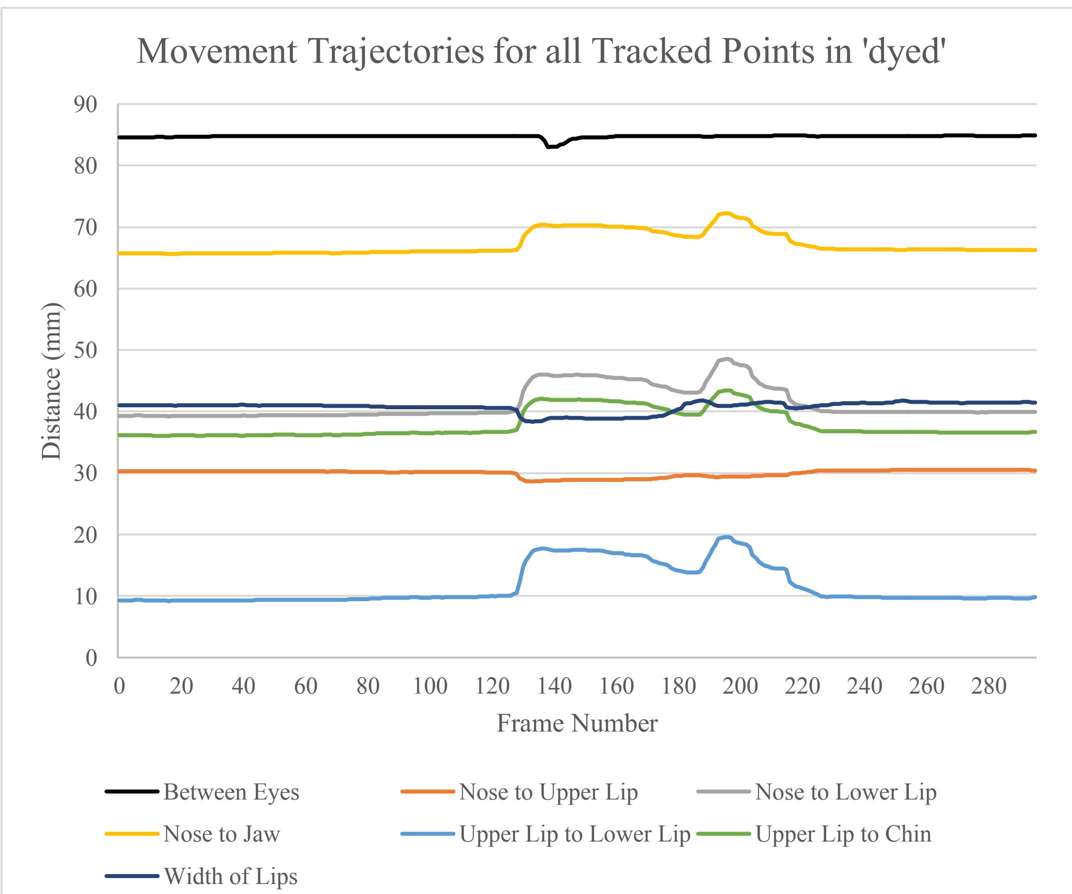
Overall, twenty-five recordings were collected for question one. Representative examples from select productions were discussed in this section, for the speech tasks, the word 'dyed' was analyzed. For the non-speech tasks, the 'ah' condition was analyzed. The movement trajectories for all recordings are in Appendix VI.

The three criteria for success of the prototype (summarized below) were evaluated in the speech and non-speech tasks. The criteria were: 1) temporal alignment of various articulators within a recording; 2) consistent movement patterns across recordings of the same stimuli; 3) the ability to represent a range of kinematic measures. Overall, the prototype was successful in tracking the face throughout all speech recordings. The prototype was successful for the majority of non-speech tasks as well. However, select recordings from the 'bite lip' condition and the 'hide lip' condition experienced recording error from an unknown source (Appendix VI).

Criterion 1: Temporal Alignment

Figure 7

Movement Trajectories 'dyed'

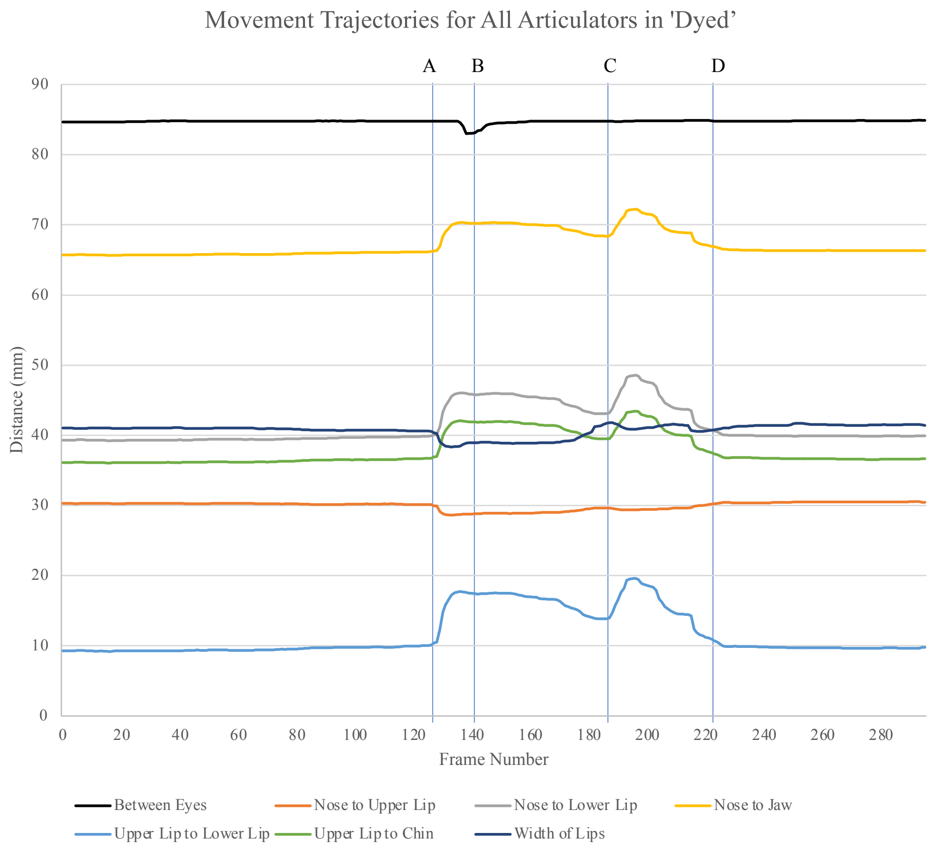


Note. Euclidean distances of all tracked points (nose to upper lip/lower lip/ chin; upper lip to lower lip; width of lips—according to figure 12 above) for one recording of the word "dyed".

The movement trajectories for all tracked points were plotted in Figure 7. The start and end points of these same trajectories were labeled in Figure 8. This figure demonstrated that the temporal alignment criteria was met for the speech tasks, as there is a clear coordination among all articulators—movement for all articulators begin at Point A and end at Point D.

Figure 8

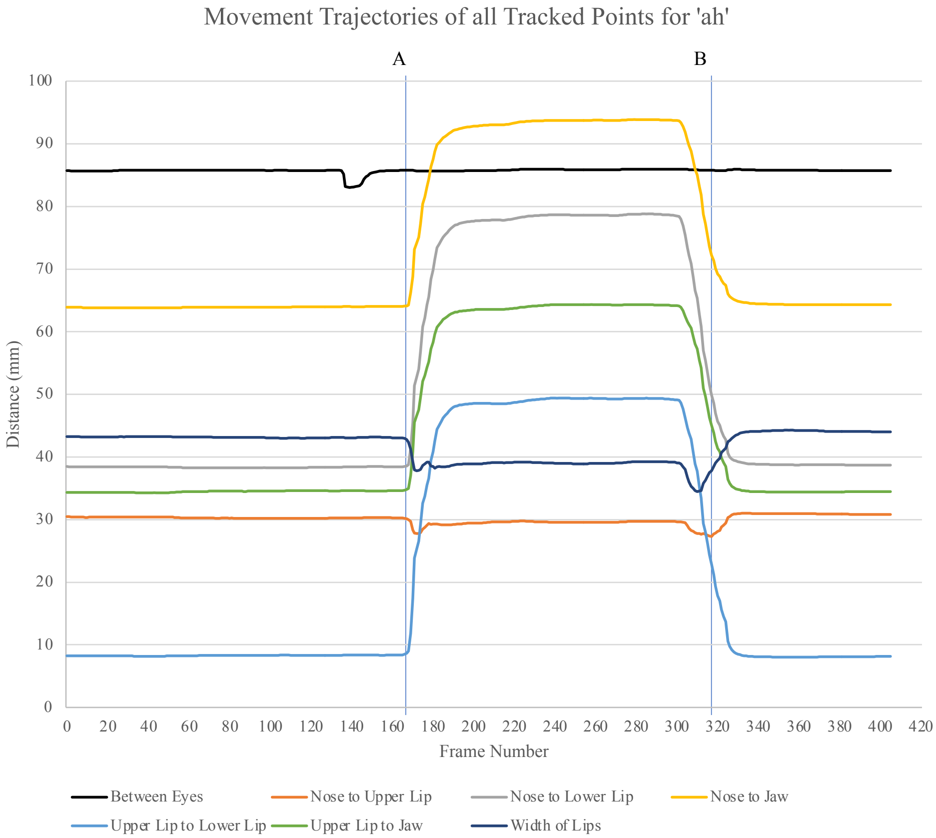
Labeled Movement Trajectories for 'dyed'



Note. Movement trajectories for articulators in a production of 'dyed'. Movement starts at Point A and ends at Point D. Points A and D were labeled through visual analysis of overall movement trajectories between measured points and are meant to be illustrative rather than delineate exact start and end points.

There was also temporal alignment in the movement trajectories for the articulators in non-speech task, ‘ah’. The labeled figure, Figure 9 showed that the movements started at Point A and ended at Point B.

Figure 9
Labeled Movement Trajectories for ‘ah’



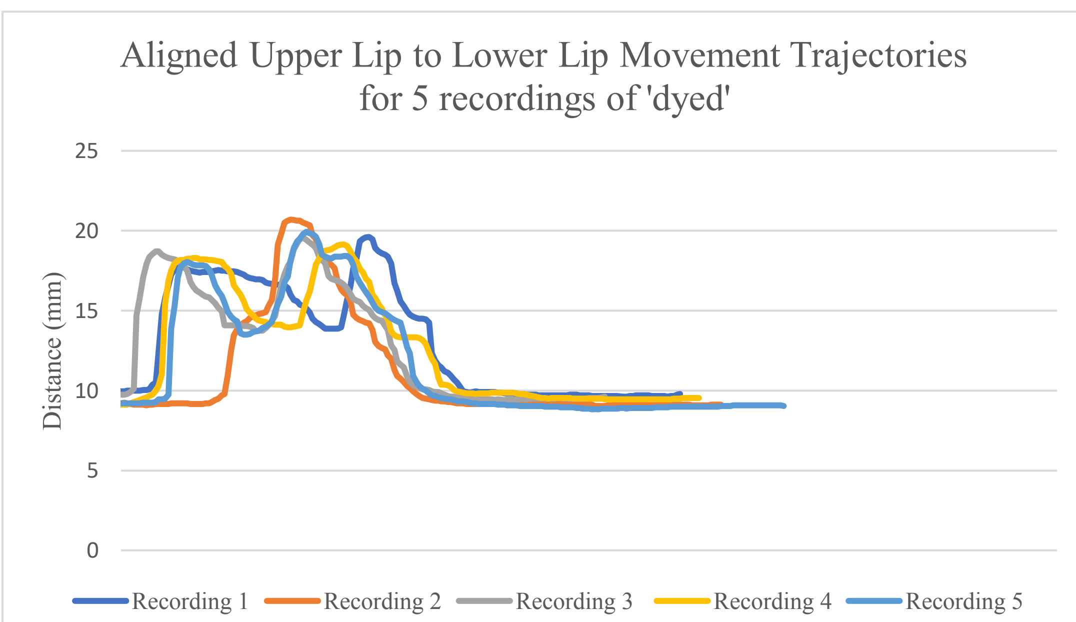
Note. Euclidean distances of all tracked points (nose to upper lip/lower lip/ chin; upper lip to lower lip; width of lips) for one recording of the non-speech task, ‘ah’.

Criterion 2: Consistent Movement Patterns Across Recordings

In order to demonstrate that the prototype had met this criterion in the speech task condition, the upper lip to lower lip movement trajectories were plotted across repeated productions of, 'dyed'. The movement pattern for the upper lip to lower lip was illustrated in Figure 8. The movement trajectory moved according to the following pattern: 1) movement began at Point A; 2) the lips were farthest from each other before Point B and shortly after Point C; 3) at Point C the lips approximated each other.

Though the timing between the two movement peaks varied across recordings, the aligned movement trajectories for five recordings of 'dyed' in Figure 10 illustrated that there was consistency in movements between recordings. The difference in timing and degree of motion captured in Figure 16 illustrated how the prototype could be used to capture difference among participants in future research.

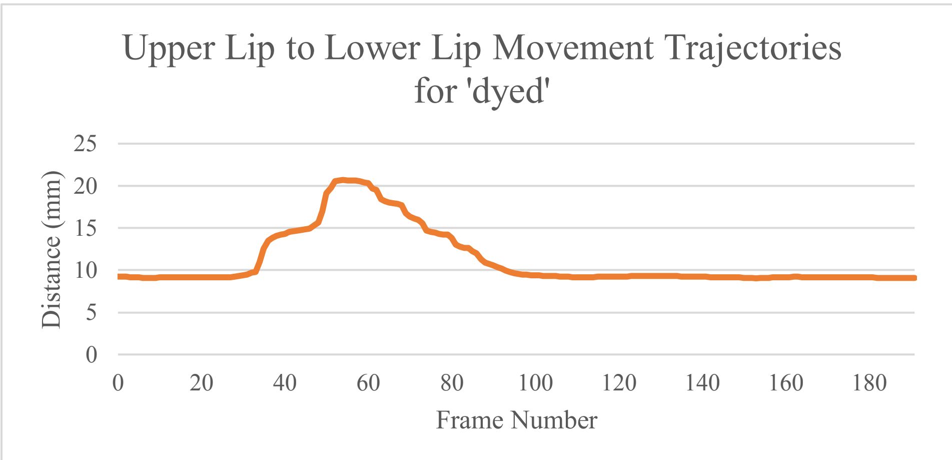
Figure 10
Aligned Movement Trajectories for Five Recordings of 'dyed'



Note. The movement trajectories across 5 recordings of 'dyed' were visually aligned based on moment that the movement began to show similarities in patterns. Recordings were aligned by removing the initial moments of rest (i.e., a series of frames before movement) so that the start points for all recordings were aligned. Therefore, frame numbers were not listed in this graph.

Figure 11

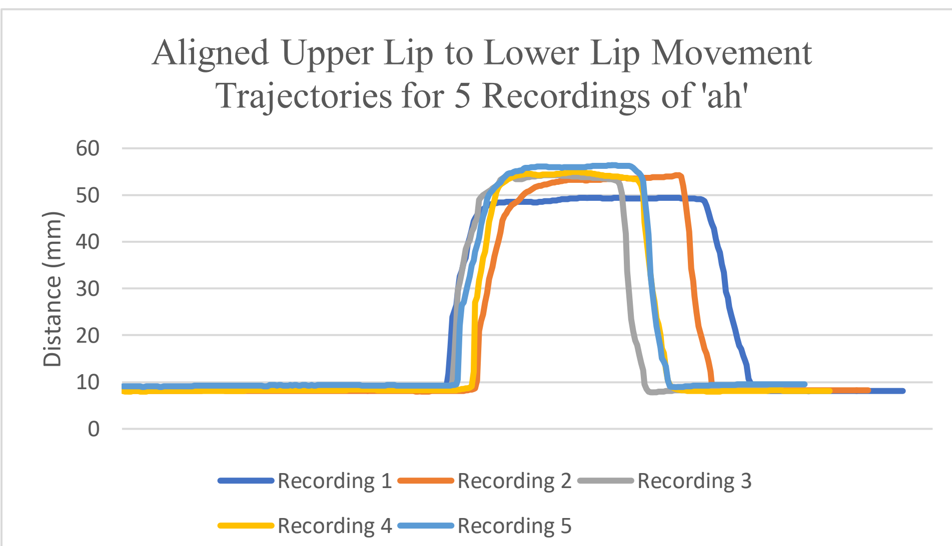
Upper Lip to Lower Lip Movement for One Recording of 'dyed'



Note. Recording number 2 demonstrates a difference in movement pattern compared to other recordings of the same word, 'dyed'.

Figure 12

Aligned Movement Trajectories for Five Recordings of 'ah'



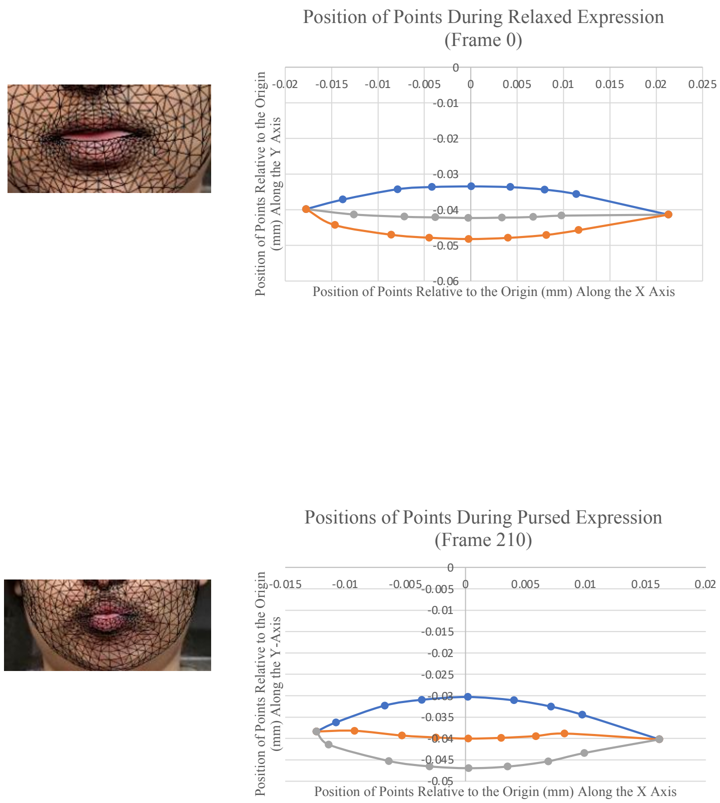
Note. Movement trajectory for Upper Lip to Lower Lip for repeated measures of the non-speech task, 'ah'. Recordings were aligned by removing the initial moments of rest (i.e., a series of frames before movement) so that the start points for all recordings were aligned. Therefore, frame numbers were not listed in this graph.

Consistency in movement patterns for repeated measures of the non-speech task were also demonstrated. The movement trajectories of the upper lip to lower lip across five recordings of ‘ah’ were presented in Figure 12. Movement trajectories across repeated measures of ‘ah’ followed the same movement pattern (maximal opening, demonstrated by the increase in distance between the lips). While the patterns between recordings were consistent, there were still differences in duration and velocity of movements, consistent with the observations made in the speech tasks.

Criterion 3: Representing Kinematic Measures

The figures presented for criteria one and two (Figures 7-12) demonstrated that the prototype can be used to represent kinematic data through movement trajectories. In addition to these movement trajectories, the prototype was able to capture visemes at specific frames. This was illustrated in the positional curve plotted in Figure 13. The tracked points around the lips were plotted at two frames, frame 0 and frame 210. These frames were chosen based on the Lip Width movement trajectory. At frame 0, the lips were at rest; and, at frame 210, the lips were maximally pursed. The vertical and horizontal displacement from the origin (along the x and y axes) for the points around the lips were plotted at these frames. The corresponding pictures were examples of the relaxed and pursed shapes of the lips for comparison. This method provided another way that kinematic data could be represented with this prototype. This is useful for future researchers.

Figure 13
Positional Curves for Two Frames



Note. At frame 0, the lips are more oblong in both the graph and the picture. At frame 210, the lips are rounded and the edges are closer together, this is seen in both the graph and the picture.

The prototype also provided information about the velocity and amplitude of movements. For instance, movement amplitudes were extracted from the recording of ‘dyed’ (Figure 7), these were presented in Table 4. Movement amplitudes have the potential to be used in future research to investigate the causes and extent of coupling between articulators.

Table 4
Maximum Amplitudes for ‘dyed’

Measurement	Amplitude
Upper Lip to Lower Lip	7.6 mm
Upper Lip to Chin	5.3 mm
Nose to Lower Lip	5.3 mm
Nose to Upper Lip	-1.0 mm

Note. Based on the labeled figure of the word ‘dyed’ (Figure 14), Point A (frame 0) was taken as the baseline point for each measurement and Point B (frame 245) was taken as the point of highest amplitude. The differences between Points A and B were presented here to represent the greatest amplitude.

Finally, because the position of the 1220 points were recorded at each frame, movement velocities for articulators were calculated from frame to frame. For instance, in a recording of a participant saying “dyed” and the non-speech correlate task, “ah”, movement velocities from frame to frame were calculated by calculating the difference between consecutive frames. Once the frame-by-frame velocity was calculated for each recording, it was possible to compare the maximum velocities of the recordings. In this case, the maximum velocity for the non-speech task was 11.5 (range: 20.0, SD 1.0) and the maximum velocity for the speech task was 4.2 (range: 5.6, SD 0.3).

2.6.2. Incidental Findings

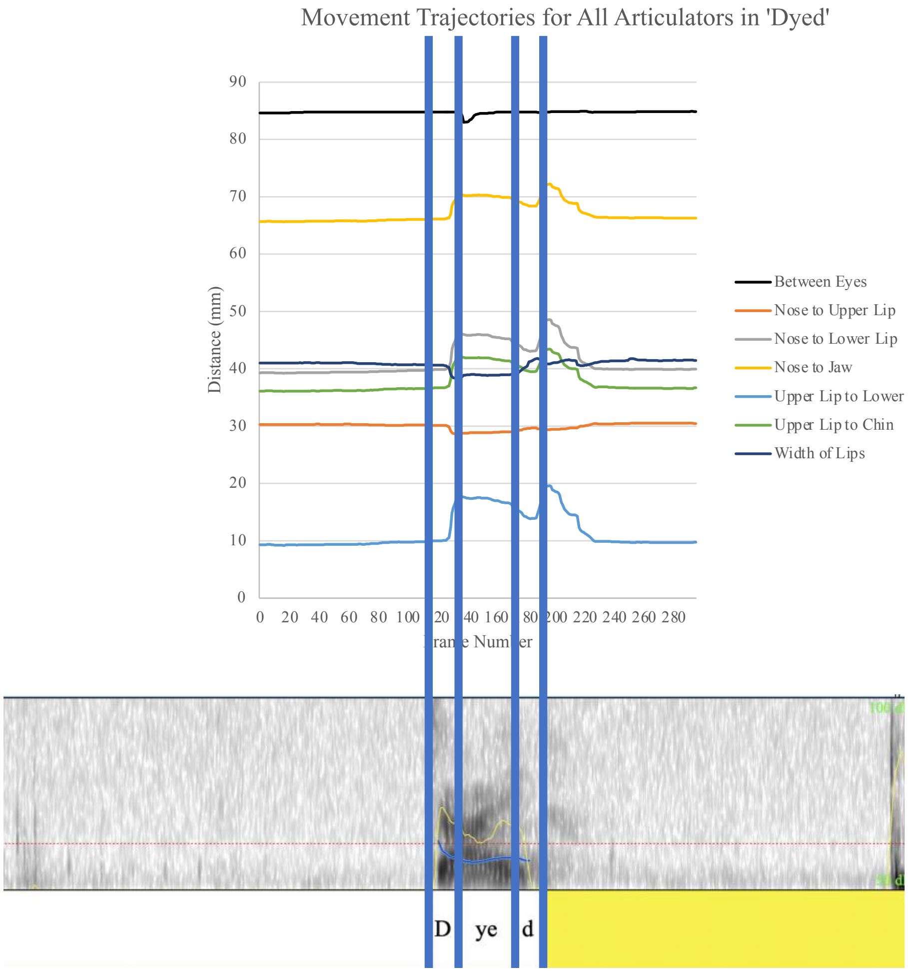
While the aim of this question was to show whether the prototype was capable of tracking speech and non-speech data, two incidental findings were presented here. The first finding, discussed in the following subsection (eye blinks), provided insight into a method that could be used to align face-tracking data and A/V data. The second finding discussed below (non-speech movement), provided an intriguing avenue for future research.

Eye Blinks

While the distance between the eyes were thought to be stable, the data indicated that the prototype was capable of tracking eye blinks. In Figure 8, the participant's blink was captured at point B, this was seen in the brief decrease in distance in the Between Eyes measurement. This information was used, in conjunction with the video file, to correctly align the face-tracking data and the acoustic data. The video revealed that the participant had blinked prior to speaking. Therefore, the data was aligned in Figure 14

Aligning the data provided context to movement trajectories. For example, it became apparent that the participant parted her lips most when producing /d/. It was also apparent that the width of the speaker's lips was the smallest when her mouth was the most open.

Figure 14
Aligned Face-Tracking and Audio Data



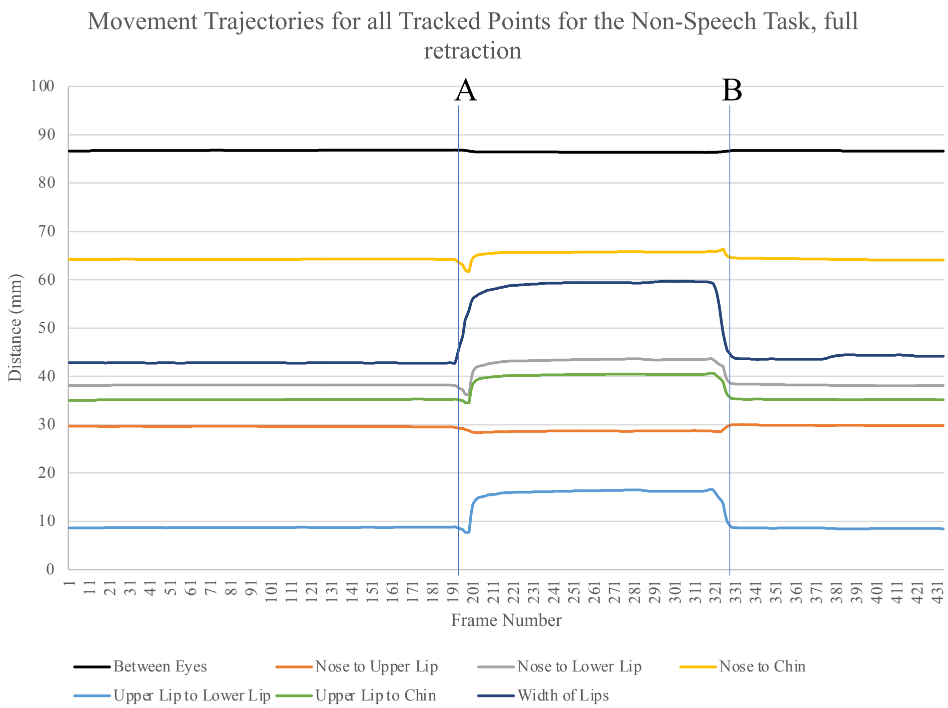
Non-Speech Movement

The incidental finding discussed in this section was a movement pattern identified in the non-speech tasks. It was not known whether this pattern was attributable to error in the prototype, or whether the prototype was capturing a particular behavior.

This pattern is evident in the movement trajectories of all tracked points for the ‘full retraction’ task (Figure 15). In this figure, two time points were highlighted; Point A occurred before the non-speech task, and Point B occurred after the task. At Point A, the distances briefly decreased before maximally increasing in the upper lip to lower lip, nose to lower lip, upper lip to chin, and nose to chin measurements. The opposite trajectory was seen in the nose to chin measurement at Point B. Across five recordings of this non-speech task (Figure 16) the same pattern of decreased distance before the behavior and increased distances following the behavior, was seen for the nose to chin measurement.

Figure 15

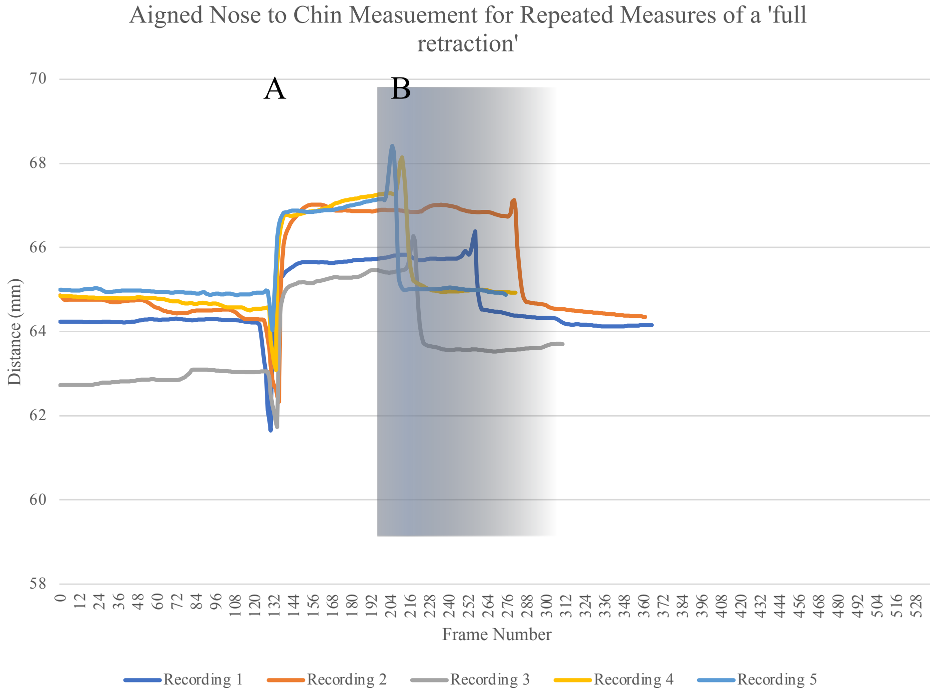
Labeled Movement Trajectories for 'full retraction'



Note. Movement trajectories for all tracked points for the non-speech task, full retraction. Point A is the beginning of the movement and Point B is the end.

Figure 16

Aligned Movement Trajectories for ‘full retraction’



Note. Movement trajectories for the Nose to Chin measurement across 5 recordings of a full retraction task. Point A indicates that all recordings have an initial decreased distance before the movement, then Point B indicates that all recordings have a final increase at the end of the movement. Though the trajectories were aligned, the movements did not occur for identical spans of time, therefore, Point B is longer than Point A, this was done to simply illustrate the pattern across recordings.

The corresponding video files were inspected for any apparent downwards chin movement prior the non-speech task; however, no such movement was discernable. Therefore, the exact reason for this movement trajectory is not known, but it is possible that the prototype is capturing muscular activation. The downwards chin movement at Point A is possibly attributed to the activation of the mentalis muscle as the participant lowers their lower lip and begins the lateral lip spread. In this case, the decreased distances would represent muscular flexion, which would bring the points closer together. It is also possible that these observations were attributable to an unknown recording error in the prototype.

Further investigation of this incidental finding is warranted to determine the cause. For instance, movement data could be collected from the surface of the face, using the prototype, and from the muscles in the face, using an EMG and compared. Regardless of the cause, this pattern was consistently visible in these tasks and may be used as a marker for this movement. That is, the presence of these rapid movements may have indicated that the participant was initiating a maximum lip spread movement. Appendix VI shows that these movements were visible in other tasks too.

2.6.3. Question 2: Error attributed to recording conditions

The benchmarks for error discussed above, were applied to each of the recording conditions for all measured distances. Across all measurements, the recordings were 'highly accurate' The standard deviations for the measured points across the different recording conditions were presented in Table 5.

Table 5
Standard Deviations of Measured Portions by Recording Condition

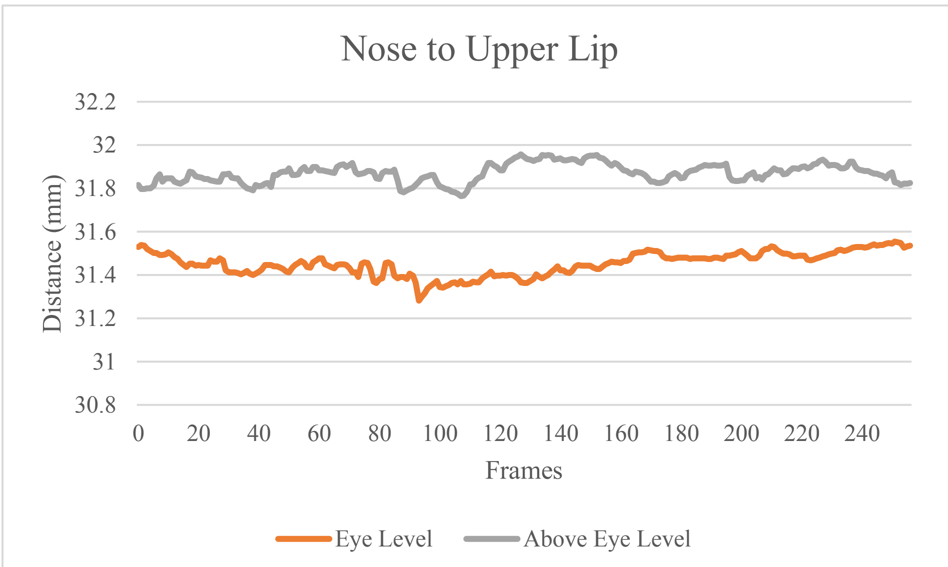
Recording Condition	Measured Portions	Standard Deviation (mm)
Basic Configuration	Width of lips	0.08
	Face length	0.05
	Nose length	0.04
	Zygofrontale	0.08
	Zygion	0.07
	Nose to upper lip	0.05

Distance Configuration	Width of lips	0.05
	Face length	0.09
	Nose length	0.04
	Zygofrontale	0.11
	Zygion	0.03
	Nose to upper lip	0.09
Angle Configuration	Width of lips	0.08
	Face length	0.08
	Nose length	0.02
	Zygofrontale	0.02
	Zygion	0.04
	Nose to upper lip	0.04
Eyes Closed Configuration	Width of lips	0.10
	Face length	0.10
	Nose length	0.50
	Zygofrontale	0.09
	Zygion	0.08
	Nose to upper lip	0.20

Note. Standard deviations of all measured points across recording conditions.

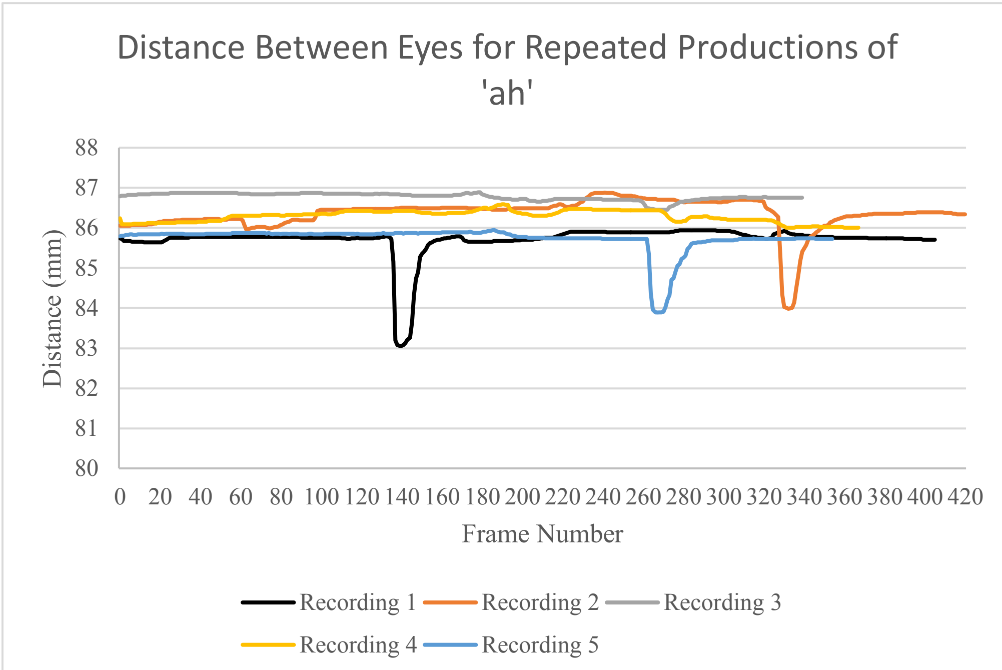
Though the level of error was low, the differences between two recording angles was evident when movement trajectories were plotted. Figure 21 demonstrated that when the recording angles were different, the distances between the same points of the face at rest appeared different. While it was possible that this observation was attributable to error in the system, it was more likely that this difference was due to the modification to the recording angle. Figure 22 showed the movement trajectories for the Between Eyes measurements in five recordings of the non-speech task, ‘ah’. Although the angle of the phone was kept constant for all five recordings, the differences in head positioning when the speaker produced the stimulus each time were slightly different. This difference was visible in the data. Therefore, future researcher may benefit from controlling participants' head positioning as well. Or they may be able to adjust the data to account for this level of error.

Figure 17
Nose to Upper Lip Distance Measures Measured at Different Recording Angles



Note. Graph enlarged for clarity.

Figure 18
Distance Between Eyes Measured at Different Participant Angles

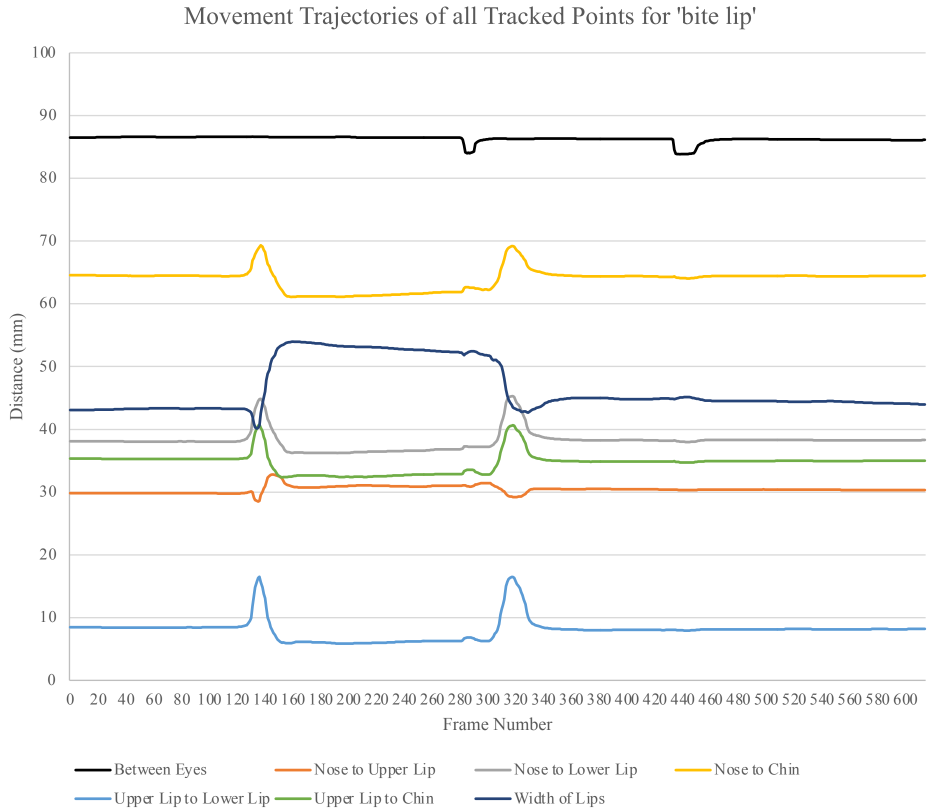


Note. Distance between eyes distances for five recordings of the non-speech task, 'ah'. Each recording was measured at a slightly different angle (participant's head relative to the camera). Graph was enlarged for clarity.

Likewise, the error attributed to eye blinks was found to be low. However, the effects of the blinks could be seen across all measurements, which indicated that the blinking may lead to some measurement error. In Figure 19, the participants' blink led to a short decrease in distance between the eyes. At the same time, there was a short increase or decrease in all other measurements as well. While a behavioral reason for this observation could not be ruled out, it was more likely that it indicated measurement error. The second eye blink in the data indicated that the error in the articulator measurements did not only occur when movement was present. Further investigation is needed to determine how large the effect of the eye blinks is on other tracked points.

Figure 19

Error Attributed to Eye Blinks for 'bite lip'



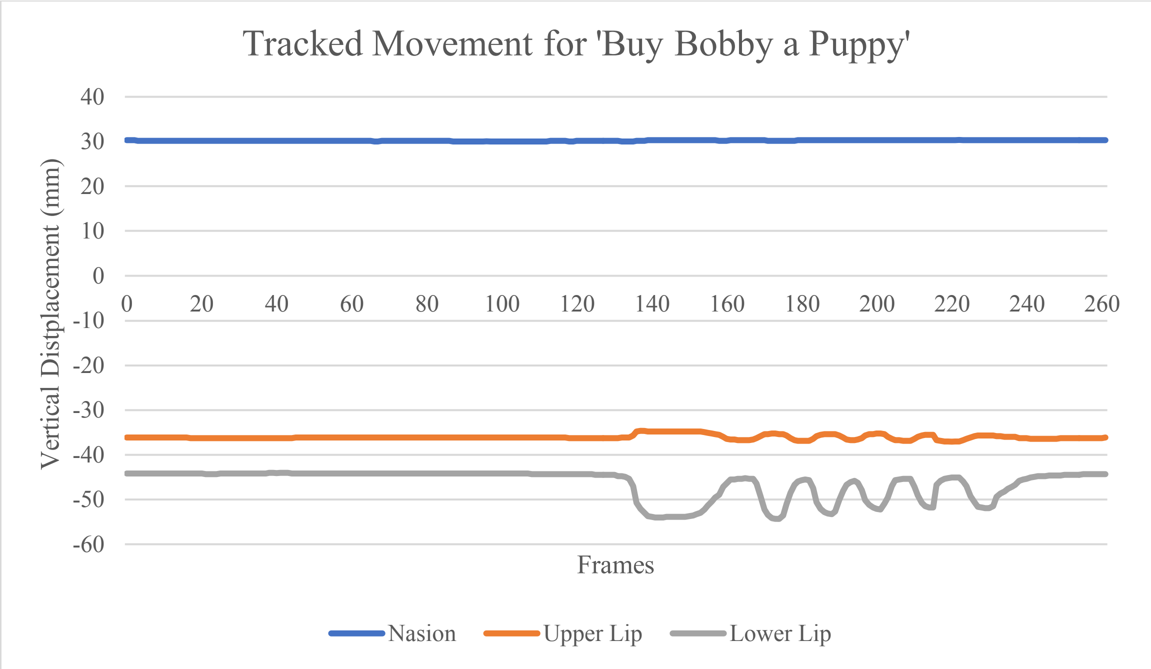
Note. The moments the eye blinks occurred were highlighted in blue.

2.6.4. Recording Error During Speech

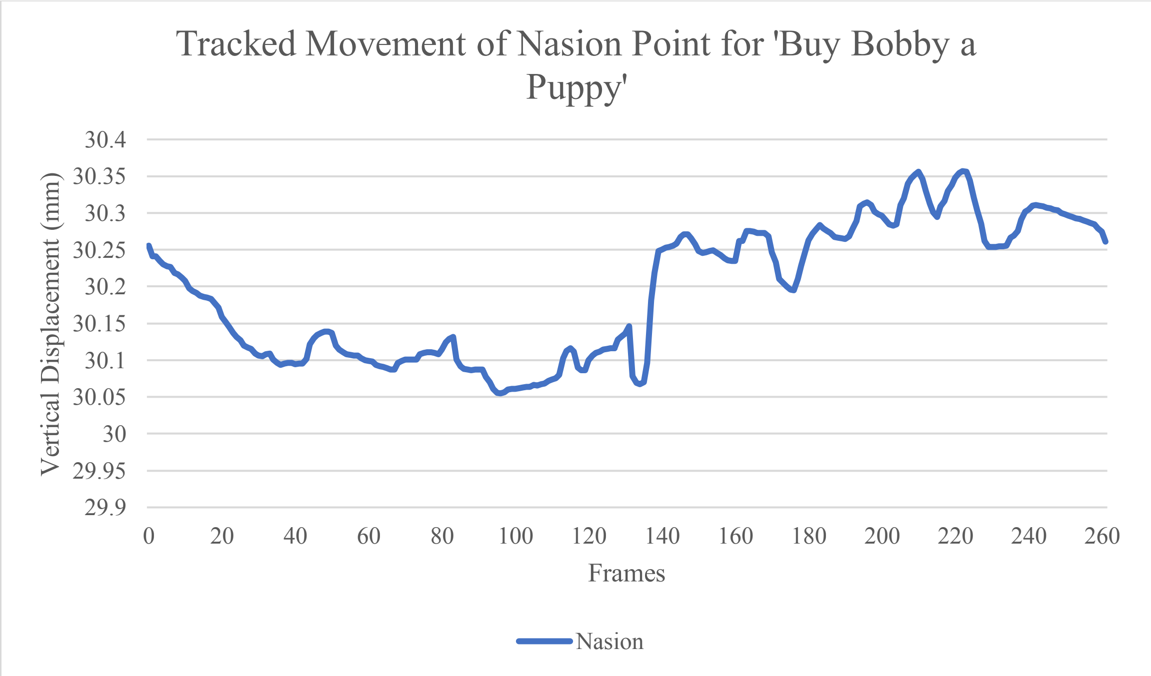
Figure 20

Tracked Movement in a Recording of, 'buy Bobby a puppy'

a.



b.



In the tracked movement points for the recording of ‘buy Bobby a puppy’, clear upper lip and lower lip motion were observed. These motions followed the expectations that the upper and lower lips would approximate each other during the bilabial sounds (/p/ and /b/) in the phrase. The movement of the nasion point, which was tracked as a measure of recording error during a speech task, had a range of 3.0 mm of movement (min: 30.1 mm; max: 30.4 mm). The standard deviation of the movement for this point was 0.09 mm. Therefore, based on the benchmarks for recording error, this task illustrated that the prototype was highly reliable.

While this thesis investigated some potential sources for recording error, others may still exist. For instance, in Appendix VI, Recording 2 for the hide lip task, had undergone significant recording error. However, further investigation is needed to determine the source of this error as recordings 1, 3, 4 and 5 in this condition did not experience the same level of error.

2.7. Final Conclusions

This thesis served to advance the field of face-tracking for research by designing a research study to track compensatory articulations in glossectomy speech in Chapter 1. This research design provides a highly controlled methodology for tracking articulatory behaviors using a prototype face-tracking application. In Chapter 2, a qualitative analysis of the prototype face-tracking application demonstrated the feasibility of a novel face-tracking method and explored sources of measurement error.

References

Abendstein, H., Nordgren, M., Boysen, M., Jannert, M., Silander, E., Ahlner-Elmqvist, M., Hammerlid, E., & Bjordal, K. (2005). Quality of life and head and neck cancer: A 5 year prospective study. *The Laryngoscope*, *115*(12), 2183–2192. https://doi.org/10.1097/01.MLG.0000181507.69620.14

Acher, A., Perrier, P., Savariaux, C., & Fougeron, C. (2014). Speech production after glossectomy: Methodological aspects. *Clinical Linguistics & Phonetics*, *28*(4), 241–256. https://doi.org/10.3109/02699206.2013.802015

Ansarin, M., Bruschini, R., Navach, V., Giugliano, G., Calabrese, L., Chiesa, F., Medina, J. E., Kowalski, L. P., & Shah, J. P. (2019). Classification of GLOSSECTOMIES: Proposal for tongue cancer resections. *Head & Neck*, *41*(3), 821–827. https://doi.org/10.1002/hed.25466

Argiris, A., Karamouzis, M. V., Raben, D., & Ferris, R. L. (2008). Head and neck cancer. *The Lancet.*, *371*(9625), 1695–1709. https://doi.org/10.1016/S0140-6736(08)60728-X

Aung, S. C., Ngim, R. C. K., & Lee, S. T. (1995). Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements. *British Journal of Plastic Surgery*, *48*(8), 551–558. https://doi.org/10.1016/0007-1226(95)90043-8

Bandini, A., Green, J. R., Taati, B., Orlandi, S., Zinman, L., & Yunusova, Y. (2018). Automatic Detection of Amyotrophic Lateral Sclerosis (ALS) from Video-Based Analysis of Facial Movements: Speech and Non-Speech Tasks. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 150–157. https://doi.org/10.1109/FG.2018.00031

Barry, W. J., & Timmermann, G. (1985). Mispronunciations and compensatory movements of tongue-operated patients. *The British Journal of Disorders of Communication*, *20*(1), 81–90.

Bennett, J. W., van Lieshout, P. H. H. M., & Steele, C. M. (2007). Tongue control for speech and swallowing in healthy younger and older subjects. *The International Journal of Orofacial Myology: Official Publication of the International Association of Orofacial Myology*, *33*, 5–18.

Bhayani, M. K., Holsinger, F. C., & Lai, S. Y. (2010). A shifting paradigm for patients with head and neck cancer: Transoral robotic surgery (TORS). *Oncology*, *24*(11), 1010. Gale Academic OneFile.

Biglioli, F., Liviero, F., Frigerio, A., Rezzonico, A., & Brusati, R. (2006). Function of the sensate free forearm flap after partial glossectomy. *Journal of Cranio-Maxillofacial Surgery*, *34*(6), 332–339. https://doi.org/10.1016/j.jcms.2006.04.005

Blyth, K. M., McCabe, P., Madill, C., & Ballard, K. J. (2015). Speech and swallow rehabilitation following partial glossectomy: A systematic review. *International Journal of Speech-Language Pathology*, *17*(4), 401–410. https://doi.org/10.3109/17549507.2014.979880

Borduas, J., Castonguay, A., Laurin, P., & Beland, D. (2020, November 17). Reliability of Mobile 3D Scanning Technologies for the Customization of Respiratory Face Masks. *Proceedings of 3DBODY.TECH 2020 - 11th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Online/Virtual, 17-18 November 2020*. 3DBODY.TECH 2020 - 11th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Online/Virtual, 17-18 November 2020, Online. https://doi.org/10.15221/20.34

Borghese, N. A., Ferrigno, G., Redolfi, M., & Pedotti, A. (1997). Automatic integrated analysis of jaw and lip movement in speech production. *The Journal of the Acoustical Society of America*, *101*(1), 482–487. https://doi.org/10.1121/1.419490

Bozkurt, E., Erdem, C. E., Erzin, E., Erdem, T., & Ozkan, M. (2007). Comparison of Phoneme and Viseme Based Acoustic Units for Speech Driven Realistic lip Animation. *2007 3DTV Conference*, 1–4. https://doi.org/10.1109/3DTV.2007.4379417

Breitbarth, A., Schardt, T., Kind, C., Brinkmann, J., Dittrich, P.-G., & Notni, G. (2019). Measurement accuracy and dependence on external influences of the iPhone X TrueDepth sensor. In B. Zagar, P. Mazurek, M. Rosenberger, & P.-G. Dittrich (Eds.), *Photonics and Education in Measurement Science 2019* (p. 7). SPIE. https://doi.org/10.1117/12.2530544

Bressmann, T., Jacobs, H., Quintero, J., & Irish, J. C. (2009). Speech Outcomes for Partial Glossectomy Surgery: Measures of speech articulation and listener perception. *CJSLPA*, *33*(02), 7.

Bressmann, T., Sader, R., Whitehill, T. L., & Samman, N. (2004). Consonant intelligibility and tongue motility in patients with partial glossectomy. *Journal of Oral and Maxillofacial Surgery*, *62*(3), 298–303. https://doi.org/10.1016/j.joms.2003.04.017

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, *49*(3–4), 155–180. https://doi.org/10.1159/000261913

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Buchaillard, S., Brix, M., Perrier, P., & Payan, Y. (2007). Simulations of the consequences of tongue surgery on tongue mobility: Implications for speech production in post-surgery conditions. *The International Journal of Medical Robotics and Computer Assisted Surgery*, *3*(3), 252–261. https://doi.org/10.1002/rcs.142

Campbell, B. H., Marbella, A., & Layde, P. M. (2000). Candidate's Thesis: Quality of Life and Recurrence Concern in Survivors of Head and Neck Cancer. *The Laryngoscope*, *110*(6), 895–906. https://doi.org/10.1097/00005537-200006000-00003

Chaturvedi, A. K., Engels, E. A., Pfeiffer, R. M., Hernandez, B. Y., Xiao, W., Kim, E., Jiang, B., Goodman, M. T., Sibug-Saber, M., Cozen, W., Liu, L., Lynch, C. F., Wentzensen, N., Jordan, R. C., Altekruse, S., Anderson, W. F., Rosenberg, P. S., & Gillison, M. L. (2011). Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of Clinical Oncology: Official Journal of the*

American Society of Clinical Oncology, *29*(32), 4294–4301. https://doi.org/10.1200/JCO.2011.36.4596

Chow, L. Q. M. (2020). Head and Neck Cancer. *New England Journal of Medicine*, *382*(1), 60–72. https://doi.org/10.1056/NEJMra1715715

Chuanjun, C., Zhiyuan, Z., Shaopu, G., Xinquan, J., & Zhihong, Z. (2002). Speech after partial glossectomy: A comparison between reconstruction and nonreconstruction patients. *Journal of Oral and Maxillofacial Surgery*, *60*(4), 404–407. https://doi.org/10.1053/joms.2002.31228

Constantinescu, G., Rieger, J., Winget, M., Paulsen, C., & Seikaly, H. (2017). Patient Perception of Speech Outcomes: The Relationship Between Clinical Measures and Self-Perception of Speech Function Following Surgical Treatment for Oral Cancer. *American Journal of Speech-Language Pathology*, *26*(2), 241–247. https://doi.org/10.1044/2016_AJSLP-15-0170

Crombie, A. K., Farah, C. S., & Batstone, M. D. (2014). Health-related quality of life of patients treated with primary chemoradiotherapy for oral cavity squamous cell carcinoma: A comparison with surgery. *British Journal of Oral and Maxillofacial Surgery*, *52*(2), 111–117. https://doi.org/10.1016/j.bjoms.2013.09.014

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 600 Million Words, 1990-Present. https://www.english-corpora.org/coca/

de Vicente, J. C., de Villalaín, L., Torre, A., & Peña, I. (2008). Microvascular Free Tissue Transfer for Tongue Reconstruction After Hemiglossectomy: A Functional Assessment of Radial Forearm Versus Anterolateral Thigh Flap. *Journal of Oral and Maxillofacial Surgery*, *66*(11), 2270–2275. https://doi.org/10.1016/j.joms.2008.01.018

Dios, P. D., Feijoo, J. Fdez., Ferreiro, M. C., & Alvarez, J. A. (1994). Functional consequences of partial glossectomy. *Journal of Oral and Maxillofacial Surgery*, *52*(1), 12–14. https://doi.org/10.1016/0278-2391(94)90005-1

Dirix, P., Nuyts, S., & Bogaert, W. V. den. (2006). Radiation-induced xerostomia in patients with head and neck cancer. *Cancer*, *107*(11), 2525–2534. https://doi.org/10.1002/cncr.22302

Do, K.-A., Johnson, M. M., Doherty, D. A., Lee, J. J., Wu, X. F., Dong, Q., Hong, W. K., Khuri, F. R., Fu, K. K., & Spitz, M. R. (2003). Second primary tumors in patients with upper aerodigestive tract cancers: Joint effects of smoking and alcohol (United States). *Cancer Causes & Control: CCC*, *14*(2), 131–138. https://doi.org/10.1023/a:1023060315781

D’Souza, G., Kreimer, A. R., Viscidi, R., Pawlita, M., Fakhry, C., Koch, W. M., Westra, W. H., & Gillison, M. L. (2007). Case-control study of human papillomavirus and oropharyngeal cancer. *The New England Journal of Medicine*, *356*(19), 1944–1956. https://doi.org/10.1056/NEJMoa065497

Duvvuri, U., Simental, A. A., D’Angelo, G., Johnson, J. T., Ferris, R. L., Gooding, W., & Myers, E. N. (2004). Elective neck dissection and survival in patients with squamous cell carcinoma of the oral cavity and oropharynx. *The Laryngoscope*, *114*(12), 2228–2234. https://doi.org/10.1097/01.mlg.0000149464.73080.20

Dwivedi, R. C., Kazi, R. A., Agrawal, N., Nutting, C. M., Clarke, P. M., Kerawala, C. J., Rhys-Evans, P. H., & Harrington, K. J. (2009). Evaluation of speech outcomes following treatment of oral and oropharyngeal cancers. *Cancer Treatment Reviews*, *35*(5), 417–424. https://doi.org/10.1016/j.ctrv.2009.04.013

Dziegielewski, P. T., Rieger, J., Shama, M. A., O'Connell, D. A., Harris, J. R., & Seikaly, H. (2019). Beavertail modification of the radial forearm free flap in total oral glossectomy reconstruction: Technique and functional outcomes. *Oral Oncology*, *96*, 71–76. https://doi.org/10.1016/j.oraloncology.2019.07.004

Dzioba, A., Aalto, D., Papadopoulos-Nydam, G., Seikaly, H., Rieger, J., Wolfaardt, J., Osswald, M., Harris, J. R., O'Connell, D. A., Lazarus, C., Urken, M., Likhterov, I., Chai, R. L., Rauscher, E., Buchbinder, D., Okay, D., Happonen, R.-P., Kinnunen, I., Irjala, H., … Laine, J. (2017). Functional and quality of life outcomes after partial glossectomy: A multi-institutional longitudinal study of the head and neck research network. *Journal of Otolaryngology - Head & Neck Surgery*, *46*(1), 56. https://doi.org/10.1186/s40463-017-0234-y

Engelbrecht, L., Van der Merwe, A., & Pretorius, J. P. (2006). The surgical management and speech and swallowing rehabilitation of patients with advanced tongue cancer in South Africa. *Health SA Gesondheid*, *11*(3). https://doi.org/10.4102/hsag.v11i3.233

Fagan, J. (2014). Partial glossectomy. *Open Access Atlas of Otolaryngology, Head & Neck Operative Surgery*, 13.

Fischbein, N. J., Kaplan, M. J., Jackler, R. K., & Dillon, W. P. (2001). *MR Imaging in Two Cases of Subacute Denervation Change in the Muscles of Facial Expression*. 5.

Forte, T., Niu, J., Lockwood, G. A., & Bryant, H. E. (2012). Incidence trends in head and neck cancers and human papillomavirus (HPV)-associated oropharyngeal cancer in Canada, 1992–2009. *Cancer Causes & Control*, *23*(8), 1343–1348. https://doi.org/10.1007/s10552-012-0013-z

Fowler, C. A., & Saltzman, E. (1993). Coordination and Coarticulation in Speech Production. *Language and Speech*, *36*(2–3), 171–195. https://doi.org/10.1177/002383099303600304

Frederick, J. W., Sweeny, L., Carroll, W. R., Peters, G. E., & Rosenthal, E. L. (2013). Outcomes in head and neck reconstruction by surgical site and donor site. *The Laryngoscope*, *123*(7), 1612–1617. https://doi.org/10.1002/lary.23775

Furia, C. L. B., Kowalski, L. P., Latorre, M. R. D. O., Angelis, E. C., Martins, N. M. S., Barros, A. P. B., & Ribeiro, K. C. B. (2001). Speech Intelligibility After Glossectomy and Speech Rehabilitation. *Archives of Otolaryngology–Head & Neck Surgery*, *127*(7), 877–883. https://doi.org/10-1001/pubs.Arch Otolaryngol. Head Neck Surg.-ISSN-0886-4470-127-7-ooa00191

Garden, A. S., Lewin, J. S., & Chambers, M. S. (2006). How to reduce radiation-related toxicity in patients with cancer of the head and neck. *Current Oncology Reports*, *8*(2), 140–145. https://doi.org/10.1007/s11912-006-0049-x

Garnier, M., Bailly, L., Dohen, M., Welby, P., & Loevenbruck, H. (2006). *An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance*. 4.

Georgian, D. A., Logemann, J. A., & Fisher, H. B. (1982). Compensatory Articulation Patterns of a Surgically Treated Oral Cancer Patient. *Journal of Speech and Hearing Disorders*, *47*(2), 154–159.

Gick, B., Francis, N., Chiu, C., Stavness, I., & Fels, S. (2012). Producing whole speech events: Differential facial stiffness across the labial stops. *The Journal of the Acoustical Society of America*, *131*(4), 3345–3345. https://doi.org/10.1121/1.4708527

Gómez, P., Mekyska, J., Gómez, A., Palacios, D., Rodellar, V., & Álvarez, A. (2019). Characterization of Parkinson's disease dysarthria in terms of speech articulation kinematics. *Biomedical Signal Processing and Control*, *52*, 312–320. https://doi.org/10.1016/j.bspc.2019.04.029

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, *65*(1). https://doi.org/10.1007/BF00243838

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech, Language, and Hearing Research : JSLHR*, *43*(1), 239–255.

Green, J. R., & Wang, Y.-T. (2003). Tongue-surface movement patterns during speech and swallowing. *The Journal of the Acoustical Society of America*, *113*(5), 2820–2833. https://doi.org/10.1121/1.1562646

Greven, A. J., Meijer, M. F., & Tiwari, R. M. (1994). Articulation after total glossectomy: A clinical study of speech in six patients. *International Journal of Language & Communication Disorders*, *29*(1), 85–93. https://doi.org/10.3109/13682829409041484

Grimm, D. L., Stone, M., Woo, J., Lee, J., Hwang, J.-H., Bedrosian, G. E., & Prince, J. L. (2017). The Effects of Palate Features and Glossectomy Surgery on /s/ Production. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3417–3425. https://doi.org/10.1044/2017_JSLHR-S-16-0425

Hagedorn, C., Lammert, A., Bassily, M., Zu, Y., Sinha, U., Goldstein, L., & Narayanan, S. S. (2014). *Characterizing Post-Glossectomy Speech Using Real-time MRI*. 4.

Hecht, S. S. (2003). Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nature Reviews Cancer*, *3*(10), 733–744. https://doi.org/10.1038/nrc1190

Hiiemae, K. M., & Palmer, J. B. (2003). Tongue Movements in Feeding and Speech. *Critical Reviews in Oral Biology & Medicine*, *14*(6), 413–429. https://doi.org/10.1177/154411130301400604

Hinni, M. L., Ferlito, A., Brandwein-Gensler, M. S., Takes, R. P., Silver, C. E., Westra, W. H., Seethala, R. R., Rodrigo, J. P., Corry, J., Bradford, C. R., Hunt, J. L., Strojan, P., Devaney, K. O., Gnepp, D. R., Hartl, D. M., Kowalski, L. P., Rinaldo, A., & Barnes,

L. (2013). Surgical margins in head and neck cancer: A contemporary review. *Head & Neck*, *35*(9), 1362–1370. https://doi.org/10.1002/hed.23110

Hoefflin, S. M. (1998). Anatomy of the Platysma and Lip Depressor Muscles. *Dermatologic Surgery*, *24*(11), 1225–1231. https://doi.org/10.1111/j.1524-4725.1998.tb04102.x

Holmes, W. (2002). *Speech Synthesis and Recognition*. CRC Press.

Huang, J.-J., Wu, C.-W., Lam, W. L., Nguyen, D. H., Kao, H.-K., Lin, C.-Y., & Cheng, M.-H. (2012). Anatomical basis and clinical application of the ulnar forearm free flap for head and neck reconstruction. *The Laryngoscope*, *122*. https://doi.org/10.1002/lary.23565

Hur, M.-S., Kim, H.-J., Choi, B.-Y., Hu, K.-S., Kim, H.-J., & Lee, K.-S. (2013). Morphology of the Mentalis Muscle and Its Relationship With the Orbicularis Oris and Incisivus Labii Inferioris Muscles. *Journal of Craniofacial Surgery*, *24*(2), 602–604. https://doi.org/10.1097/SCS.0b013e318267bcc5

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256. https://doi.org/10.1017/S0954394509990093

Janik, S., Pyka, J., Stanisz, I., Wachholbinger, T., Leonhard, M., Roesner, I., Denk-Linnert, D.-M., Miles, B. A., Schneider-Stickler, B., & Erovic, B. M. (2019). Use of the myocutaneous serratus anterior free flap for reconstruction after salvage glossectomy. *European Archives of Oto-Rhino-Laryngology*, *276*(2), 559–566. https://doi.org/10.1007/s00405-018-5245-4

Kashino, M. (2006). Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology*, *27*(6), 318–321. https://doi.org/10.1250/ast.27.318

Kier, W. M., & Smith, K. K. (1985). Tongues, tentacles and trunks: The biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, *83*(4), 307–324. https://doi.org/10.1111/j.1096-3642.1985.tb01178.x

Kini, E. (2015). Free Flap Procedures for Reconstruction After Head and Neck Cancer. *AORN Journal*, *102*(6), 644.e1-644.e6. https://doi.org/10.1016/j.aorn.2015.10.013

Korpijaakko-Huuhka, A.-M., Söderholm, A.-L., & Lehtihalmes, M. (1999). Long-lasting speech and oral-motor deficiencies following oral cancer surgery: A retrospective study. *Logopedics Phoniatrics Vocology*, *24*(3), 97–106. https://doi.org/10.1080/140154399435048

Kovacs, L., Zimmermann, A., Brockmann, G., Gühring, M., Baurecht, H., Papadopulos, N. A., Schwenzer-Zimmerer, K., Sader, R., Biemer, E., & Zeilhofer, H. F. (2006). Three-dimensional recording of the human face with a 3D laser scanner. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, *59*(11), 1193–1202. https://doi.org/10.1016/j.bjps.2005.10.025

Kreeft, A., Tan, I. B., van den Brekel, M. W. M., Hilgers, F. J., & Balm, A. J. M. (2009). The surgical dilemma of "functional inoperability" in oral and oropharyngeal cancer:

Current consensus on operability with regard to functional results. *Clinical Otolaryngology: Official Journal of ENT-UK ; Official Journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, *34*(2), 140–146. https://doi.org/10.1111/j.1749-4486.2009.01884.x

Kumar, M., Nanavati, R., Modi, T., & Dobariya, C. (2016). *Oral cancer: Etiology and risk factors: A review*. *12*(2), 7.

LaBlance, G. R., Kraus, K., & Steckol, K. F. (1991). Rehabilitation of Swallowing and Communication Following Glossectomy. *Rehabilitation Nursing*, *16*(5), 266–270. https://doi.org/10.1002/j.2048-7940.1991.tb01231.x

Lewin, J. S., & Hutcheson, K. A. (2014). Evaluation and Rehabilitation of Speech, Voice, and Swallowing Functions After Treatment of Head and Neck Cancer. In *Head and neck cancer: A multidisciplinary approach.* (4th ed., p. 11). Lippincott Williams & Wilkins.

Ley, B. (1999). The Physics Factbook. In *The Physics Hyper Textbook*. http://hypertextbook.com/facts/1999/BrianLey.shtm

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Linowes, J., & Babilinski, K. (2017). *Augmented Reality for Developers: Build practical augmented reality applications with Unity, ARCore, ARKit, and Vuforia*. Packt Publishing Ltd.

List, M. A., Siston, A., Haraf, D., Schumm, P., Kies, M., Stenson, K., & Vokes, E. E. (1999). Quality of Life and Performance in Advanced Head and Neck Cancer Patients on Concomitant Chemoradiotherapy: A Prospective Examination. *Journal of Clinical Oncology*, *17*(3), 1020–1020. https://doi.org/10.1200/JCO.1999.17.3.1020

Loewen, I. J., Boliek, C. A., Harris, J., Seikaly, H., & Rieger, J. M. (2010). Oral sensation and function: A comparison of patients with innervated radial forearm free flap reconstruction to healthy matched controls. *Head & Neck*, *32*(1), 85–95. https://doi.org/10.1002/hed.21155

Löfqvist, A., & Gracco, V. L. (1997). Lip and Jaw Kinematics in Bilabial Stop Consonant Production. *Journal of Speech, Language, and Hearing Research*, *40*(4), 877–893. https://doi.org/10.1044/jslhr.4004.877

Logemann, J. A. (1999). Rehabilitation of head and neck cancer patients. In C. F. von Gunten (Ed.), *Palliative Care and Rehabilitation of Cancer Patients* (pp. 91–105). Springer US. https://doi.org/10.1007/978-1-4615-5003-7_6

Logemann, J. A., Pauloski, B. R., Rademaker, A. W., & Colangelo, L. A. (1997). Speech and Swallowing Rehabilitation for Head and Neck Cancer Patients. *Oncology*, *11*(5), 7.

Lowe, A. A. (1980). The neural regulation of tongue movements. *Progress in Neurobiology*, *15*(4), 295–344. https://doi.org/10.1016/0301-0082(80)90008-8

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review*, *77*(3), 182–196.

Martin-Harris, B., Garand, K. L. (Focht), & McFarland, D. (2017). Optimizing Respiratory-Swallowing Coordination in Patients With Oropharyngeal Head and Neck Cancer. *Perspectives of the ASHA Special Interest Groups*, *2*(13), 103–110. https://doi.org/10.1044/persp2.SIG13.103

Mayer, C., Abel, J., Barbosa, A., Black, A., & Vatikiotis-Bateson, E. (2011). The labial viseme reconsidered: Evidence from production and perception. *The Journal of the Acoustical Society of America*, *129*(4), 2456–2456. https://doi.org/10.1121/1.3588075

McConnel, F. M. S., Pauloski, B. R., Logemann, J. A., Rademaker, A. W., Colangelo, L., Shedd, D., Carroll, W., Lewin, J., & Johnson, J. (1998). Functional results of primary closure vs flaps in oropharyngeal reconstruction: A prospective study of speech and swallowing. *Archives of Otolaryngology - Head and Neck Surgery*, *124*(6), 625–630. https://doi.org/10.1001/archotol.124.6.625

Mercante, G., Ruscito, P., Pellini, R., Cristalli, G., & Spriano, G. (2013). Transoral robotic surgery (TORS) for tongue base tumours. *Acta Otorhinolaryngologica Italica*, *33*(4), 230–235.

Michiwaki, Y., Schmelzeisen, R., Hacki, T., & Michi, K. (1993). Functional effects of a free jejunum flap used for reconstruction in the oropharyngeal region. *Journal of Cranio-Maxillofacial Surgery*, *21*(4), 153–156. https://doi.org/10.1016/S1010-5182(05)80104-0

Moorhouse, T., & Edwards, D. (2018). Head and neck cancer. *InnovAiT*, *11*(6), 342–346. https://doi.org/10.1177/1755738018758495

Morrish, L. (1984). Compensatory vowel articulation of the glossectomee: Acoustic and videofluoroscopic evidence. *International Journal of Language & Communication Disorders*, *19*(2), 125–134. https://doi.org/10.3109/13682828409007183

Nair, U., Bartsch, H., & Nair, J. (2004). Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: A review of agents and causative mechanisms. *Mutagenesis*, *19*(4), 251–262. https://doi.org/10.1093/mutage/geh036

Neligan, P. C. (2013). Head and Neck Reconstruction. *Plastic and Reconstructive Surgery*, *131*(2), 260e. https://doi.org/10.1097/PRS.0b013e3182778938

Nguyen-Tan, P. F., Zhang, Q., Ang, K. K., Weber, R. S., Rosenthal, D. I., Soulieres, D., Kim, H., Silverman, C., Raben, A., Galloway, T. J., Fortin, A., Gore, E., Westra, W. H., Chung, C. H., Jordan, R. C., Gillison, M. L., List, M., & Le, Q.-T. (2014). Randomized phase III trial to test accelerated versus standard fractionation in combination with concurrent cisplatin for head and neck carcinomas in the Radiation Therapy Oncology Group 0129 trial: Long-term report of efficacy and toxicity. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, *32*(34), 3858–3866. https://doi.org/10.1200/JCO.2014.55.3925

Nouraei, S. A. R., Middleton, S. E., Hudovsky, A., Branford, O. A., Lau, C., Clarke, P. M., Wood, S. H., Aylin, P., Mace, A., Jallali, N., & Darzi, A. (2015). Role of reconstructive surgery in the management of head and neck cancer: A national outcomes analysis of 11,841 reconstructions. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, *68*(4), 469–478. https://doi.org/10.1016/j.bjps.2014.11.011

Öhman, S. E. G. (1967). Numerical Model of Coarticulation. *The Journal of the Acoustical Society of America*, *41*(2), 310–320. https://doi.org/10.1121/1.1910340

O'Malley, B. W., Weinstein, G. S., Snyder, W., & Hockstein, N. G. (2006). Transoral Robotic Surgery (TORS) for Base of Tongue Neoplasms. *The Laryngoscope*, *116*(8), 1465–1472. https://doi.org/10.1097/01.mlg.0000227184.90514.1a

Ostry, D. J., & Keller, E. (1983). Similarities in the Control of the Speech Articulators and the Limbs: Kinematics of Tongue Dorsum Movement in Speech. *Journal of Experimental Psychology:Human Perception and Performance*, *9*(4), 622–636.

Palmer, P. M., Jaffe, D. M., McCulloch, T. M., Finnegan, E. M., Van Daele, D. J., & Luschei, E. S. (2008). Quantitative contributions of the muscles of the tongue, floor-of-mouth, jaw, and velum to tongue-to-palate pressure generation. *Journal of Speech, Language, and Hearing Research*, *51*(4), 828-. Gale Academic OneFile.

Parkin, D. M., Bray, F., Ferlay, J., & Pisani, P. (2005). Global Cancer Statistics, 2002. *CA: A Cancer Journal for Clinicians*, *55*(2), 74–108. https://doi.org/10.3322/canjclin.55.2.74

Parrell, B., & Narayanan, S. (2018). Explaining Coronal Reduction: Prosodic Structure and Articulatory Posture. *Phonetica*, *75*(2), 151–181. https://doi.org/10.1159/000481099

Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *The Journal of the Acoustical Society of America*, *114*(3), 1582–1599. https://doi.org/10.1121/1.1587737

Petruson, K., Mercke, C., Lundberg, L.-M., Silander, E., & Hammerlid, E. (2005). Longitudinal evaluation of patients with cancer in the oral tongue, tonsils, or base of tongue – Does interstitial radiation dose affect quality of life? *Brachytherapy*, *4*(4), 271–277. https://doi.org/10.1016/j.brachy.2005.06.001

Piccinin, M. A., & Zito, P. M. (2020). Anatomy, Head and Neck, Lips. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK507900/

Popat, H., Zhurov, A. I., Toma, A. M., Richmond, S., Marshall, D., & Rosin, P. L. (2012). Statistical modelling of lip movement in the clinical context. *Orthodontics & Craniofacial Research*, *15*(2), 92–102. https://doi.org/10.1111/j.1601-6343.2011.01539.x

Radford, K., Woods, H., Lowe, D., & Rogers, S. N. (2004). A UK multi-centre pilot study of speech and swallowing outcomes following head and neck cancer. *Clinical Otolaryngology & Allied Sciences*, *29*(4), 376–381. https://doi.org/10.1111/j.1365-2273.2004.00823.x

Rentschler, G. J., & Mann, M. B. (1980). The effects of glossectomy on intelligibility of speech and oral perceptual discrimination. *Journal of Oral Surgery (American Dental Association: 1965)*, *38*(5), 348–354.

Rhodus, N. L., Kerr, A. R., & Patel, K. (2014). Oral Cancer. *Dental Clinics of North America*, *58*(2), 315–340. https://doi.org/10.1016/j.cden.2013.12.004

Richardson, A., Addington-Hall, J., Amir, Z., Foster, C., Stark, D., Armes, J., Brearley, S. G., Hodges, L., Hook, J., Jarrett, N., Stamataki, Z., Scott, I., Walker, J., Ziegler, L., & Sharpe, M. (2011). Knowledge, ignorance and priorities for research in key areas of cancer survivorship: Findings from a scoping review. *British Journal of Cancer*, *105*(S1), S82–S94. https://doi.org/10.1038/bjc.2011.425

Rieger, J., Dickson, N., Lemire, R., Bloom, K., Wolfaardt, J., Wolfaardt, U., & Seikaly, H. (2006). Social Perception of Speech in Individuals with Oropharyngeal Reconstruction. *Journal of Psychosocial Oncology*, *24*(4), 33–51. https://doi.org/10.1300/J077v24n04_03

Rogers, S. N., Ahad, S. A., & Murphy, A. P. (2007). A structured review and theme analysis of papers published on ‘quality of life’ in head and neck cancer: 2000–2005. *Oral Oncology*, *43*(9), 843–868. https://doi.org/10.1016/j.oraloncology.2007.02.006

Rosenthal, D. I., Lewin, J. S., & Eisbruch, A. (2006). Prevention and Treatment of Dysphagia and Aspiration After Chemoradiation for Head and Neck Cancer. *Journal of Clinical Oncology*, *24*(17), 2636–2643. https://doi.org/10.1200/JCO.2006.06.0079

Sanders, I., & Mu, L. (2013). A three-dimensional atlas of human tongue muscles. *Anatomical Record (Hoboken, N.J.: 2007)*, *296*(7), 1102–1114. https://doi.org/10.1002/ar.22711

Santosh, A. B. R., Cumberbatch, K., & Jones, T. (2017). Post-glossectomy in lingual carcinomas: A scope for sign language in rehabilitation. *Contemporary Oncology*, *21*(2), 123–130. https://doi.org/10.5114/wo.2017.68620

Scherr, S. A., Kammler, C., & Elberzhager, F. (2019). Detecting User Emotions with the True-Depth Camera to Support Mobile App Quality Assurance. *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 169–173. https://doi.org/10.1109/SEAA.2019.00034

Schötz, S., Frid, J., & Löfqvist, A. (2013). Development of speech motor control: Lip movement variability. *The Journal of the Acoustical Society of America*, *133*(6), 4210–4217. https://doi.org/10.1121/1.4802649

Schwartz, D. L., & Hayes, D. N. (2020). The Evolving Role of Radiotherapy for Head and Neck Cancer. *Hematology/Oncology Clinics of North America*, *34*(1), 91–108. https://doi.org/10.1016/j.hoc.2019.08.019

Scully, C., & Bagan, J. (2009). Oral squamous cell carcinoma overview. *Oral Oncology*, *45*(4–5), 301–308. https://doi.org/10.1016/j.oraloncology.2009.01.004

Shiboski, C. H., Shiboski, S. C., & Silverman, S. (2000). Trends in oral cancer rates in the United States, 1973–1996. *Community Dentistry and Oral Epidemiology*, *28*(4), 249–256. https://doi.org/10.1034/j.1600-0528.2000.280402.x

Shin, Y. S., Koh, Y. W., Kim, S.-H., Jeong, J. H., Ahn, S., Hong, H. J., & Choi, E. C. (2012). Radiotherapy Deteriorates Postoperative Functional Outcome After Partial Glossectomy With Free Flap Reconstruction. *Journal of Oral and Maxillofacial Surgery*, *70*(1), 216–220. https://doi.org/10.1016/j.joms.2011.04.014

Šimko, J., Beňuš, Š., & Vainio, M. (2016). Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *The Journal of the Acoustical Society of America*, *139*(1), 151–162. https://doi.org/10.1121/1.4939495

Solomon, N. P., & Munson, B. (2004). The effect of jaw position on measures of tongue strength and endurance. *Journal of Speech, Language, and Hearing Research*, *47*(3), 584-. Gale Academic OneFile.

Stone, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *The Journal of the Acoustical Society of America*, *87*(5), 2207–2217. https://doi.org/10.1121/1.399188

Stone, M., Langguth, J. M., Woo, J., Chen, H., & Prince, J. L. (2014). Tongue Motion Patterns in Post-Glossectomy and Typical Speakers: A Principal Components Analysis. *Journal of Speech, Language, and Hearing Research*, *57*(3), 707–717. https://doi.org/10.1044/1092-4388(2013/13-0085)

Stone, M., Rizk, S., Woo, J., Murano, E. Z., Chen, H., & Prince, J. L. (2012). Frequency of Apical and Laminal /s/ in Normal and Postglossectomy Patients. *Journal of Medical Speech-Language Pathology*, *20*(4). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4492454/

Stone, M., Woo, J., Lee, J., Poole, T., Seagraves, A., Chung, M., Kim, E., Murano, E. Z., Prince, J. L., & Blemker, S. S. (2018). Structure and variability in human tongue muscle anatomy. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *6*(5), 499–507. https://doi.org/10.1080/21681163.2016.1162752

Story, B. H., & Bunton, K. (2010). Relation of vocal tract shape, formant transitions, and stop consonant identification. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1514-. Gale Academic OneFile.

Sunila, J., Hassuji, R., & Rajashekhar, B. (2011). Speech and swallowing outcomes in buccal mucosa carcinoma. *Indian Journal of Palliative Care*, *17*(3), 238. https://doi.org/10.4103/0973-1075.92344

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and Mandibular Dynamics during the Production of Bilabial Consonants: Preliminary Observations. *Journal of Speech and Hearing Research*, *16*(3), 397–420. https://doi.org/10.1044/jshr.1603.397

Swiftable. (2021). *Jabberwocky*. Swiftable LLC.

Trotti, A., Bellm, L. A., Epstein, J. B., Frame, D., Fuchs, H. J., Gwede, C. K., Komaroff, E., Nalysnyk, L., & Zilberberg, M. D. (2003). Mucositis incidence, severity and associated outcomes in patients with head and neck cancer receiving radiotherapy with or without chemotherapy: A systematic literature review. *Radiotherapy and Oncology*, *66*(3), 253–262. https://doi.org/10.1016/S0167-8140(02)00404-8

Tsai, M.-S., Lai, C.-H., Lee, C.-P., Yang, Y.-H., Chen, P.-C., Kang, C.-J., Chang, G.-H., Tsai, Y.-T., Lu, C.-H., Chien, C.-Y., Young, C.-K., Fang, K.-H., Liu, C.-J., Yeh, R.-M. A., & Chen, W.-C. (2016). Mortality in tongue cancer patients treated by curative surgery: A retrospective cohort study from CGRD. *PeerJ*, *4*, e2794. https://doi.org/10.7717/peerj.2794

Vainio, M., Järvikivi, J., Aalto, D., & Suni, A. (2010). Phonetic tone signals phonological quantity and word structure. *The Journal of the Acoustical Society of America*, *128*(1313), 10.

Walsh, B., Mettel, K. M., & Smith, A. (2015). Speech motor planning and execution deficits in early childhood stuttering. *Journal of Neurodevelopmental Disorders*, *7*(1), 27. https://doi.org/10.1186/s11689-015-9123-8

Westbury, J. R., Lindstrom, M. J., & McClean, M. D. (2002). Tongues and Lips Without Jaws: A Comparison of Methods for Decoupling Speech Movements. *Journal of Speech, Language, and Hearing Research*, *45*(4), 651–662. https://doi.org/10.1044/1092-4388(2002/052)

Whiting, D. M. (1965). Glossectomy. *South African Journal of Communication Disorders*, *11*(1), 25–26.

Wong, R. K., Poon, E. S.-M., Woo, C. Y.-M., Chan, S. C.-S., Wong, E. S.-P., & Chu, A. W.-S. (2007). Speech outcomes in Cantonese patients after glossectomy. *Head & Neck*, *29*(8), 758–764. https://doi.org/10.1002/hed.20566

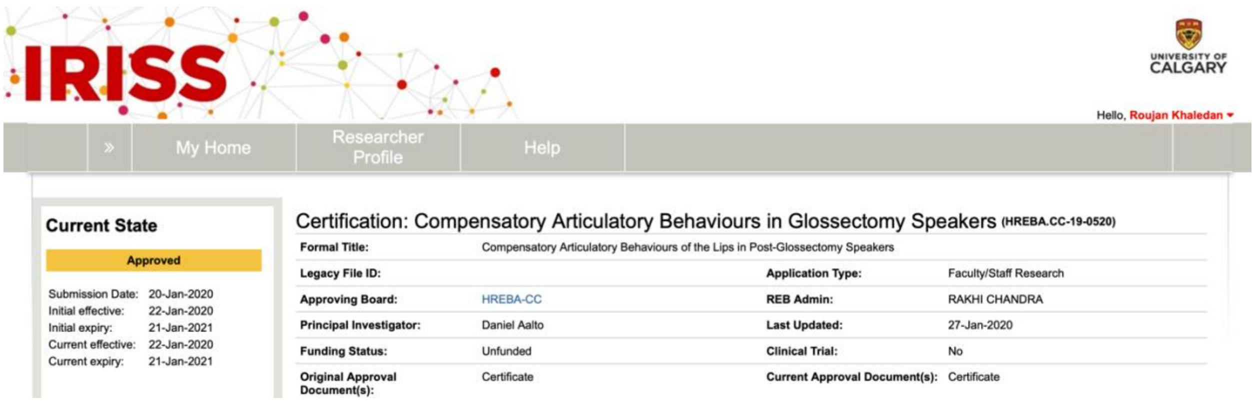
Yadav, P. (2013). Head and neck reconstruction. *Indian Journal of Plastic Surgery : Official Publication of the Association of Plastic Surgeons of India*, *46*(2), 275–282. https://doi.org/10.4103/0970-0358.118604

Zhuang, Z., Landsittel, D., Benson, S., Roberge, R., & Shaffer, R. (2010). Facial Anthropometric Differences among Gender, Ethnicity, and Age Groups. *The Annals of Occupational Hygiene*. https://doi.org/10.1093/annhyg/meq007

Appendix I. Ethics Approval for this Project

Figure 21

Confirmation of Ethics Approval


IRISS

UNIVERSITY OF CALGARY

Hello, Roujan Khaledan

» My Home | Researcher Profile | Help

Current State

Approved

Submission Date: 20-Jan-2020
Initial effective: 22-Jan-2020
Initial expiry: 21-Jan-2021
Current effective: 22-Jan-2020
Current expiry: 21-Jan-2021

Certification: Compensatory Articulatory Behaviours in Glossectomy Speakers (HREBA.CC-19-0520)

Formal Title:	Compensatory Articulatory Behaviours of the Lips in Post-Glossectomy Speakers		
Legacy File ID:		**Application Type:**	Faculty/Staff Research
Approving Board:	HREBA-CC	**REB Admin:**	RAKHI CHANDRA
Principal Investigator:	Daniel Aalto	**Last Updated:**	27-Jan-2020
Funding Status:	Unfunded	**Clinical Trial:**	No
Original Approval Document(s):	Certificate	**Current Approval Document(s):**	Certificate

Note. Confirmation of Ethics Approval for this Thesis

Appendix II. Instructions for Participants

Figure 22
Instructions for Participants

Thank you for participating in this experiment! I will be available by phone or video chat for the duration of this experiment and can be reached at: ______.

Using the Phone:

Step 1. Locate and open the ARFace Tracker app.

Step 2. Sit 55cm away from the phone. The distance is written on the bottom of the screen.

Step 3. Once you are comfortably seated, put on your headphones and turn up your headphones to a comfortable listening level, the sound should not be so loud that it causes any discomfort.

Step 4. Press the GREEN “start” button

Step 5. The iPhone will prompt you to record video and audio, press “accept”

Step 6. Read the first stimulus item

Step 7. Once you are done the recording, press the RED “stop” button

Note. The participants will receive the instructions above. The researcher will explain the steps verbally prior to recording as well. Following question 7, the researcher will force quit the application. Participants will be asked to press the buttons themselves because it allows them to set the pace for data collection.

Appendix III. Example of a Pseudo Randomized Phrase List

Table 6
Example of a Phrase List

Sound Type	Position	Sentence
Experimental voiceless /t/	Initial	The **tracks** ran all the way to the end of the block.
	Final	Don't worry about your shoes, there is a **rat** in the house, so you'd better leave them on.
	Initial	Just put the **table** in the corner.
	Final	She was mortified to be standing there with a food-**mart** badge on her shirt when her ex-boyfriend walked in.
	Initial	Life was better when she was a **teen** athlete.
	Initial	The **train** could only fit one driver.
	Final	The coast guard determined that the boat's **weight** was too high.
	Final	When it was her turn to **bait** the fish, she really impressed the crowd.
	Initial	Sally is a novice contractor, "I want to use that **tool** thing over there when I set up my deck", she said.
	Initial	Patrick had to swerve to avoid the **trash** on the road.
Experimental voiced /d/	Final	In the spring there was only a single **bud** to be found on the trees.
	Initial	I watched the mystery **dough** slowly rise.
	Final	The passengers gasped because the **bird** was fast approaching.
	Final	It was regarded as a **road** model; it was meant only for the city.
	Initial	There was a **dye** in the store that she knew would make her hair red.
	Final	It was a **bid** one would be proud of.
	Initial	The way it works is that you get your nails done while they **dote** on you.
	Initial	She had a somewhat **dummy** smile.
	Initial	There was **dew** all over the metal hand rails.
	Final	I want to **rid** myself of all those negative vibes.
Control voiceless /k/	Final	The dry **rock** in the ground made a nice bench.
	Initial	The **cop** was out of place in the funeral hall.
	Initial	The shopper put a **coat** back on the rack
	Initial	The water splashing on the **cap** will leave a watermark
	Final	When I go **back** there next time, I'll be ready!
	Initial	Sherry is a new zookeeper: Sherry looked and looked all day she finally found the **cub** near the corner store.
	Initial	The bride's father paid the **cab** for the party.
	Initial	I will only calm down once I find the **key** I am looking for.
	Final	Her **neck** was crooked
	Initial	What she saw was a **crew** team after all.
Control voiced /g/	Initial	The **gain** was too big to be overlooked.
	Initial	I don't know how I will ever get through this **gate** without incentive.
	Final	When you **dig** that up, disaster will ensue.
	Initial	It was not hard to see that the **gear** was out of place in the middle of the watch.

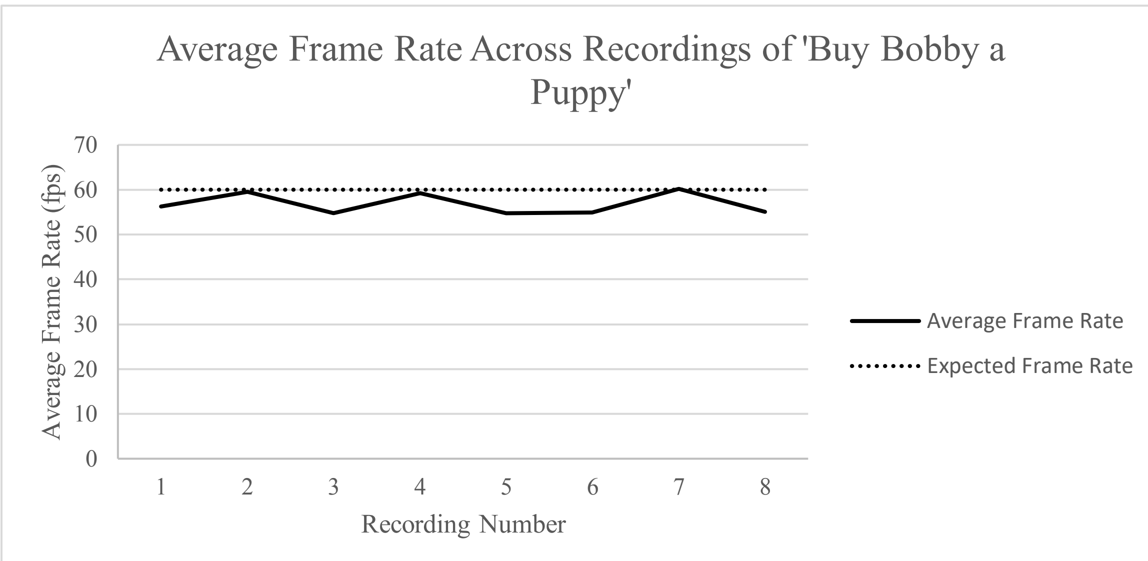
	Initial	The **grain** had filled the cup.
	Final	There was a **lag** in the train station computers
	Final	Dean's gear was now being transported by **sag** wagon like everyone else's.
	Initial	She admired how fluffy her **gown** looked as she passed the shop window.
	Final	When the machine is cold it doesn't work was well, for example if the **cog** gets cold, it may jam.
	Initial	The **gust** on the shore was particularly evident today wasn't it?

Note. This table provides an example of an experimental list. Each list was pseudo-randomly generated from a set of experimental words (voiceless: /t/ and voiced: /d/) and control words (voiceless: /k/ and voiced: /g/). In for each of the word types, 10 sentences were chosen. Of the 40 words chosen, 0.13% had consonant clusters word initially.

Appendix IV. Frame Rates

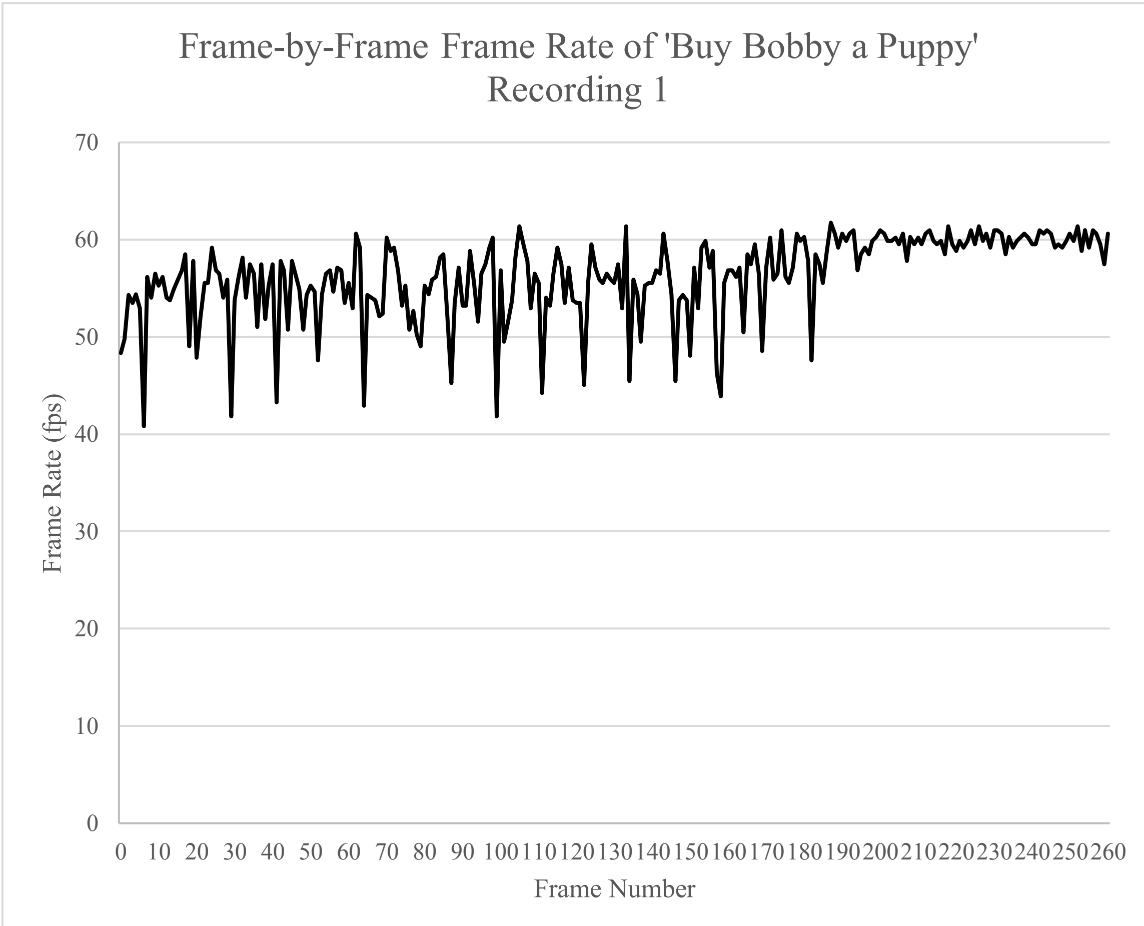
In order to accurately analyze kinematic data from the prototype, it was important to first ascertain whether it was truly capturing data at a rate of 60 fps. To determine whether this was the case, the UNIX timestamps of each frame were used to determine stability of frame rates across 8 recordings of the participant saying, ‘buy bobby a puppy’ (Figure 23). While an average frame rate of 60 fps was expected, an actual average of 56.8 fps (min: 54.7 fps, max: 60.2 fps) was observed across the 8 recordings.

Figure 23
Average Frame Rate Across Recordings of ‘buy Bobby a puppy’



While the average frame rates for these recordings were within 6 fps of the expected frame rate, Figure 24 shows that the frame-to-frame rate varied by up to 20.9 frames (min: 40.8 fps, max: 61.7fps).

Figure 24
Frame-by-Frame Frame Rate of ‘buy Bobby a puppy’



In this example, the frame rates became more stable after frame 190, however this pattern did not hold across other recordings, therefore it was not possible to predict when frame rates would become more stable.

Appendix V. Basics of the Prototype

The following figures (Figure 25-27) illustrate the numbered face mesh, the orientation of the axes, and the misalignment between the face-tracking data and the A/V data.

Figure 25
Numbered Mesh

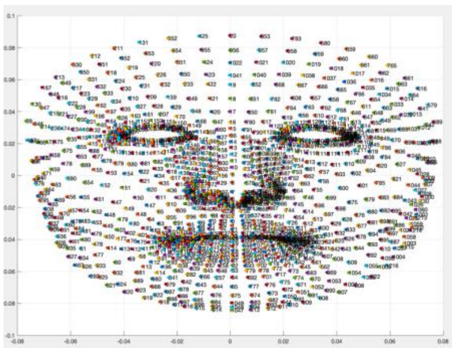
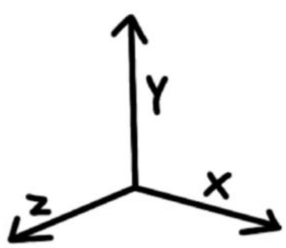
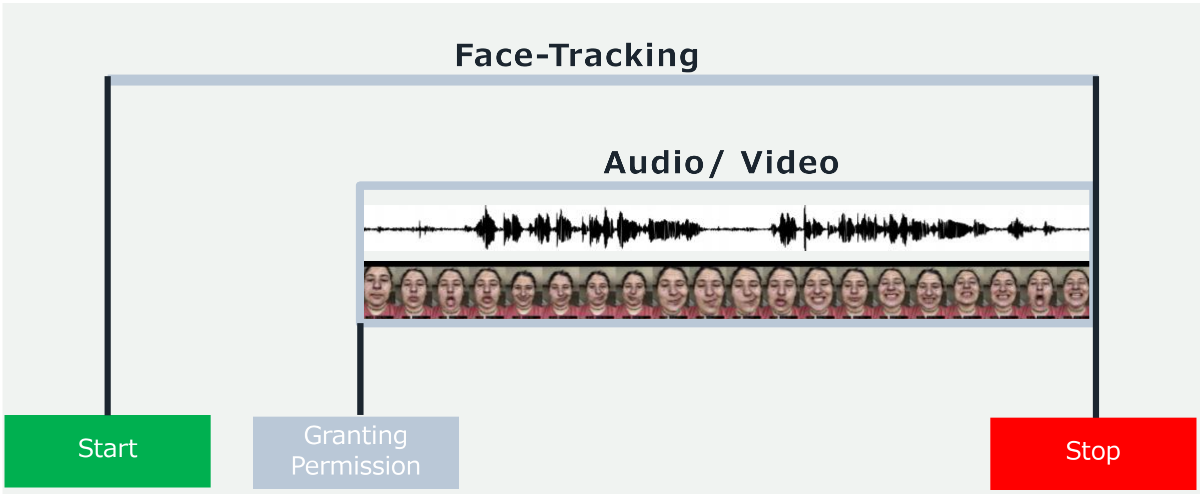


Figure 26
Orientation of Axes



Note. Orientation of axes. The x axis indicates left-right movement, the y axis indicates up-down movement, and the z axis indicates frontward-backward movement.

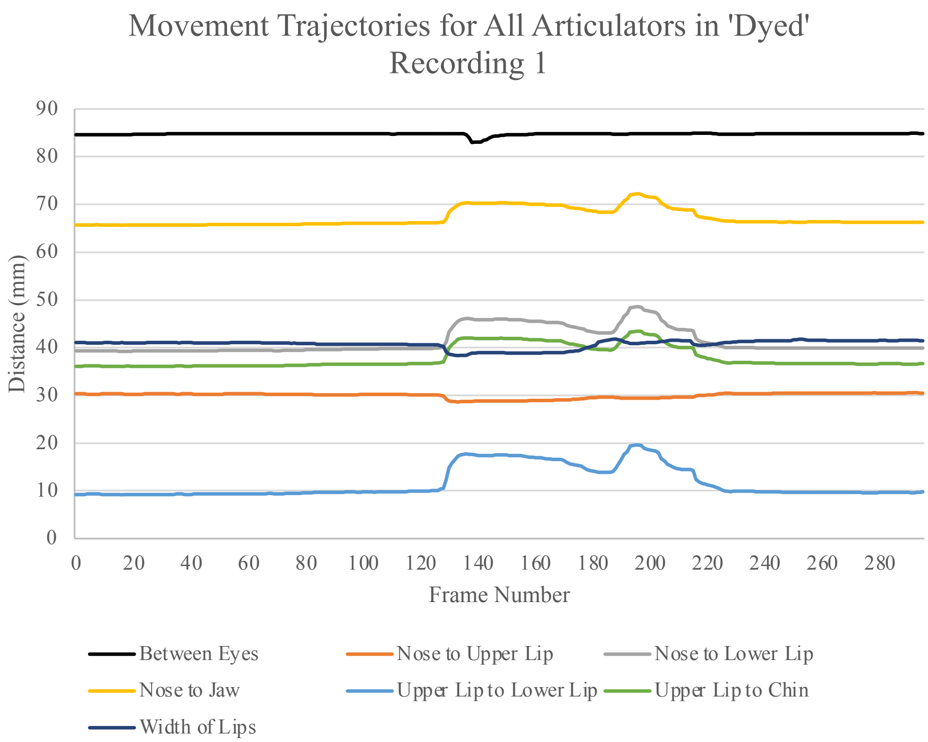
Figure 27
Misalignment Between Face-Tracking and A/V Data

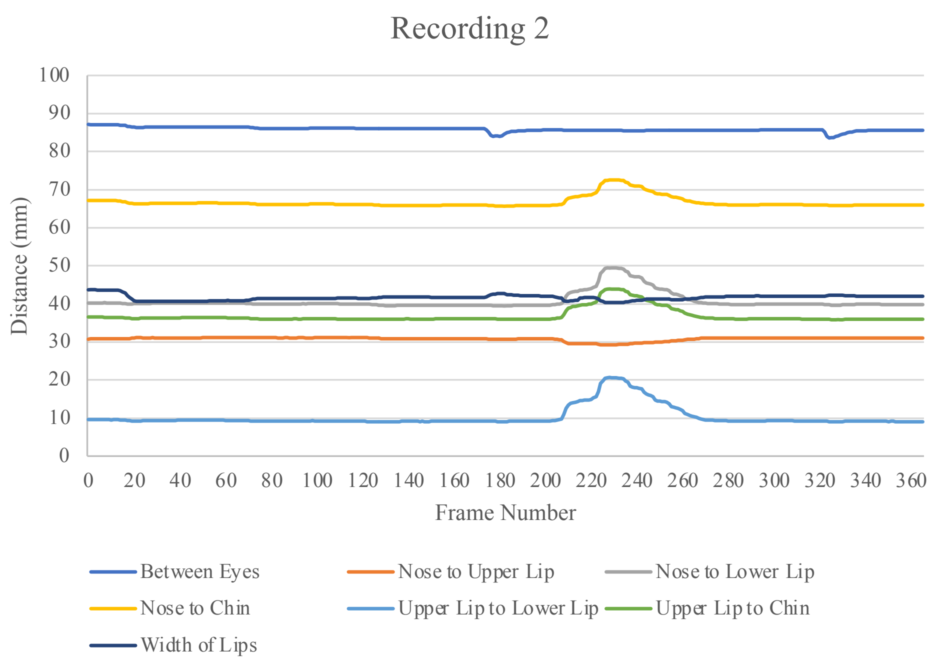


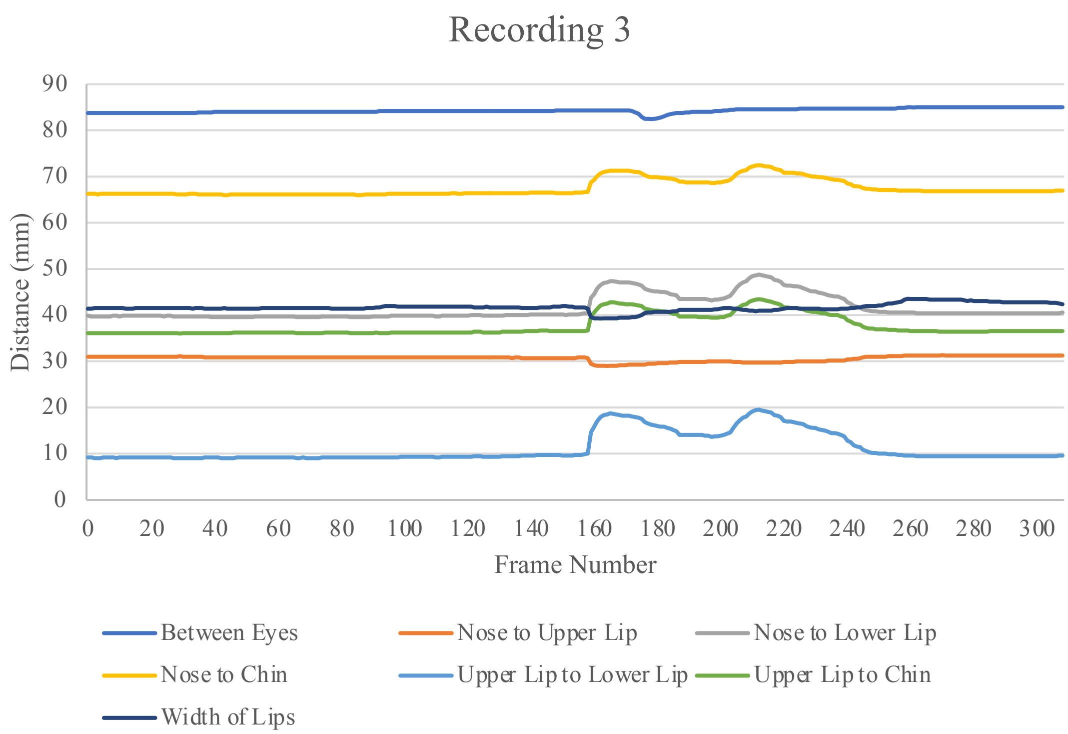
Appendix VI. Movement Trajectories for All Stimulus Items in Chapter 2, Question 1

Figure 28

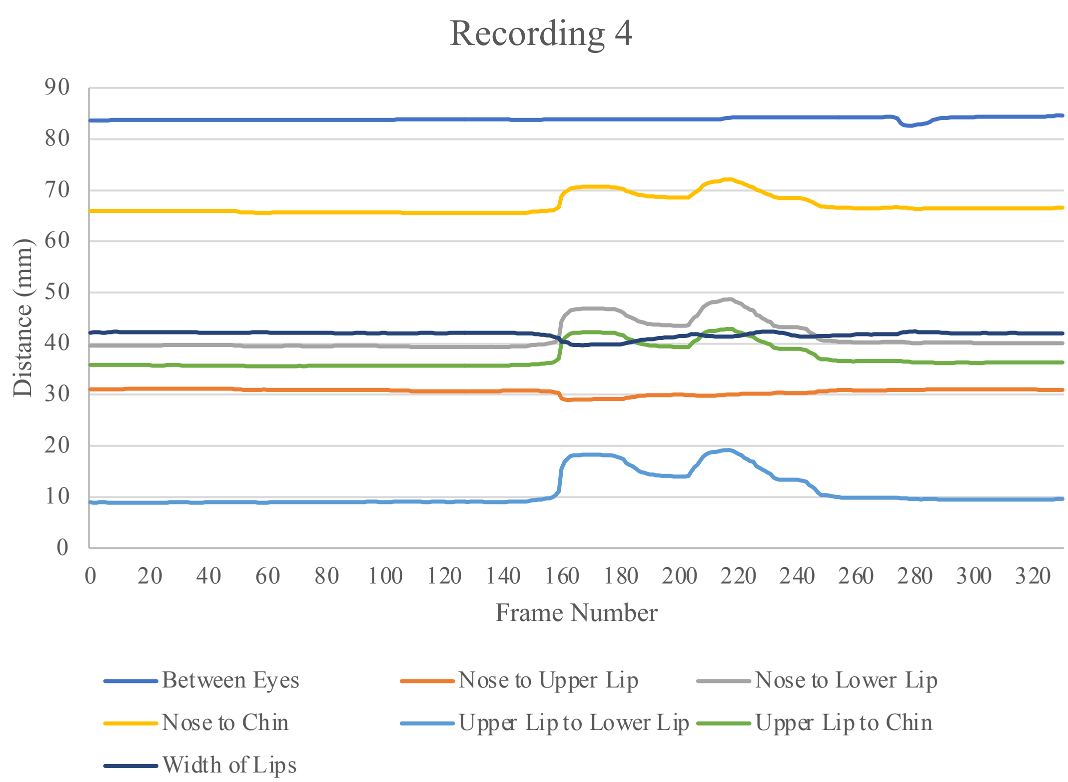
Movement Trajectories for All Stimulus Items



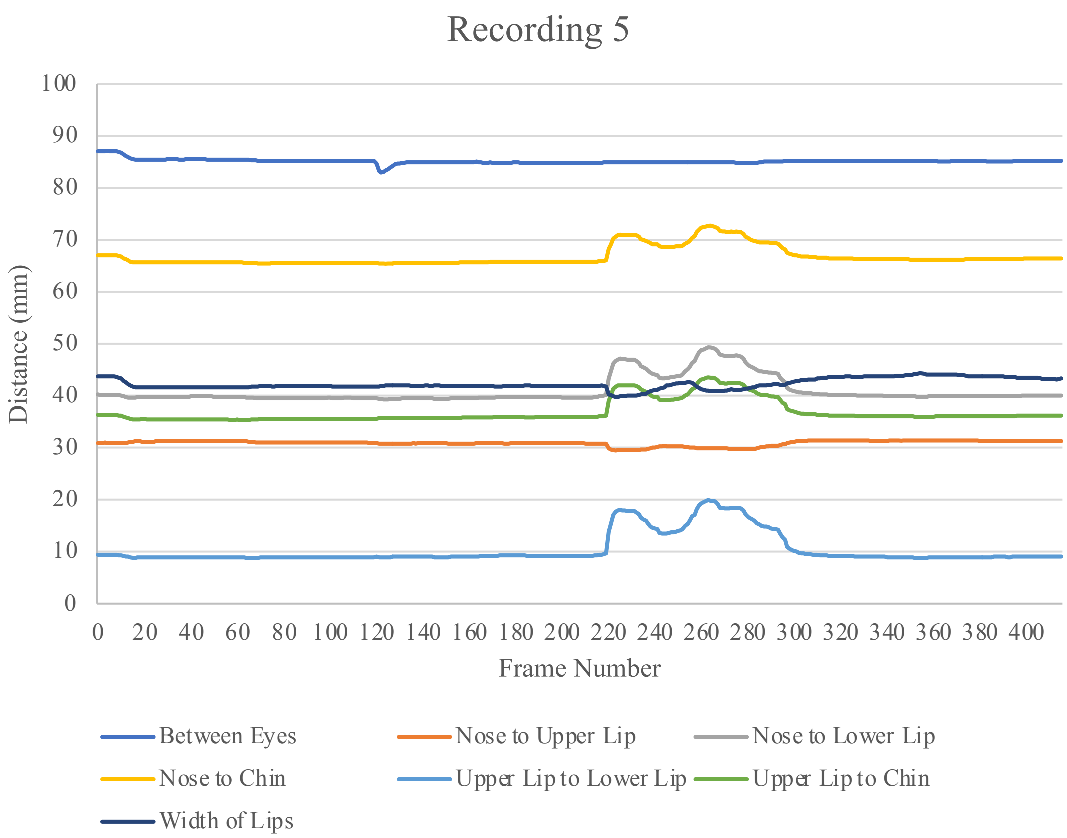




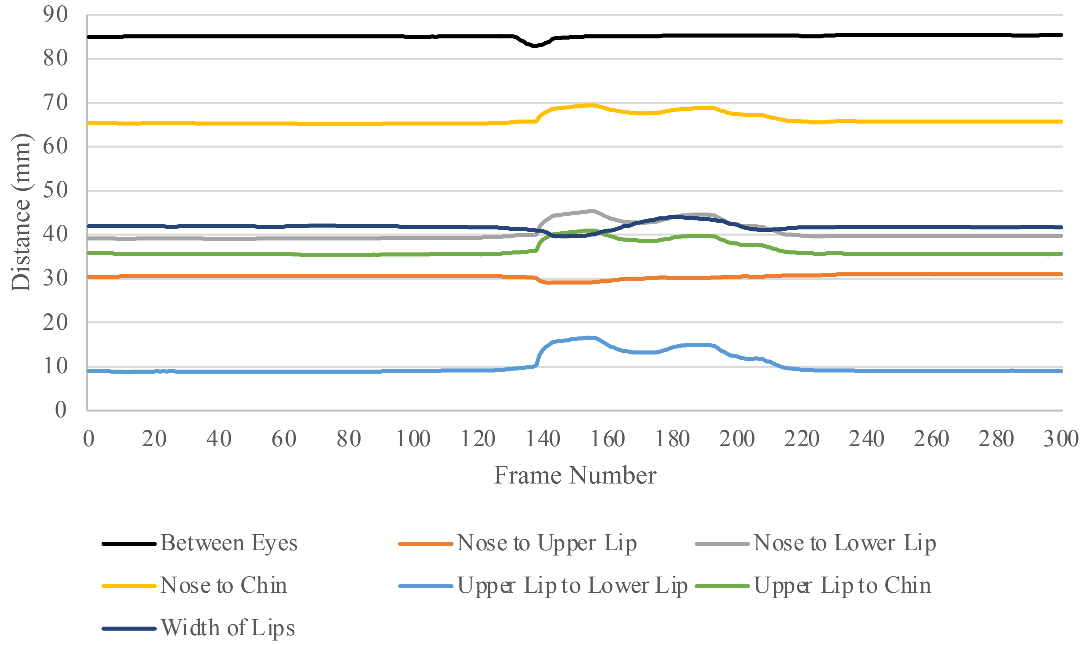
Recording 3
Distance (mm)
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips



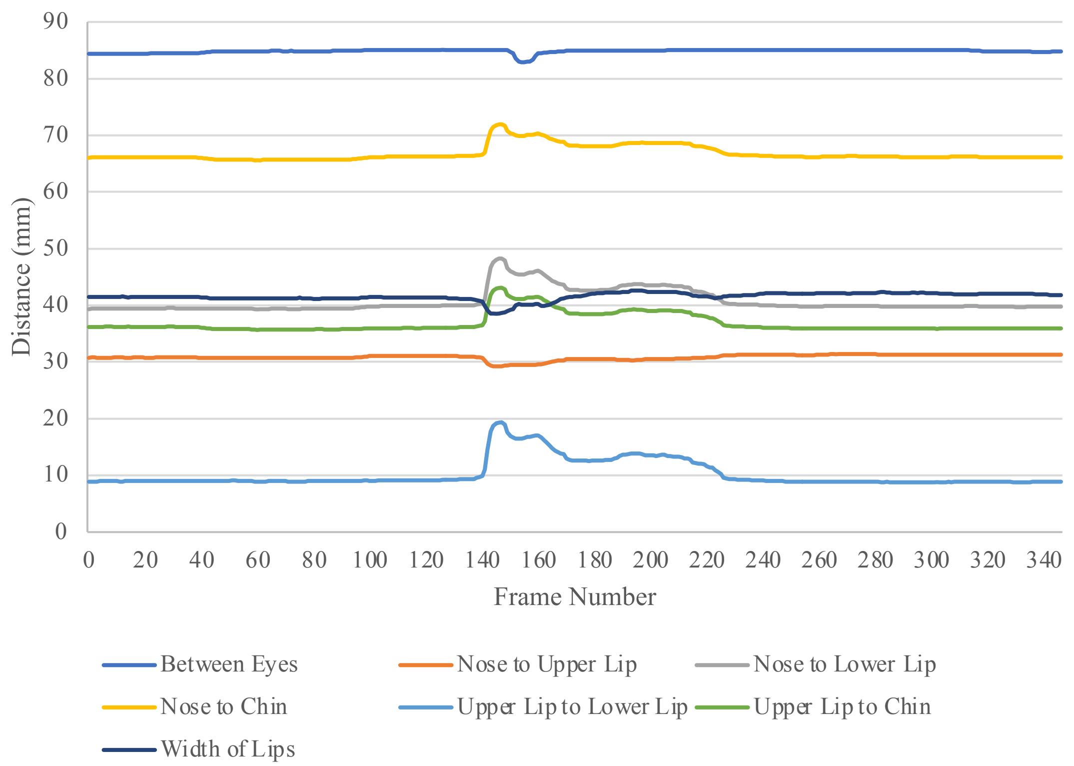
Recording 4
Distance (mm)
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

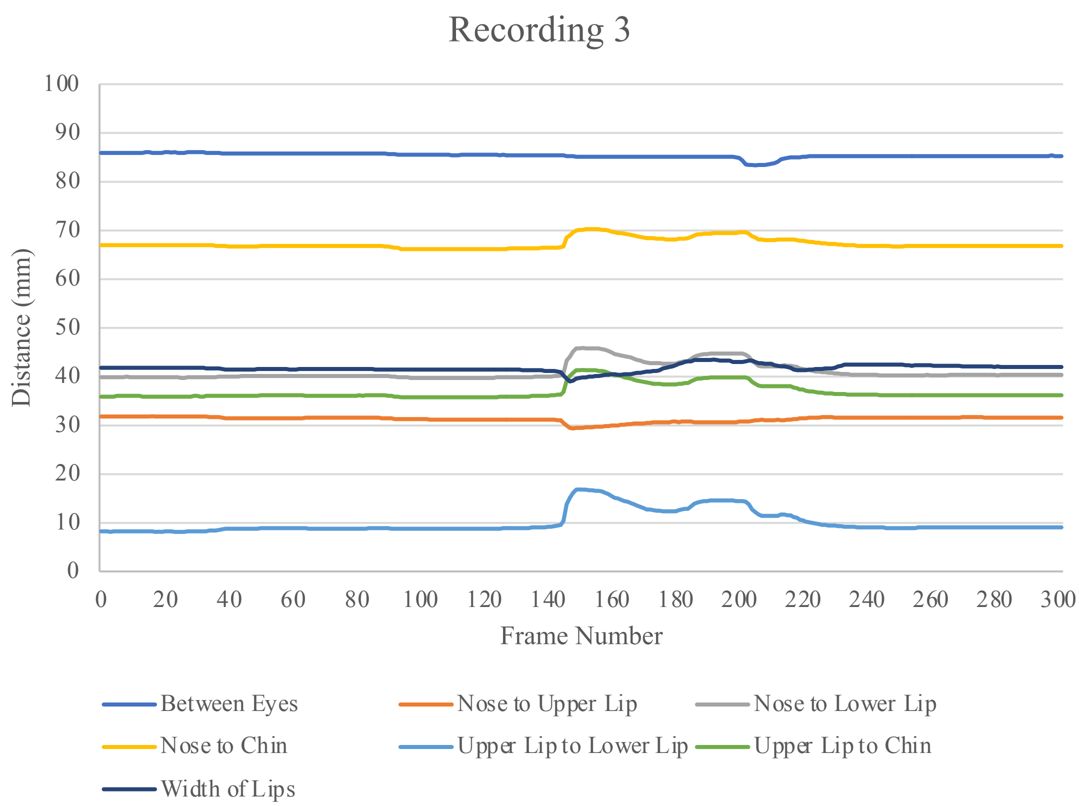

Recording 5
Distance (mm)
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
360
380
400
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

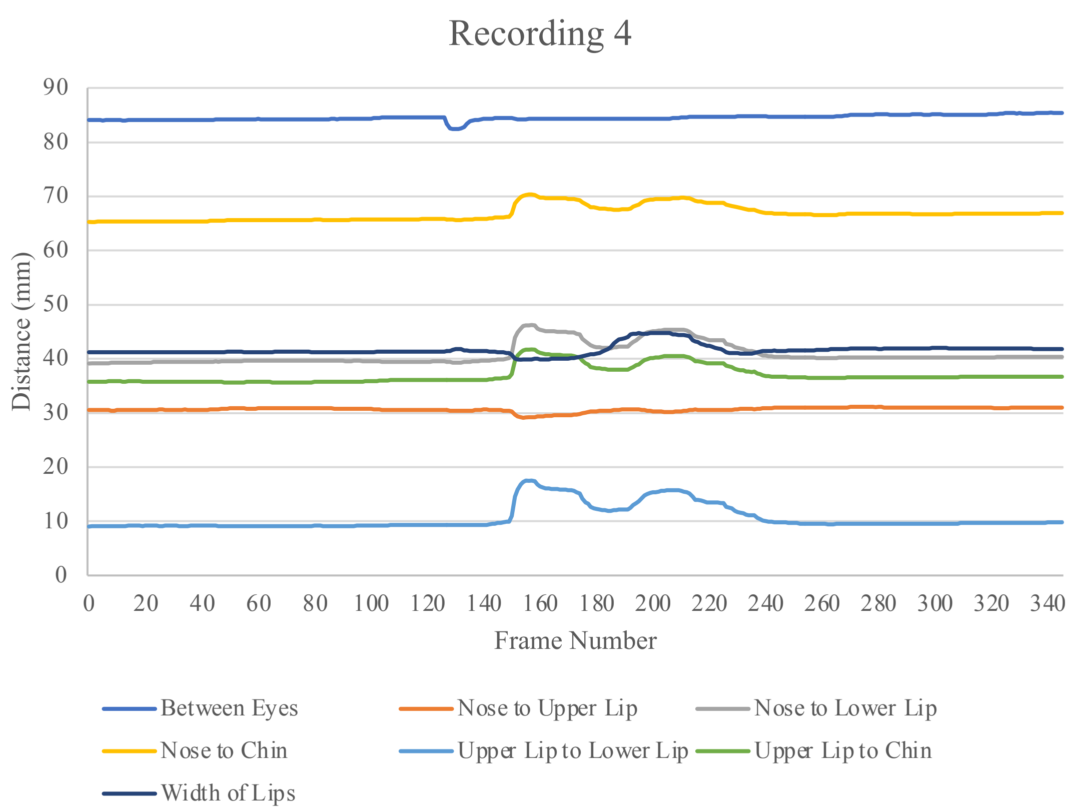
Movement Trajectories for all Tracked Points in 'deed'
Recording 1

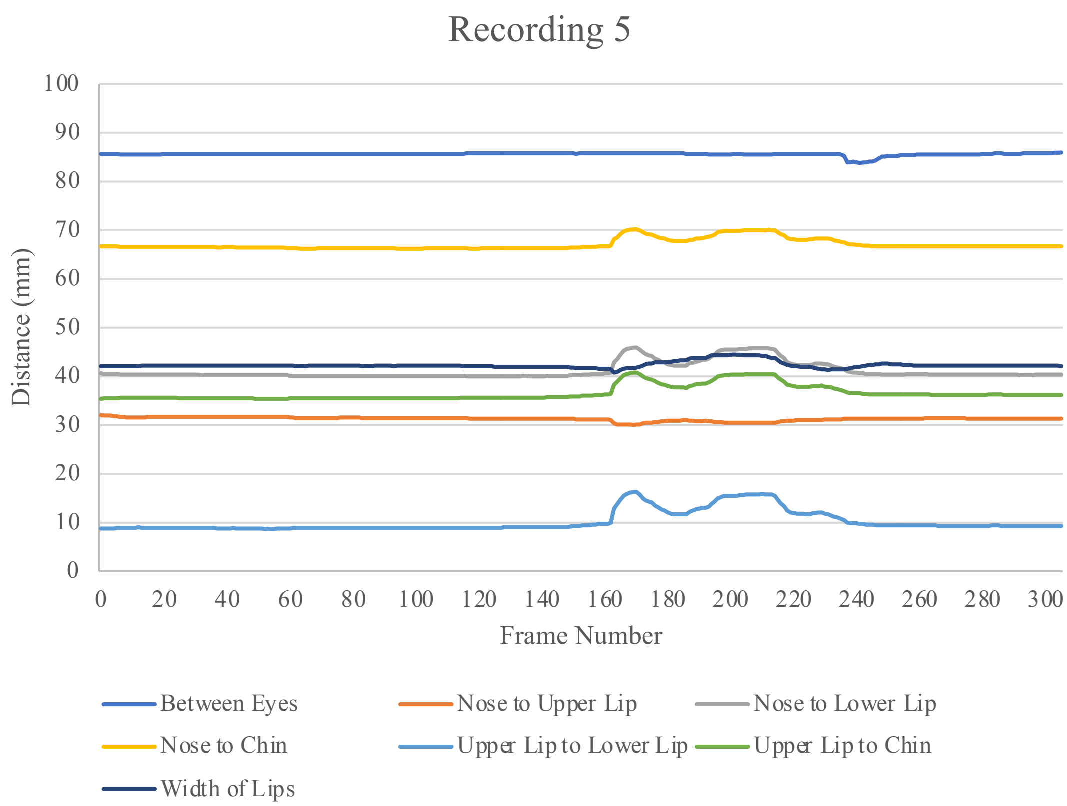


Recording 2



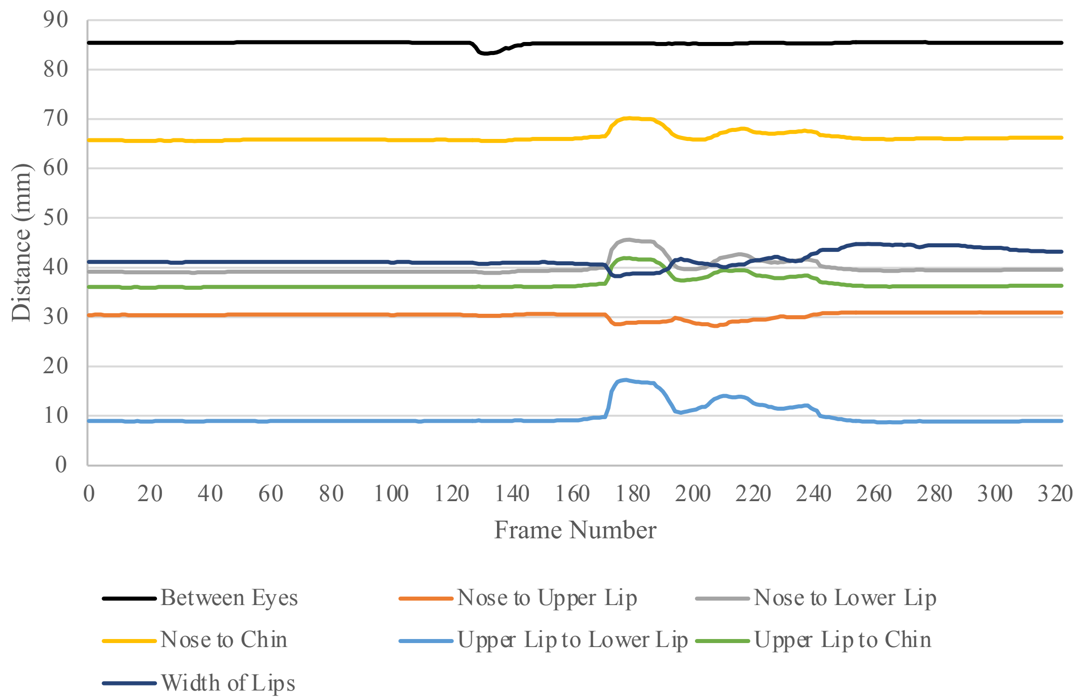

Recording 3
100
90
80
70
60
50
40
30
20
10
0
Distance (mm)
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips


Recording 4
90
80
70
60
50
40
30
20
10
0
Distance (mm)
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

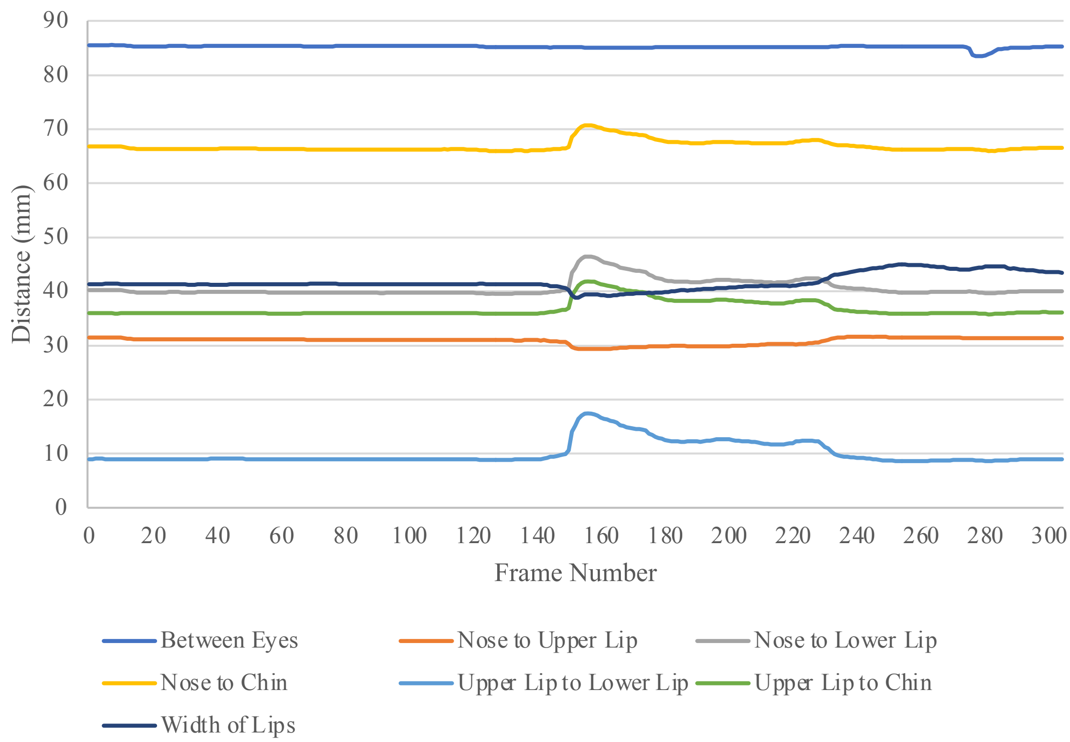

Recording 5
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

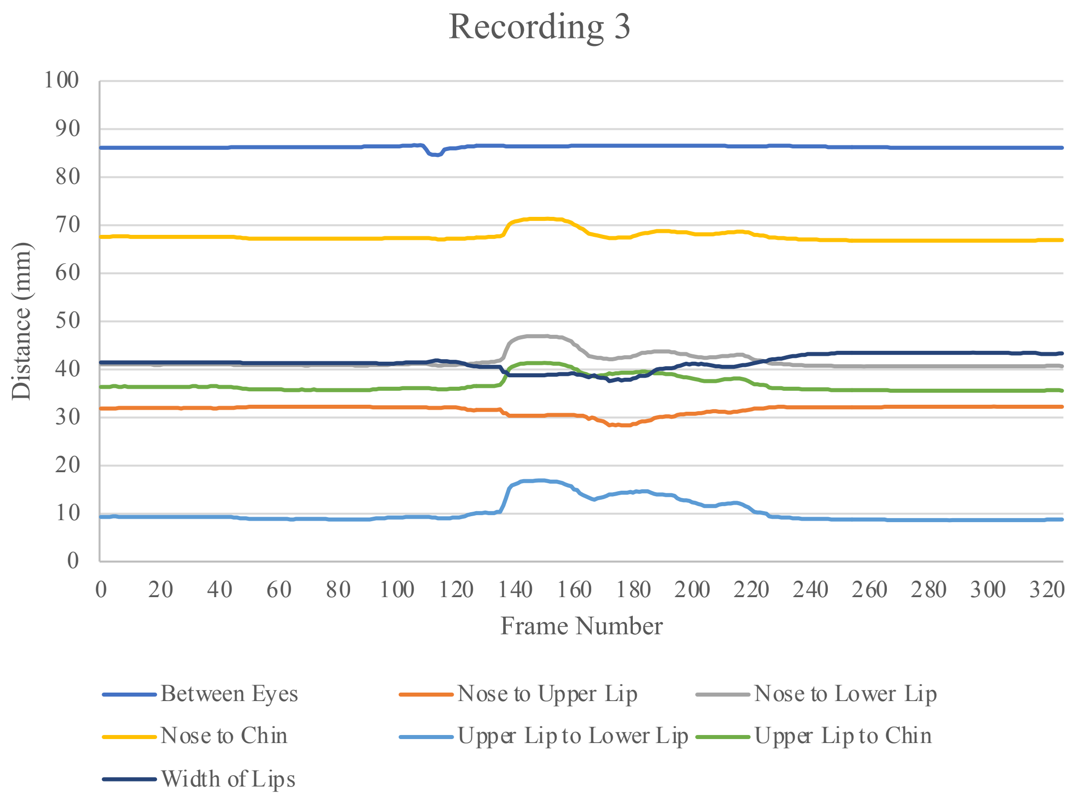
Movement Trajectories for all Tracked Points of 'dude'

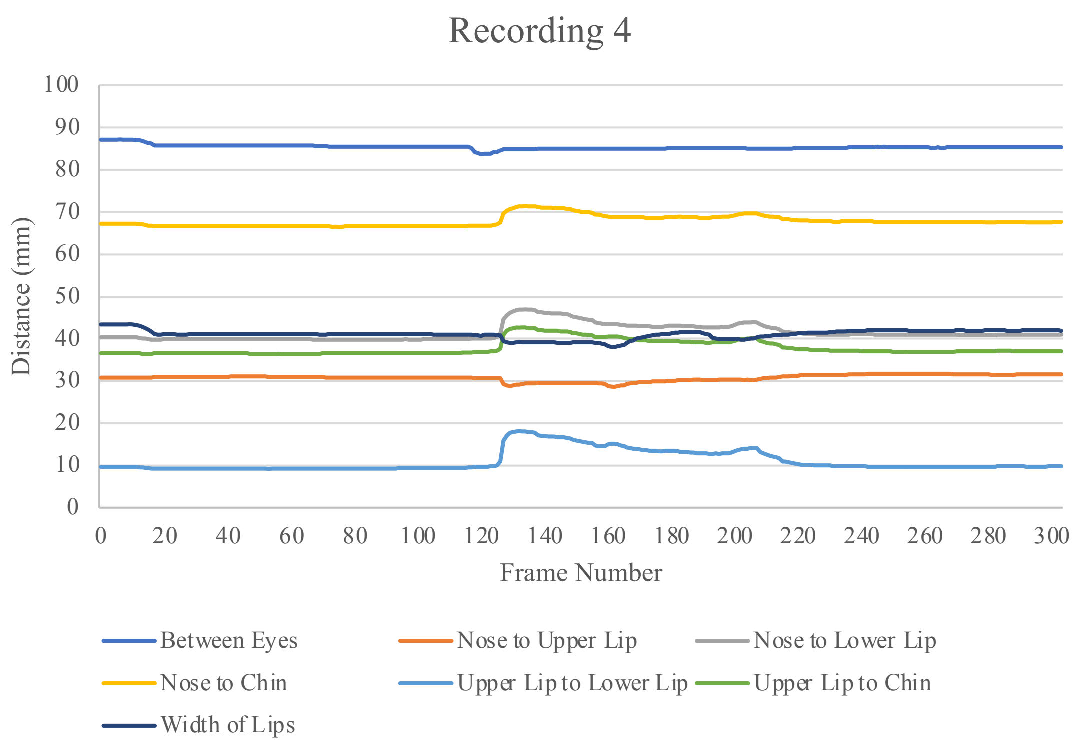
Recording 1



Recording 2

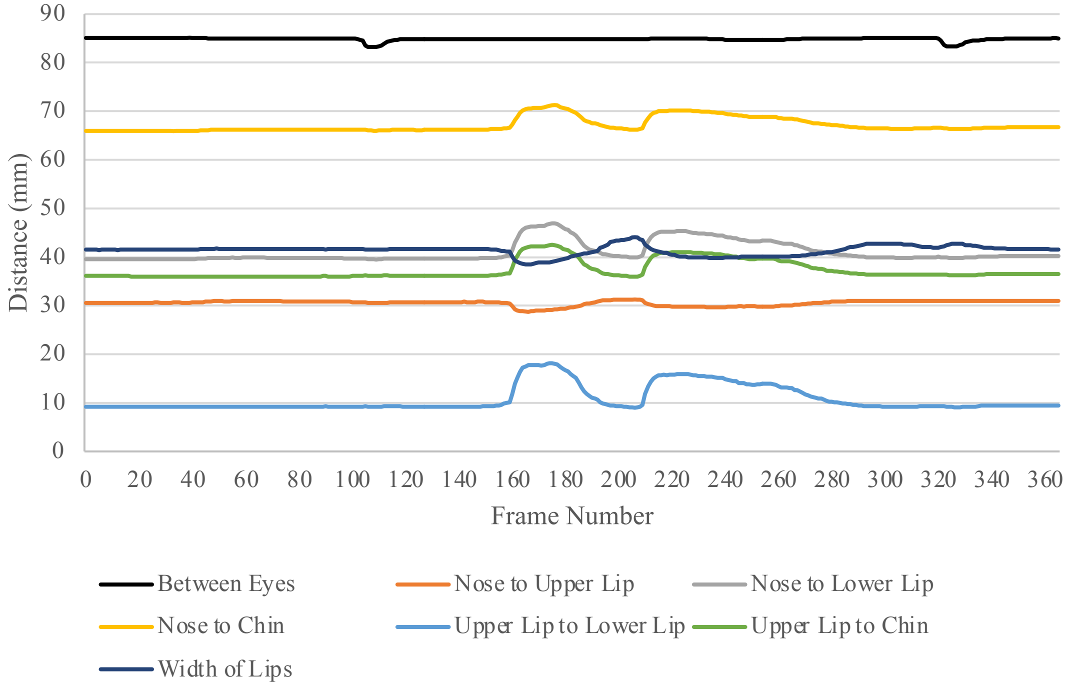




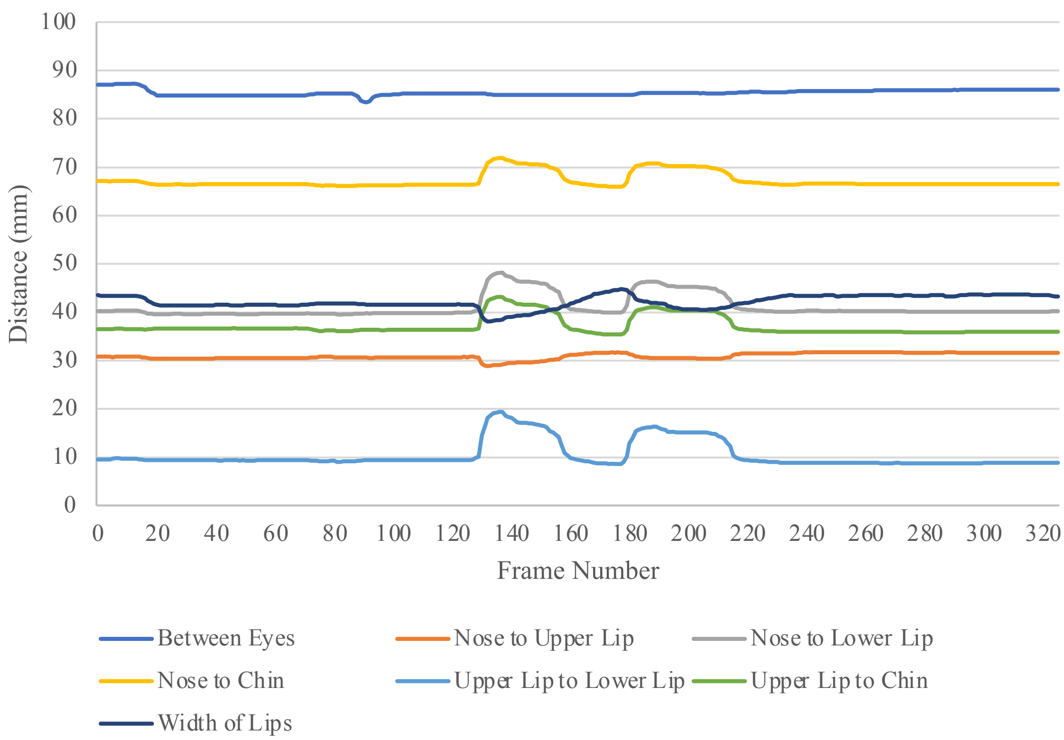


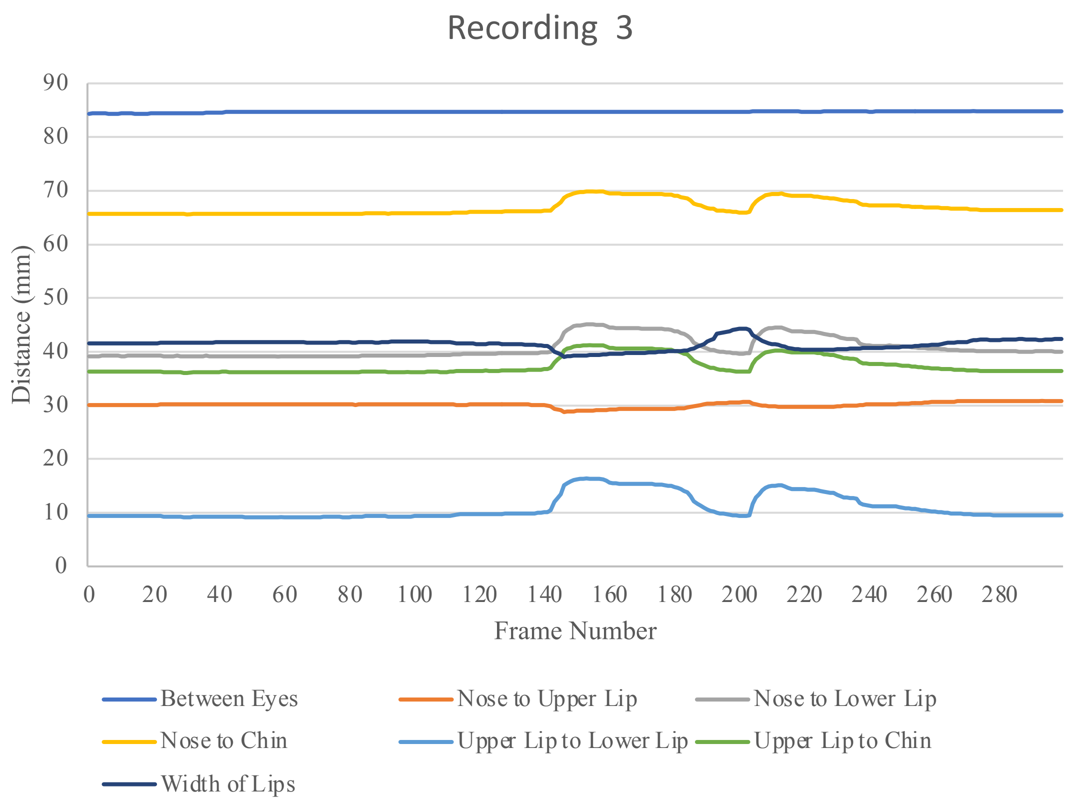
Due to unexpected data loss, recording 5 of ‘dude’ was not saved.

Movement Trajetories for all Tracked Points in 'fin'
Recording 1

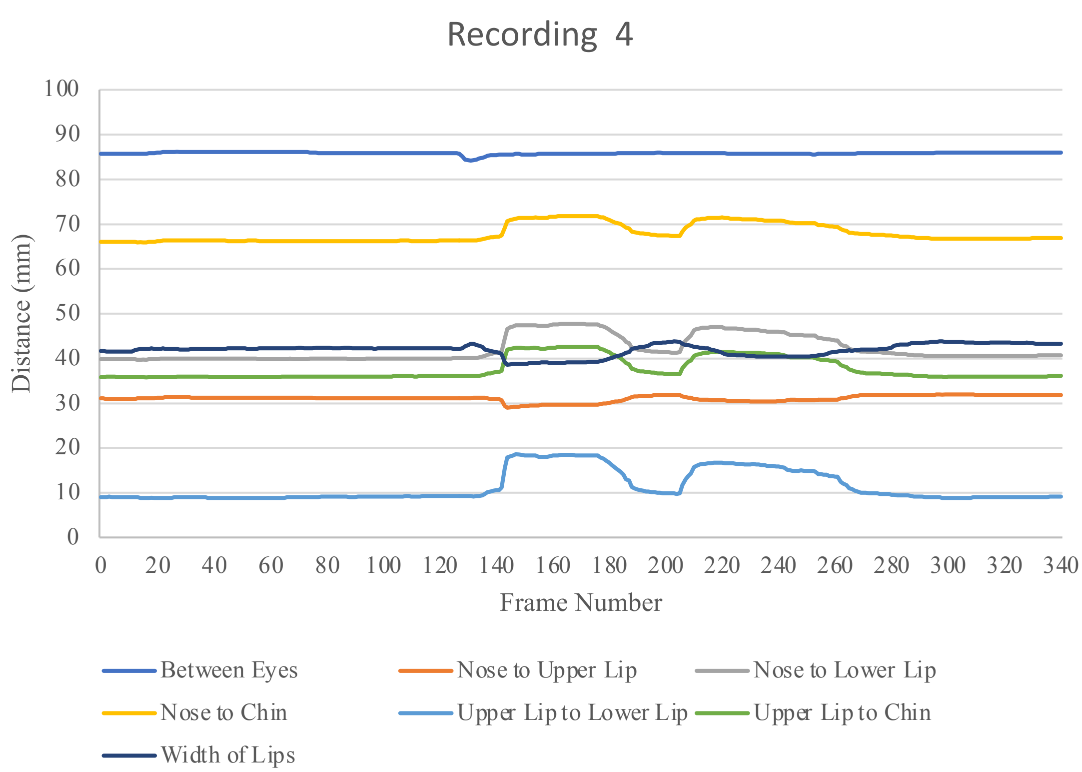


Recording 2

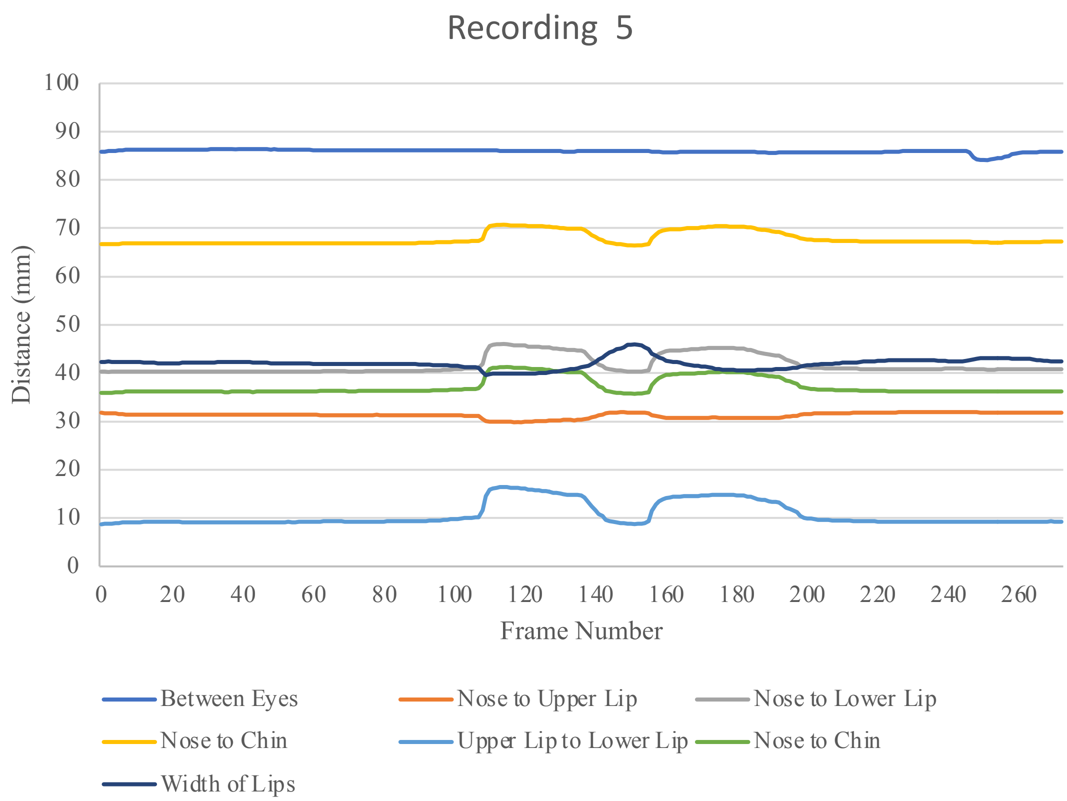




Recording 3
Distance (mm)
0
10
20
30
40
50
60
70
80
90
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

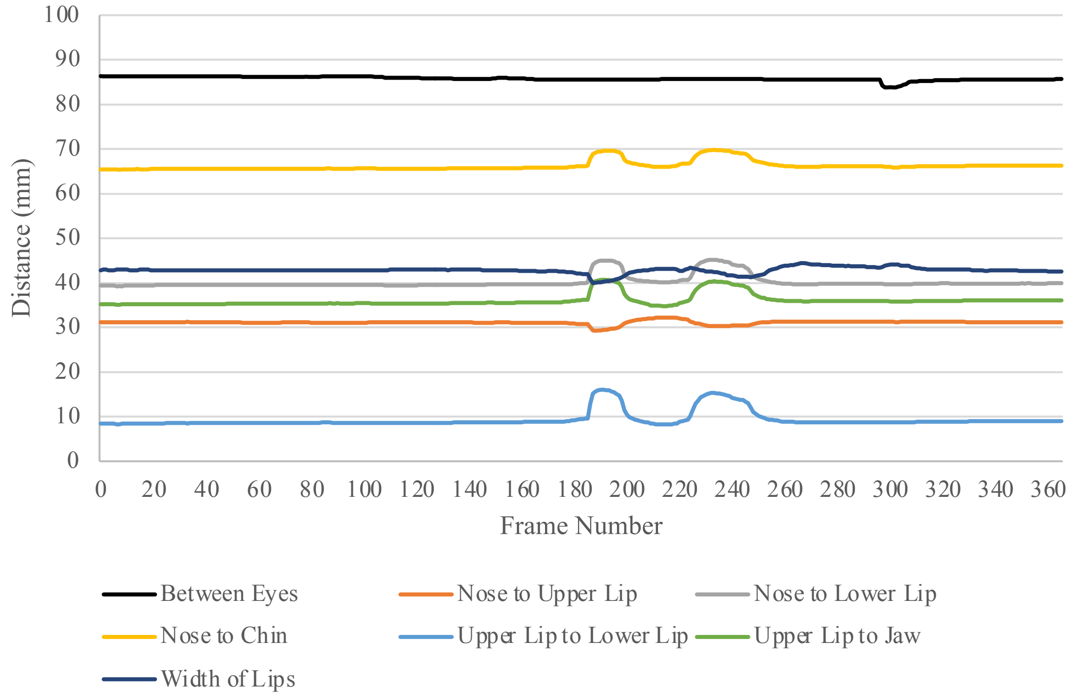


Recording 4
Distance (mm)
0
10
20
30
40
50
60
70
80
90
100
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

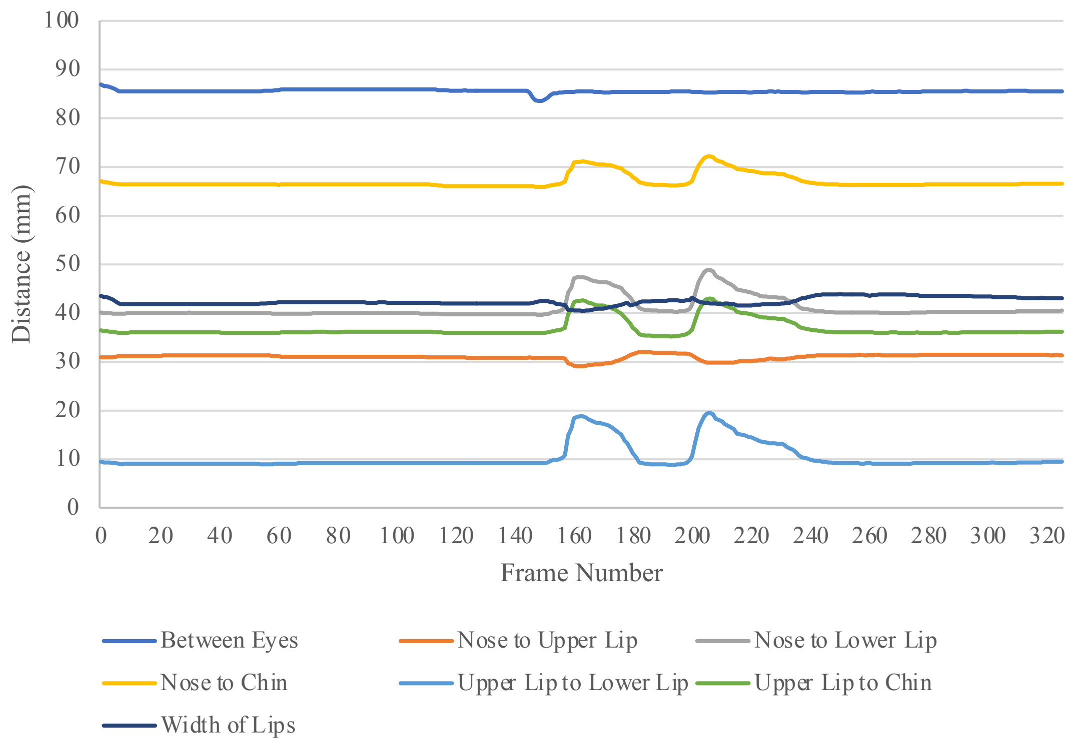

Recording 5
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Nose to Chin
Width of Lips

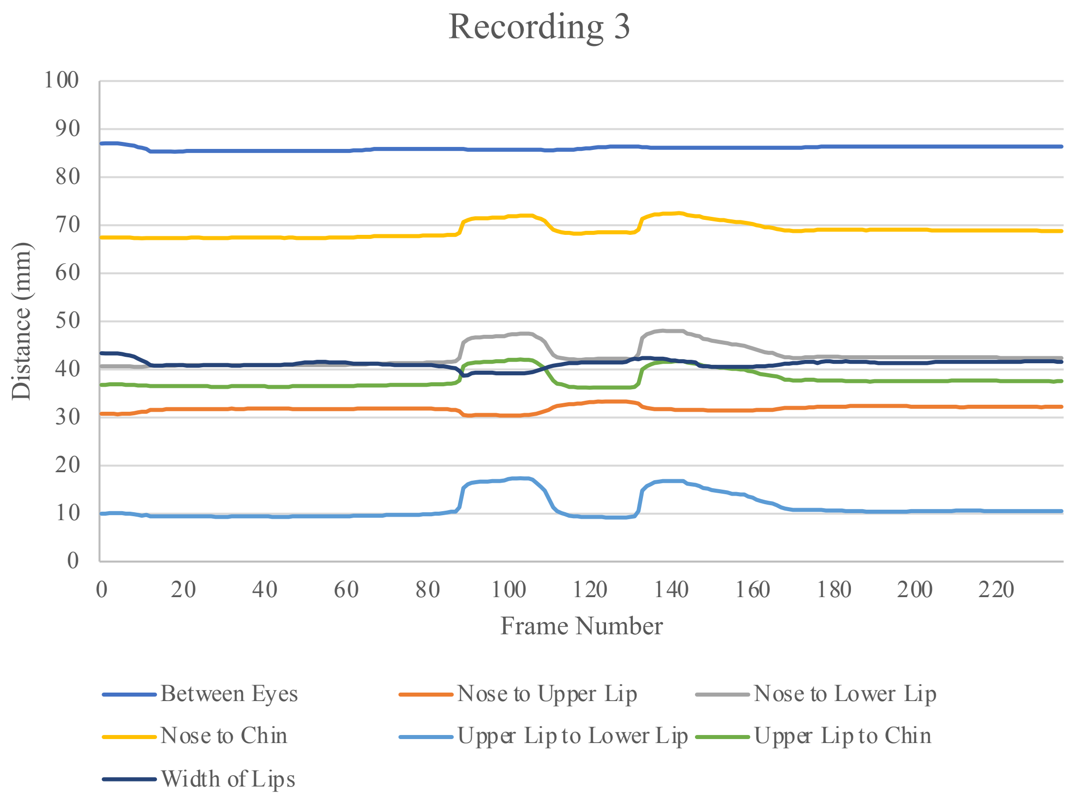
Movement Trajectories of all Tracked Points for 'bin'

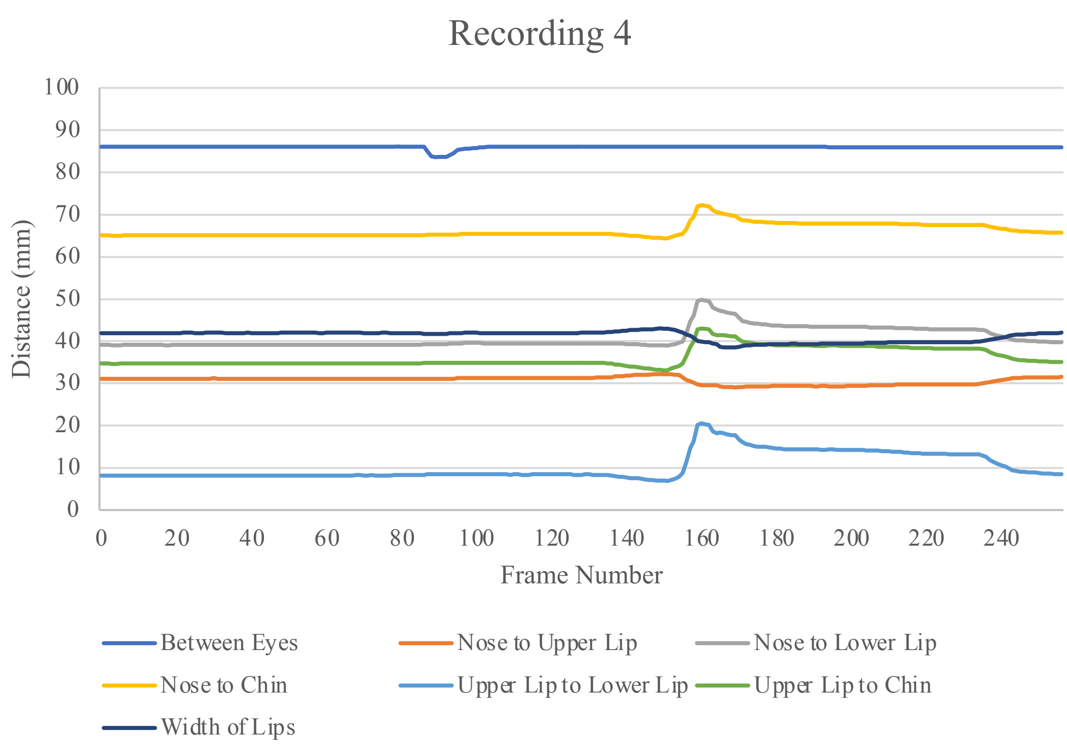
Recording 1



Recording 2



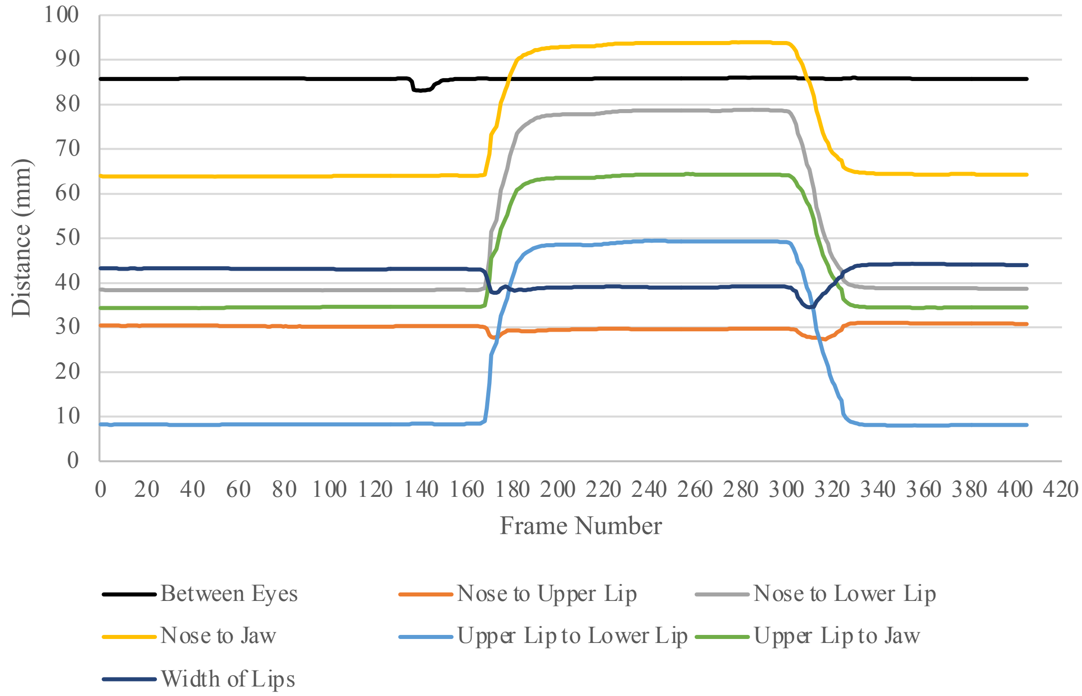




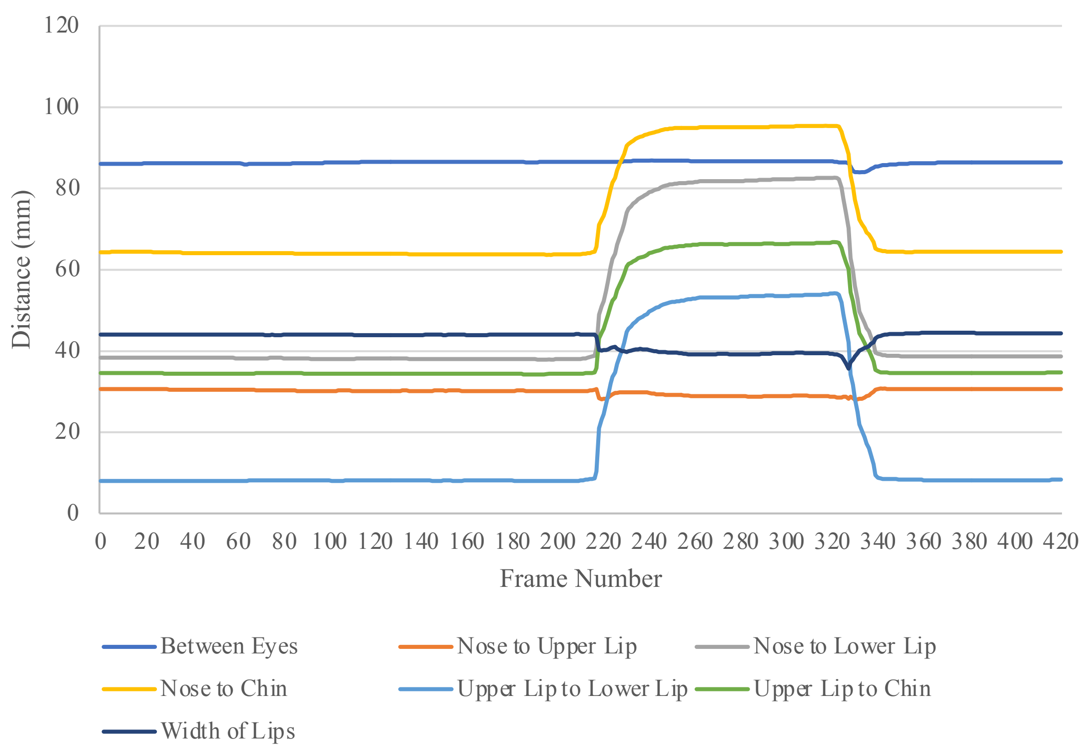
Due to unexpected data loss, recording 5 of ‘bin’ was not saved.

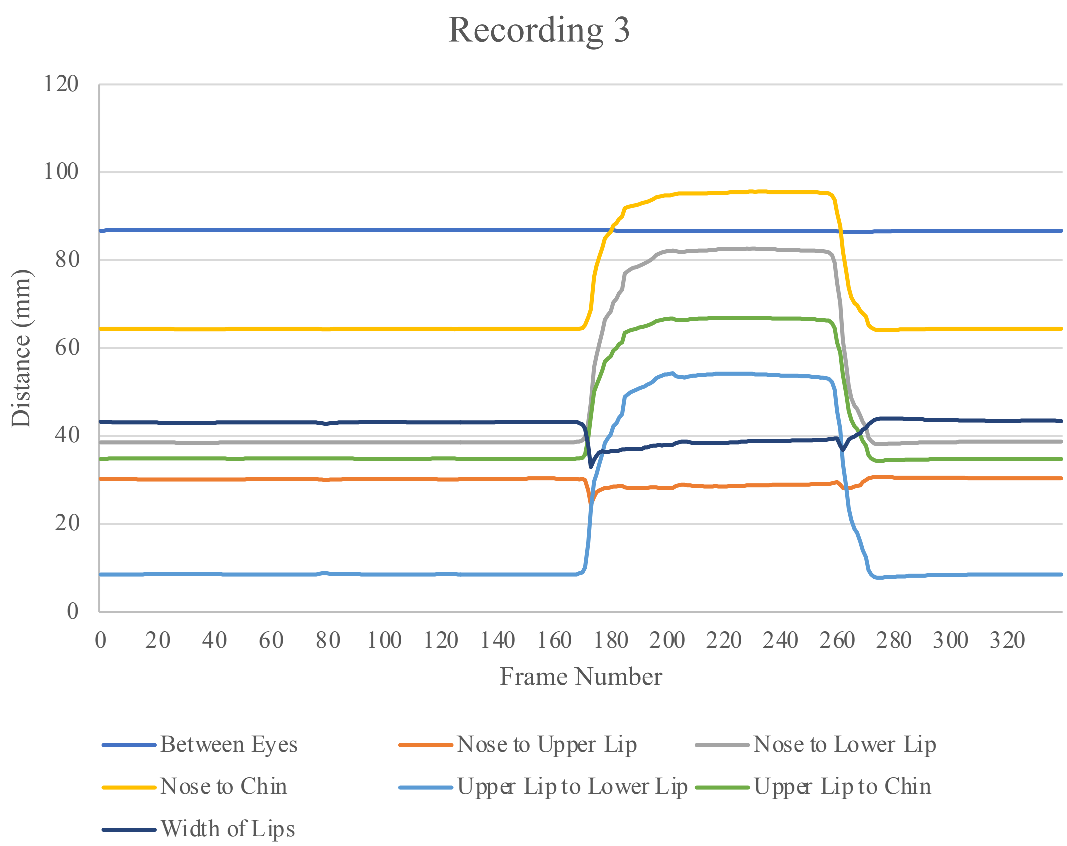
Movement Trajectories of all Tracked Points for 'ah'

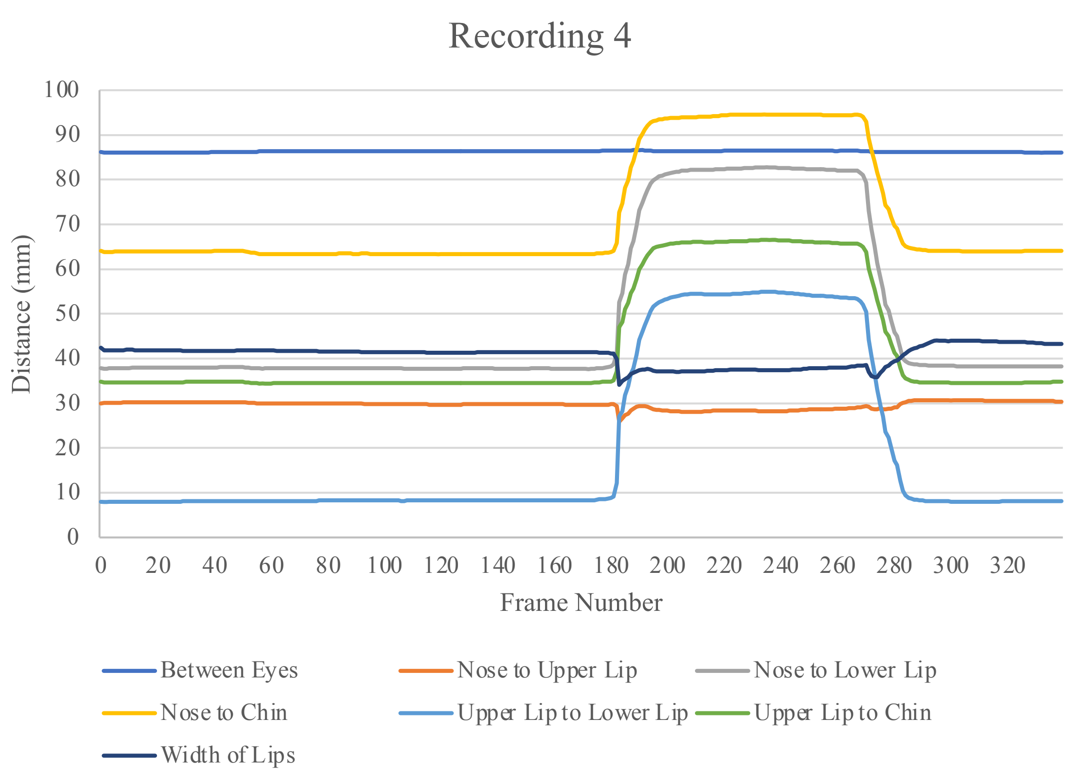
Recording 1

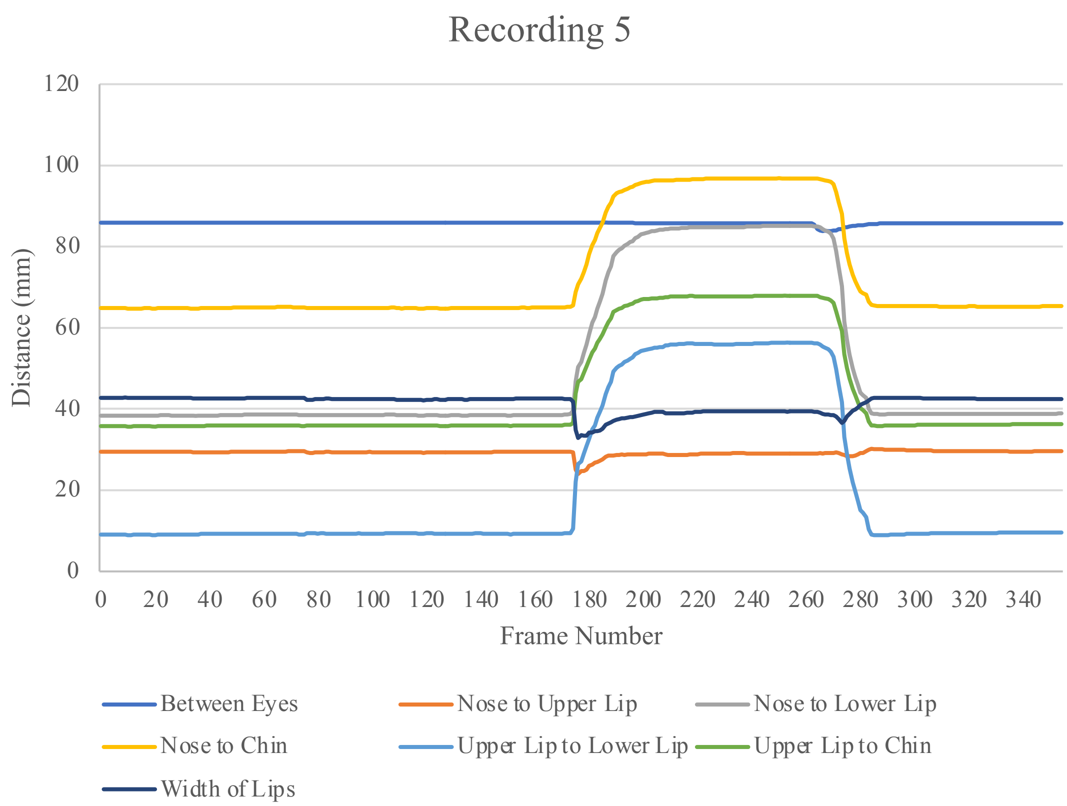


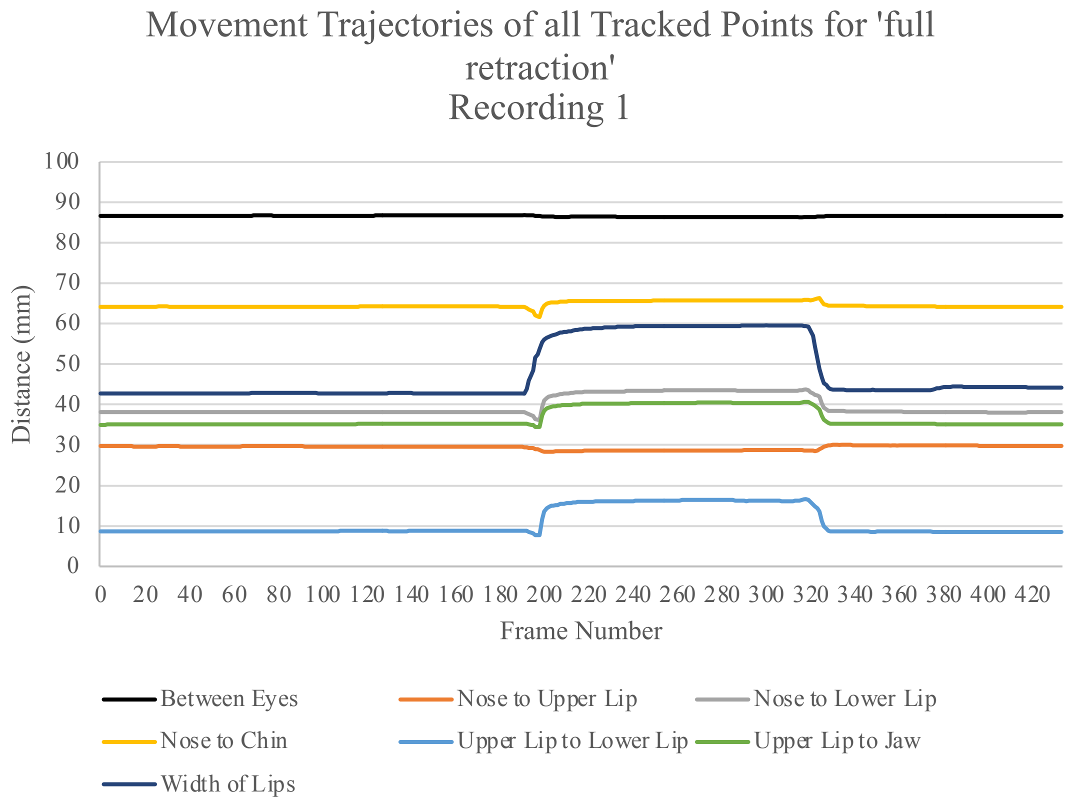
Recording 2

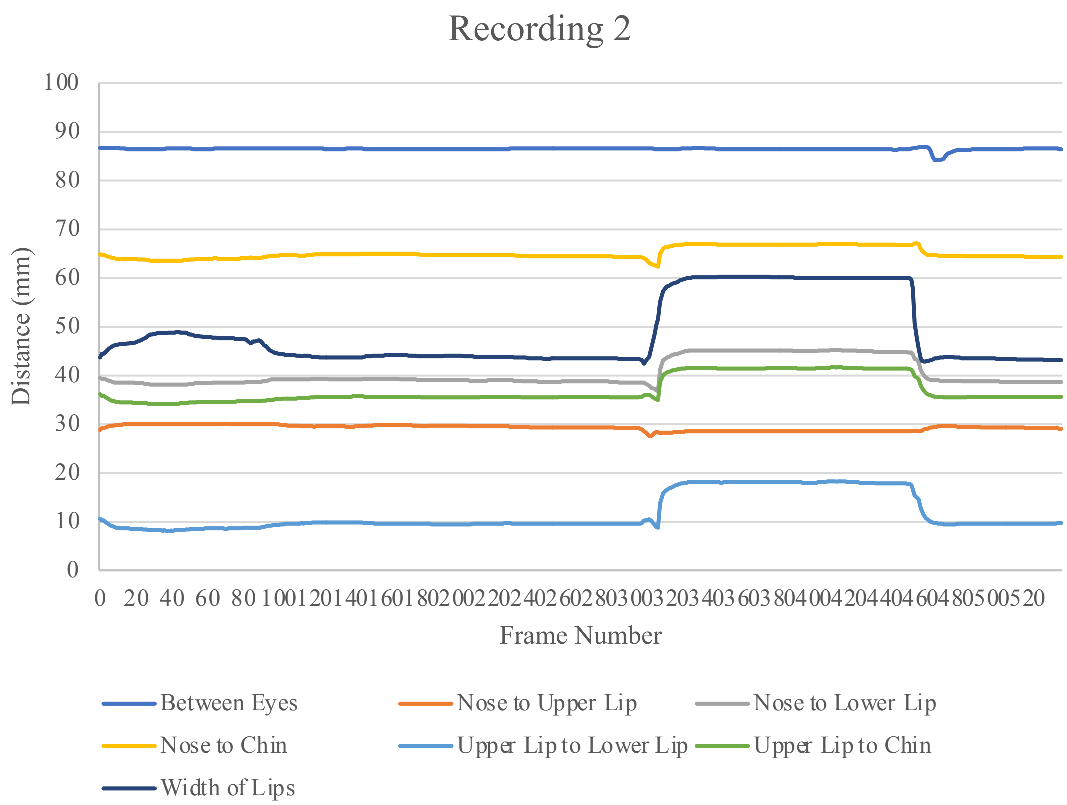


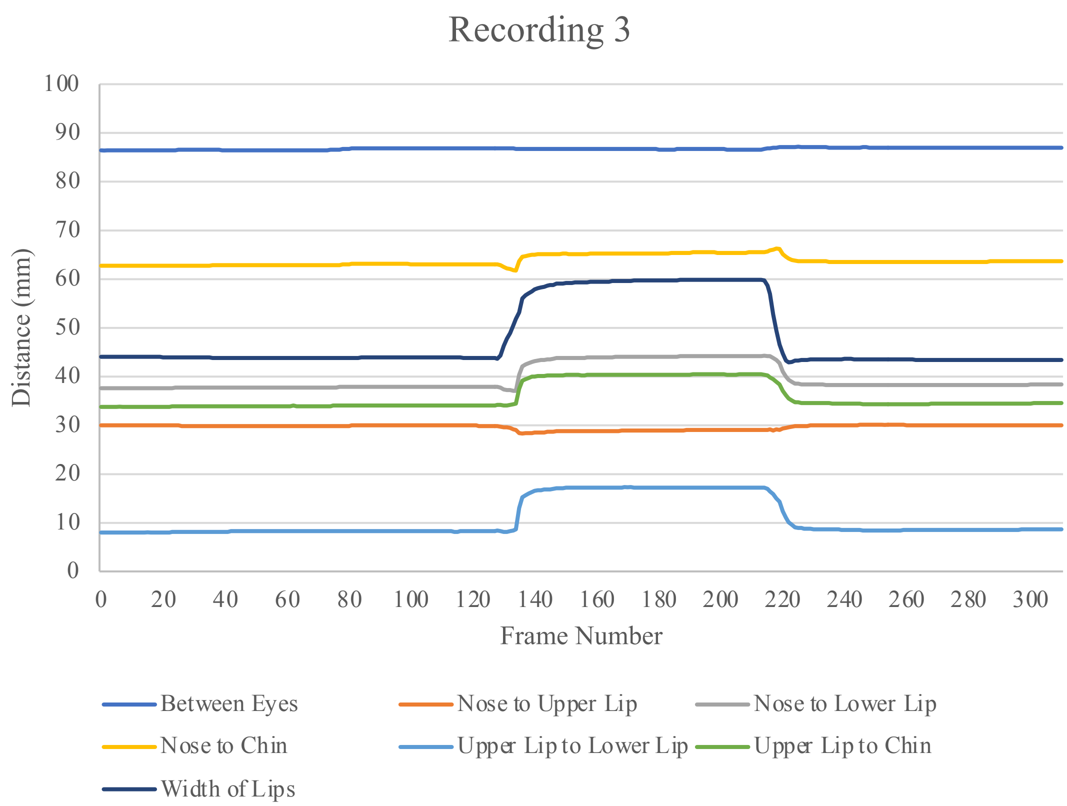

Recording 3
120
100
80
60
40
20
0
Distance (mm)
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

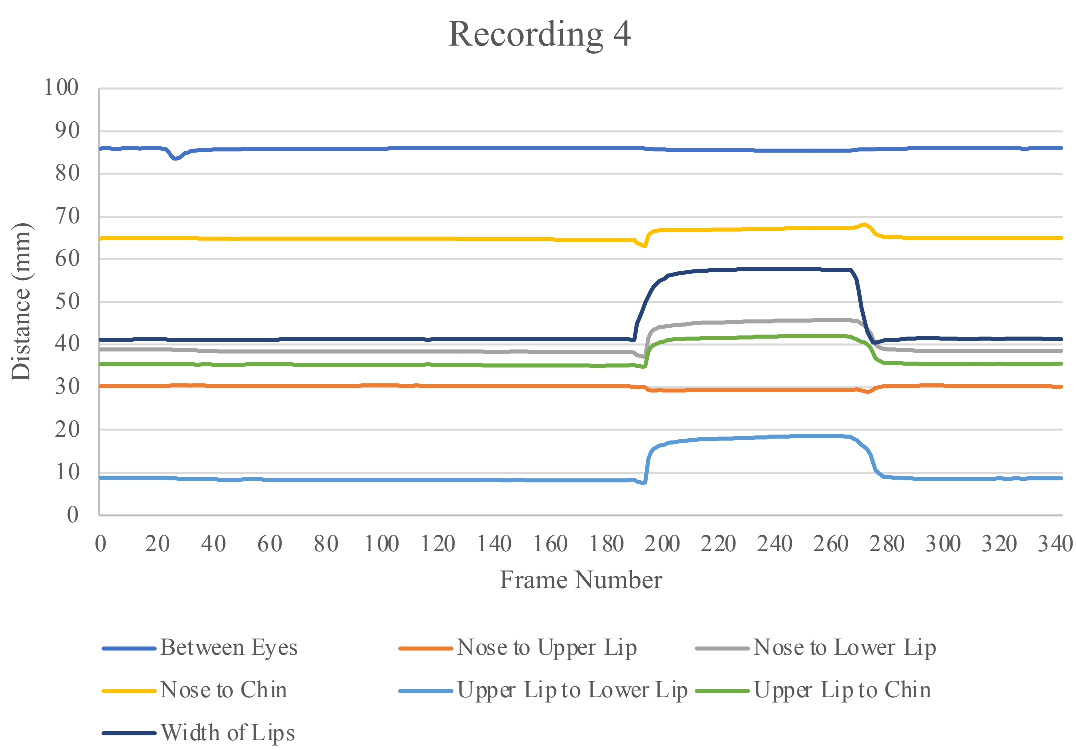

Recording 4
100
90
80
70
60
50
40
30
20
10
0
Distance (mm)
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

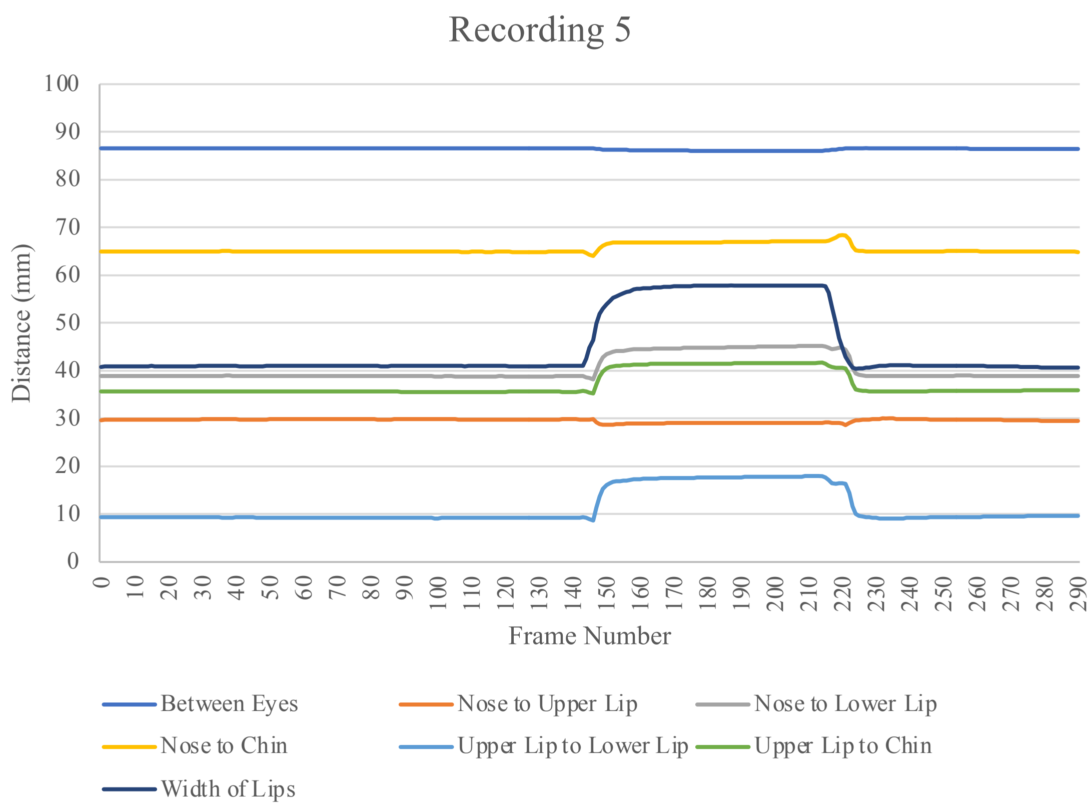

Recording 5
Distance (mm)
0
20
40
60
80
100
120
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips


Movement Trajectories of all Tracked Points for 'full retraction'
Recording 1
Distance (mm)
Frame Number
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420
0 10 20 30 40 50 60 70 80 90 100
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Jaw
Width of Lips


Recording 2
Distance (mm)
Frame Number
0 10 20 30 40 50 60 70 80 90 100
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips


Recording 3
100
90
80
70
60
50
40
30
20
10
0
Distance (mm)
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

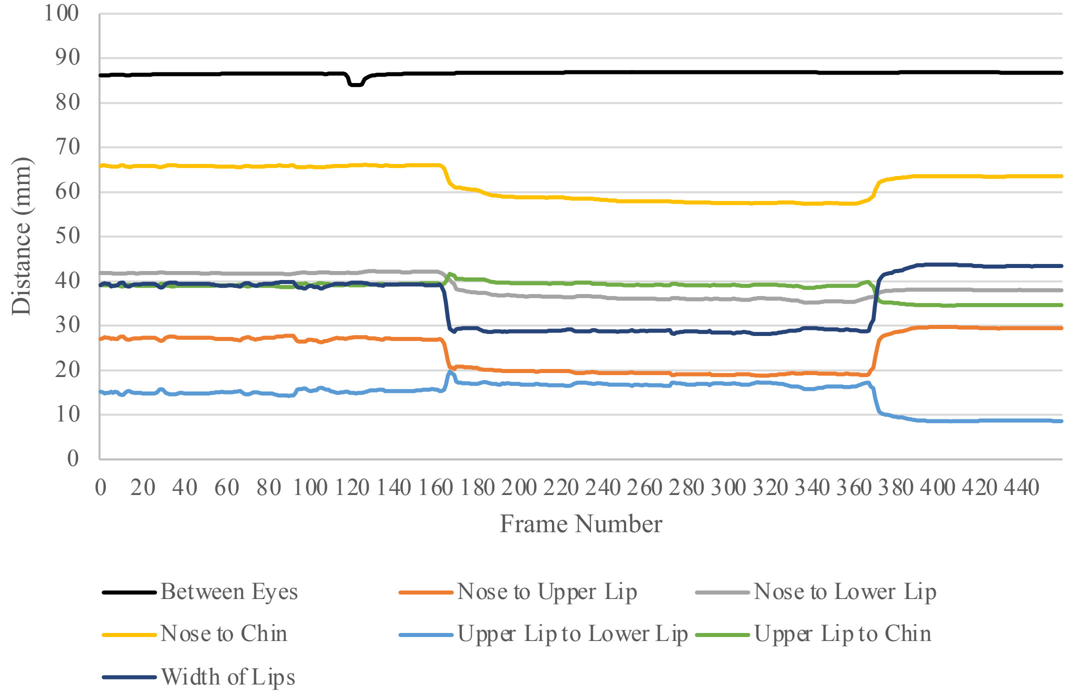

Recording 4
100
90
80
70
60
50
40
30
20
10
0
Distance (mm)
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips



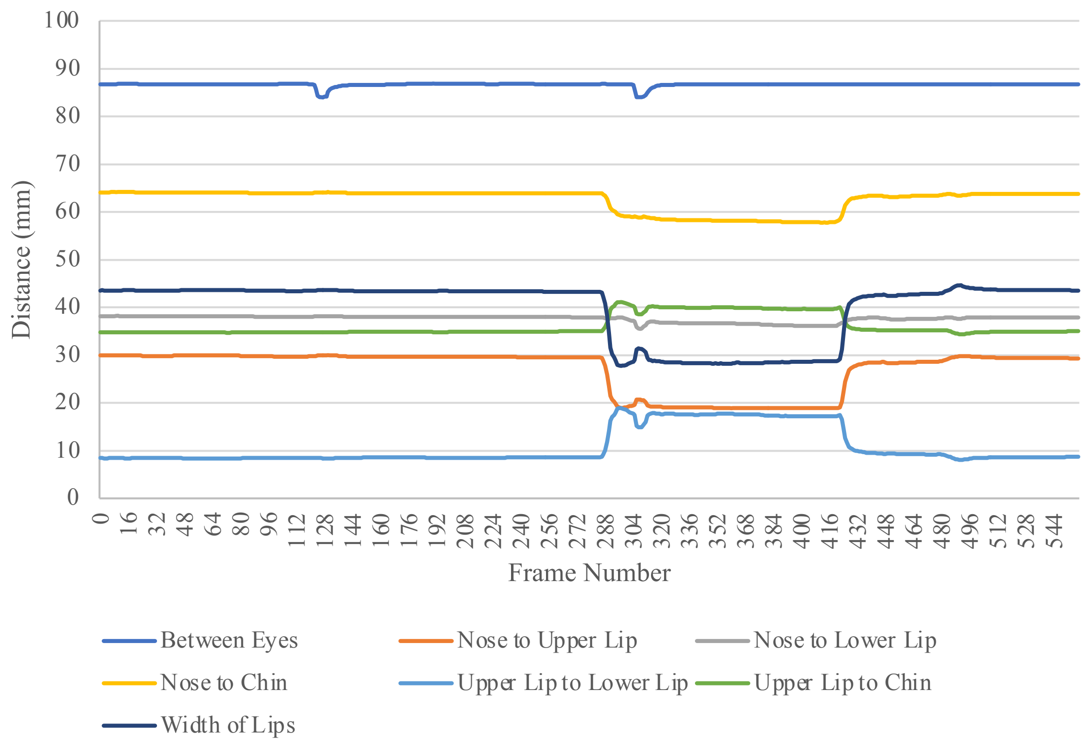
Recording 5
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

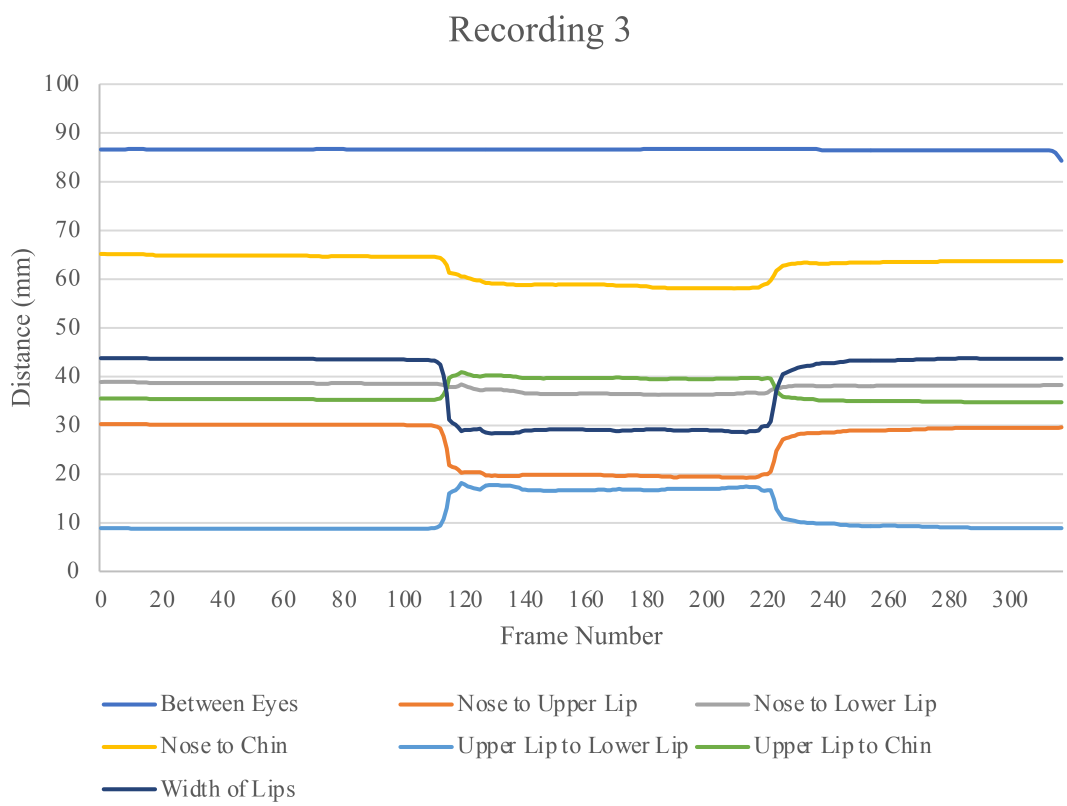
Movement Trajectories of all Tracked Points for 'purse'

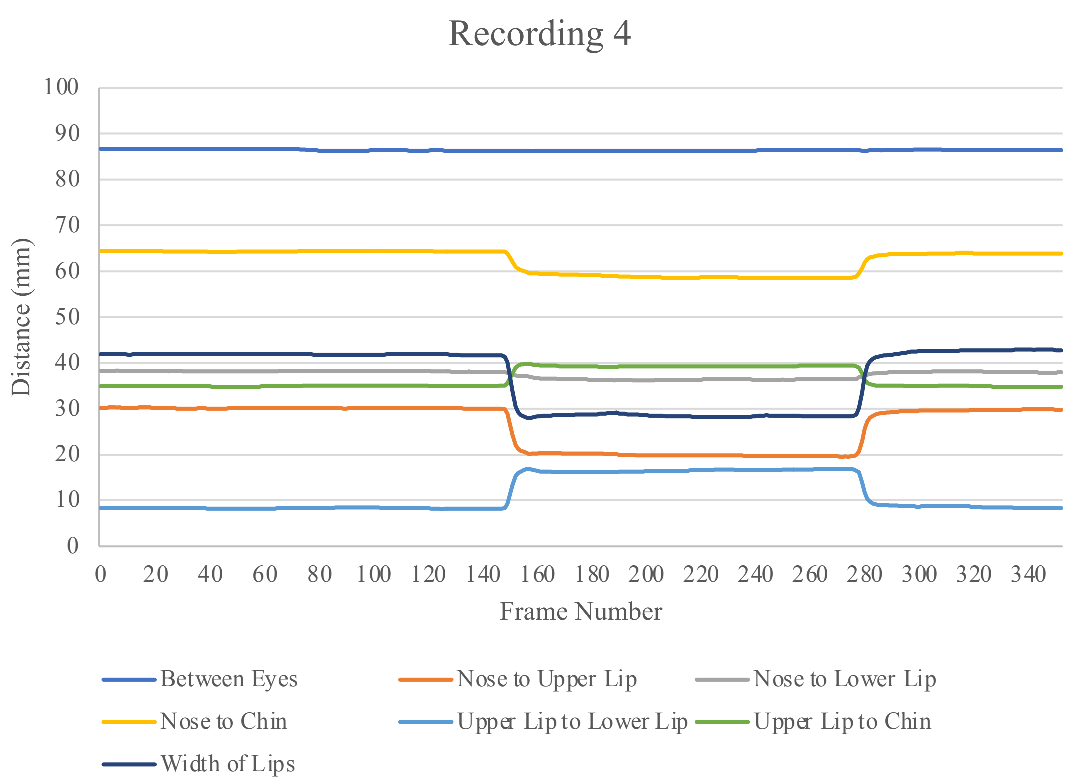
Recording 1

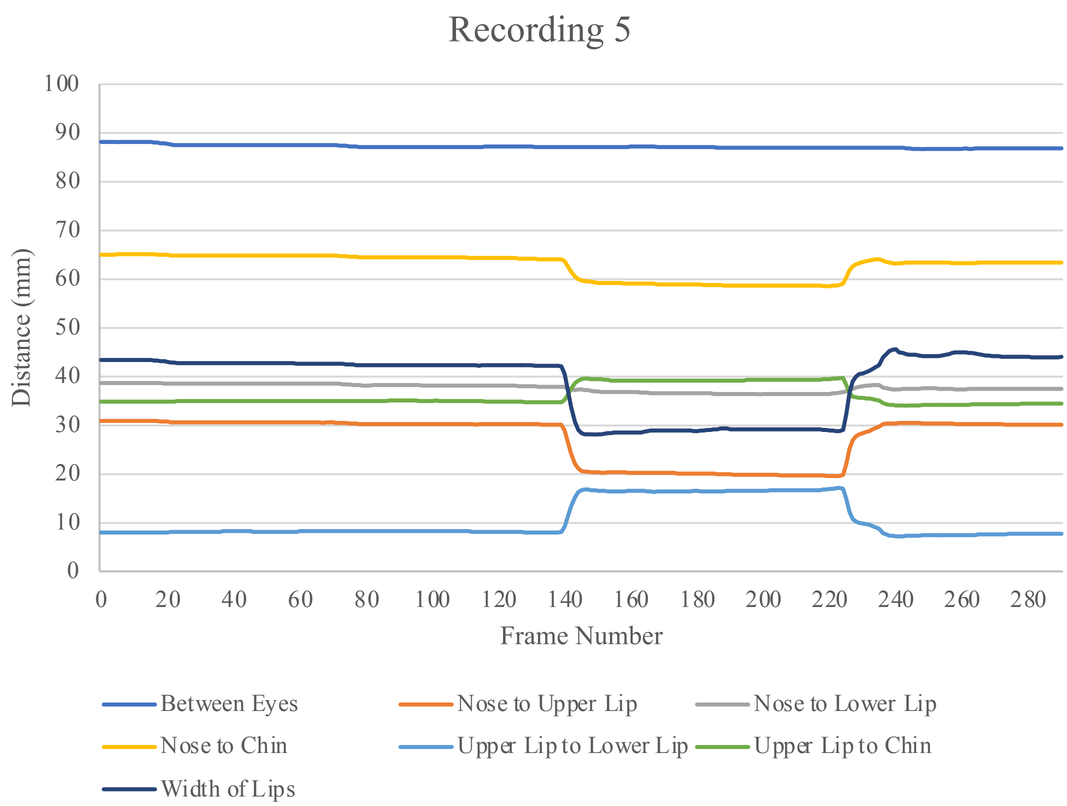


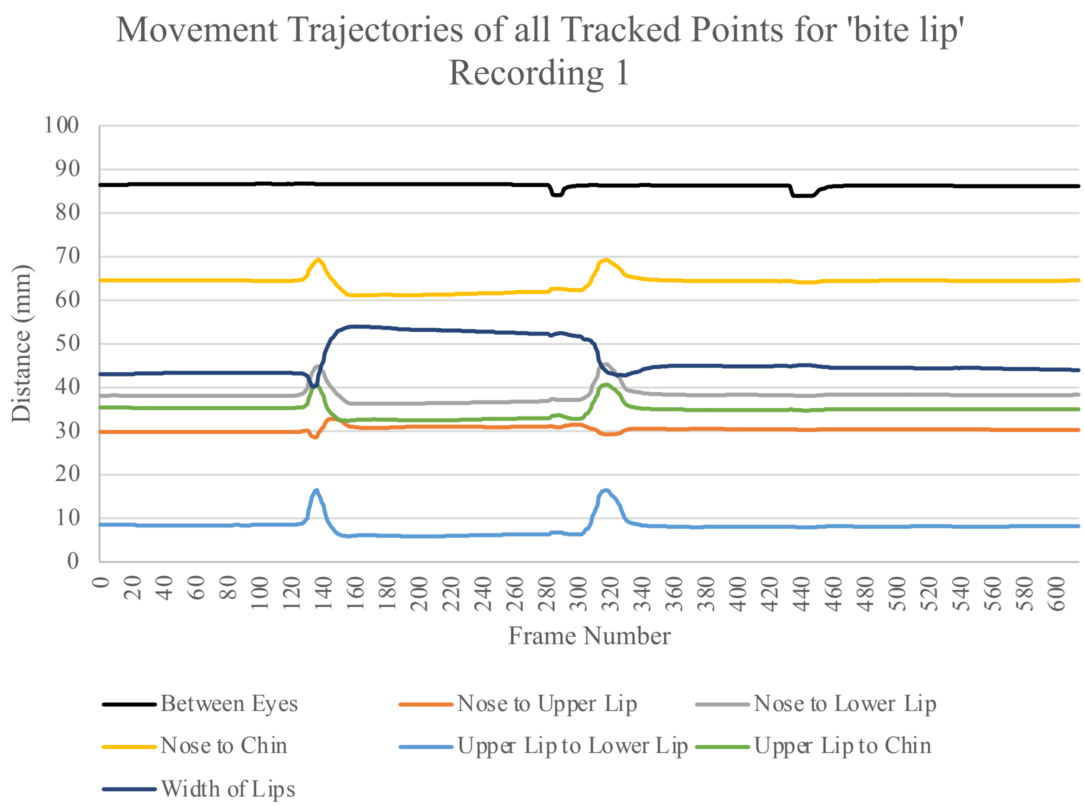
Recording 2



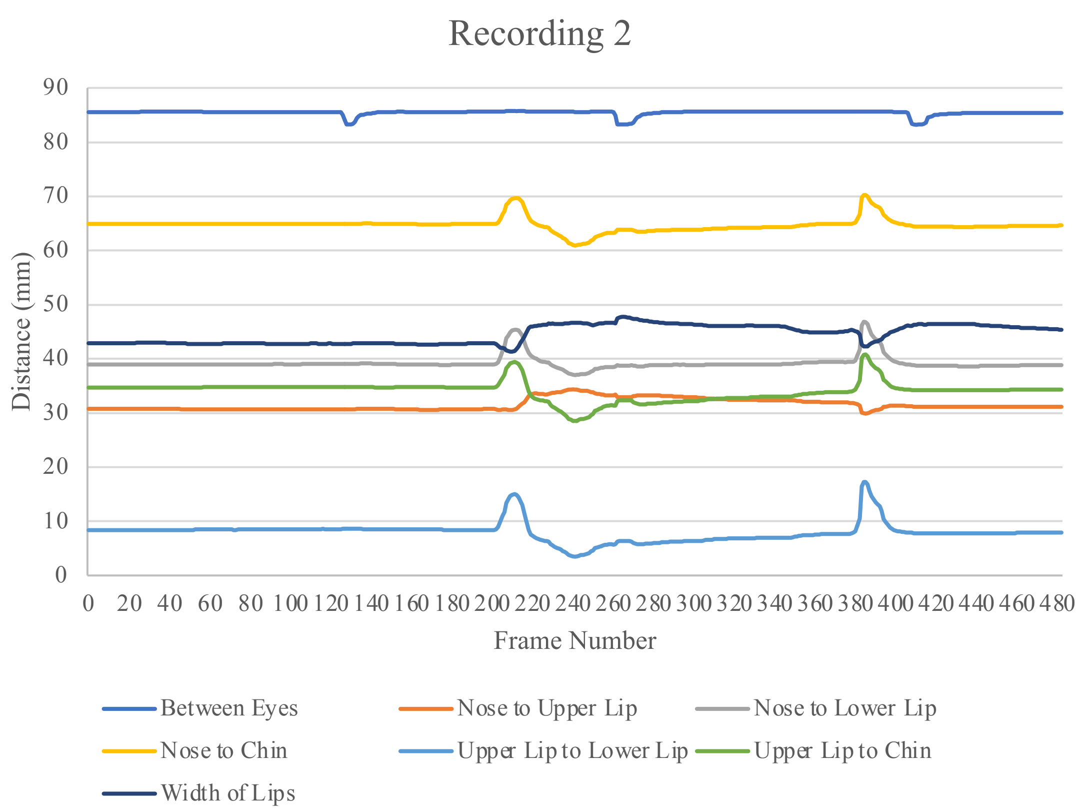

Recording 3
100
90
80
70
60
50
40
30
20
10
0
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300
Frame Number
Distance (mm)
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips


Recording 4
100
90
80
70
60
50
40
30
20
10
0
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340
Frame Number
Distance (mm)
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

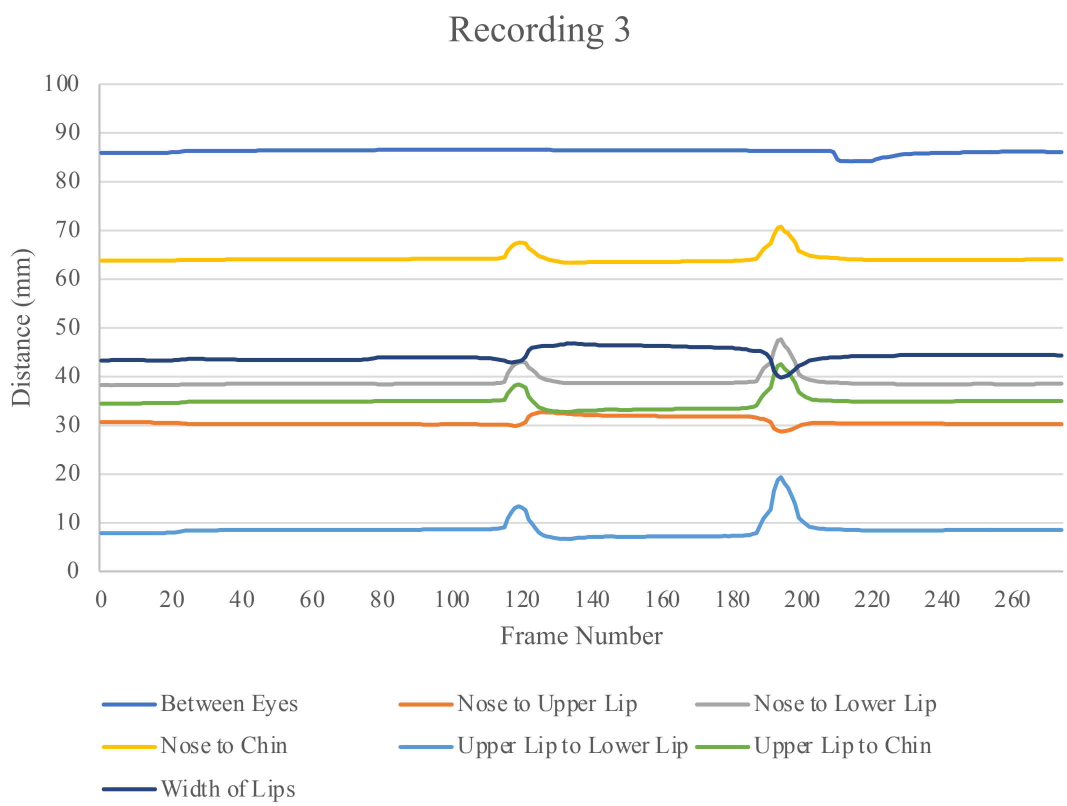

Recording 5
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

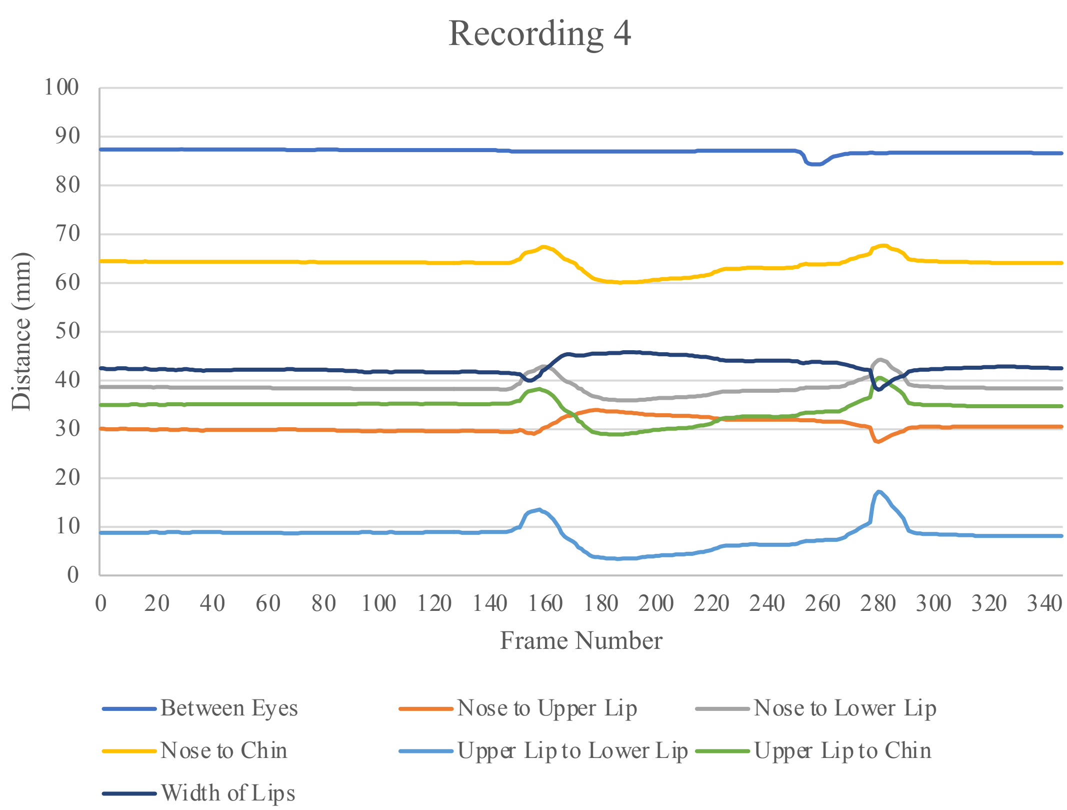


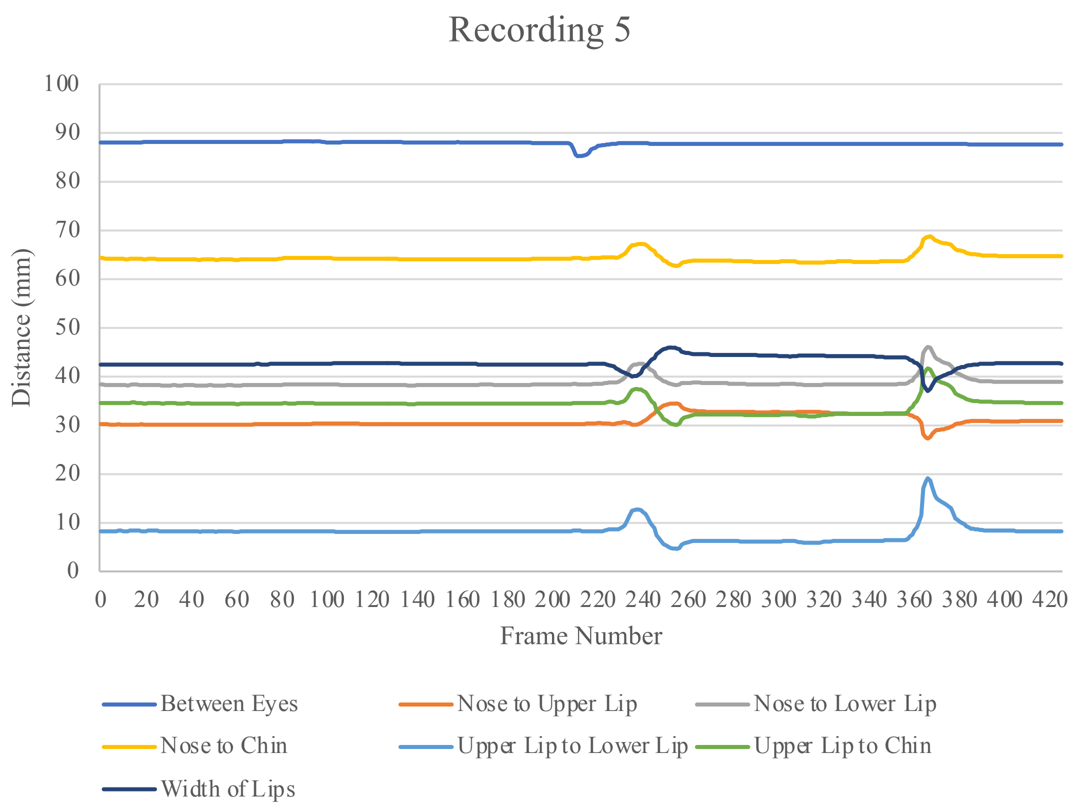
Movement Trajectories of all Tracked Points for 'bite lip'
Recording 1
Distance (mm)
0
10
20
30
40
50
60
70
80
90
100
Frame Number
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips



Recording 2
Distance (mm)
0
10
20
30
40
50
60
70
80
90
Frame Number
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

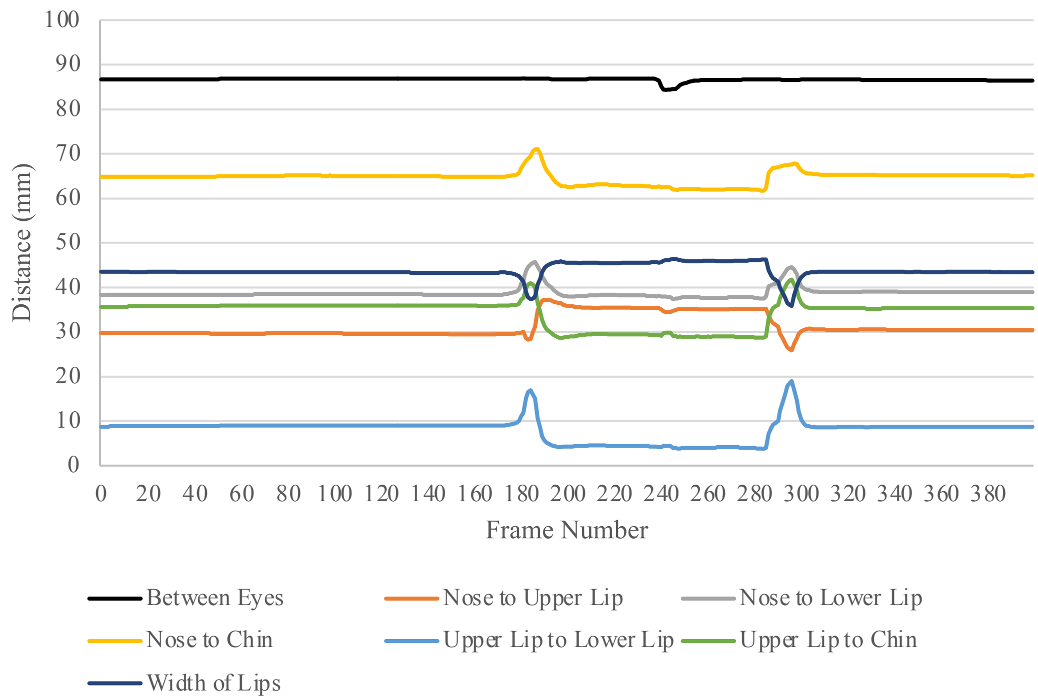

Recording 3
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips


Recording 4
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
340
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

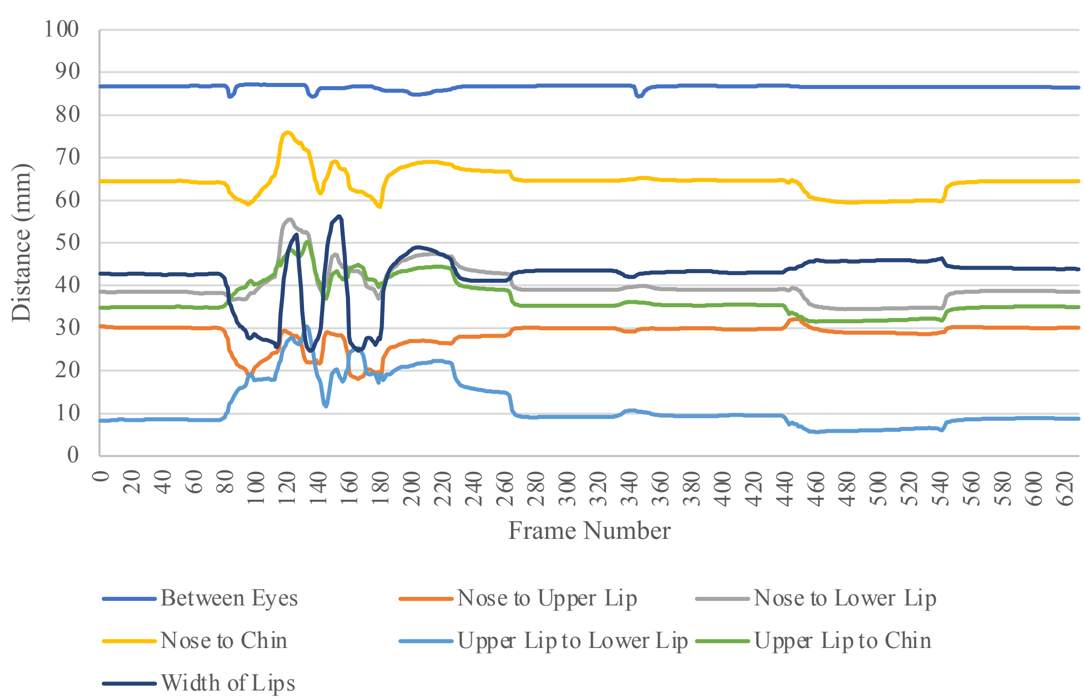


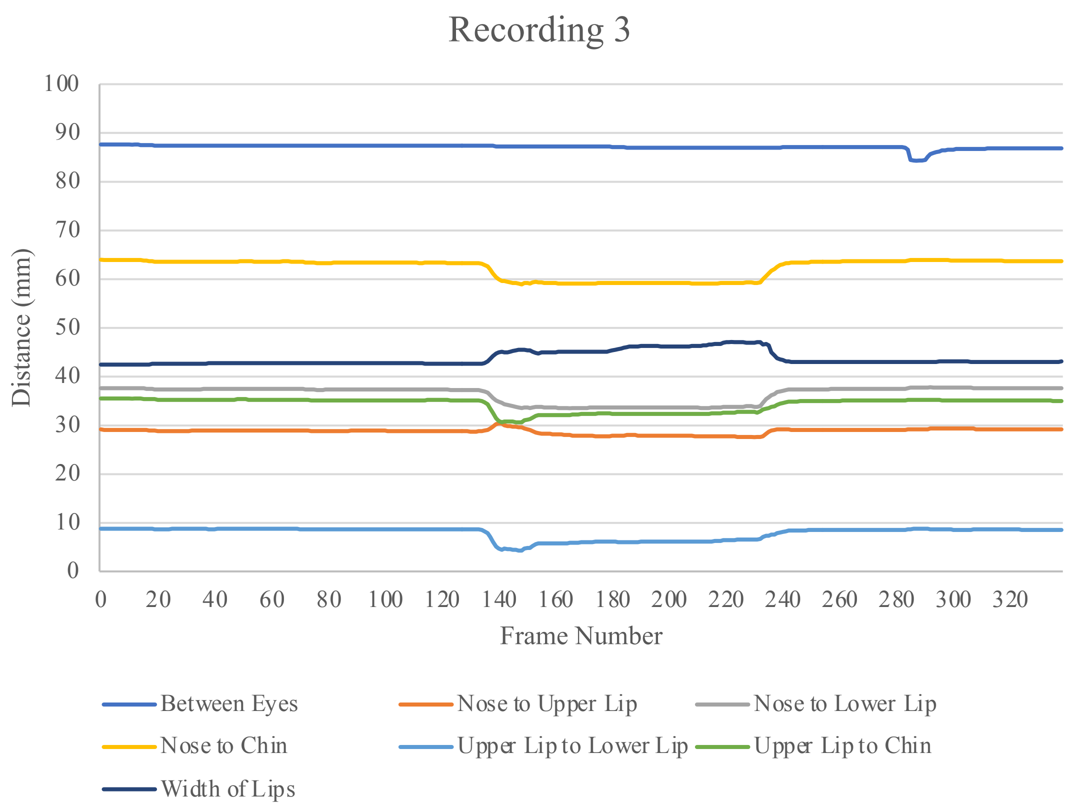
Recording 5
Distance (mm)
100
90
80
70
60
50
40
30
20
10
0
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips

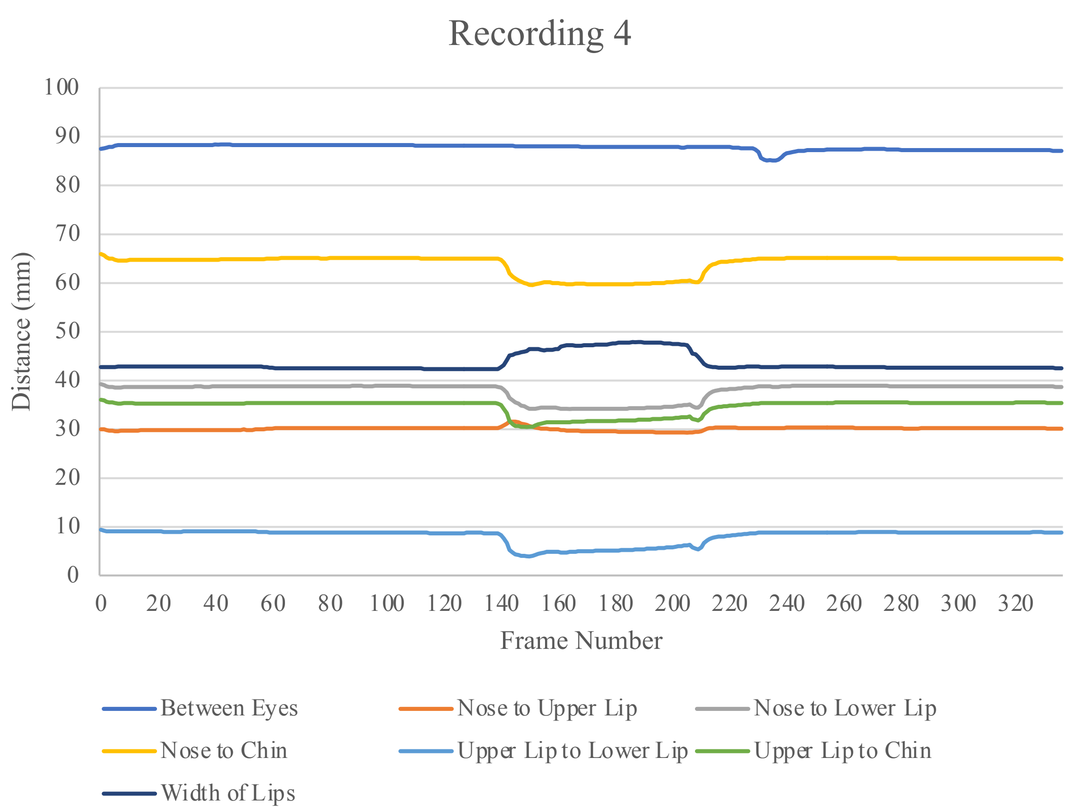
Movemejectories of all Tracked Points for ' hide lip'
Recording 1

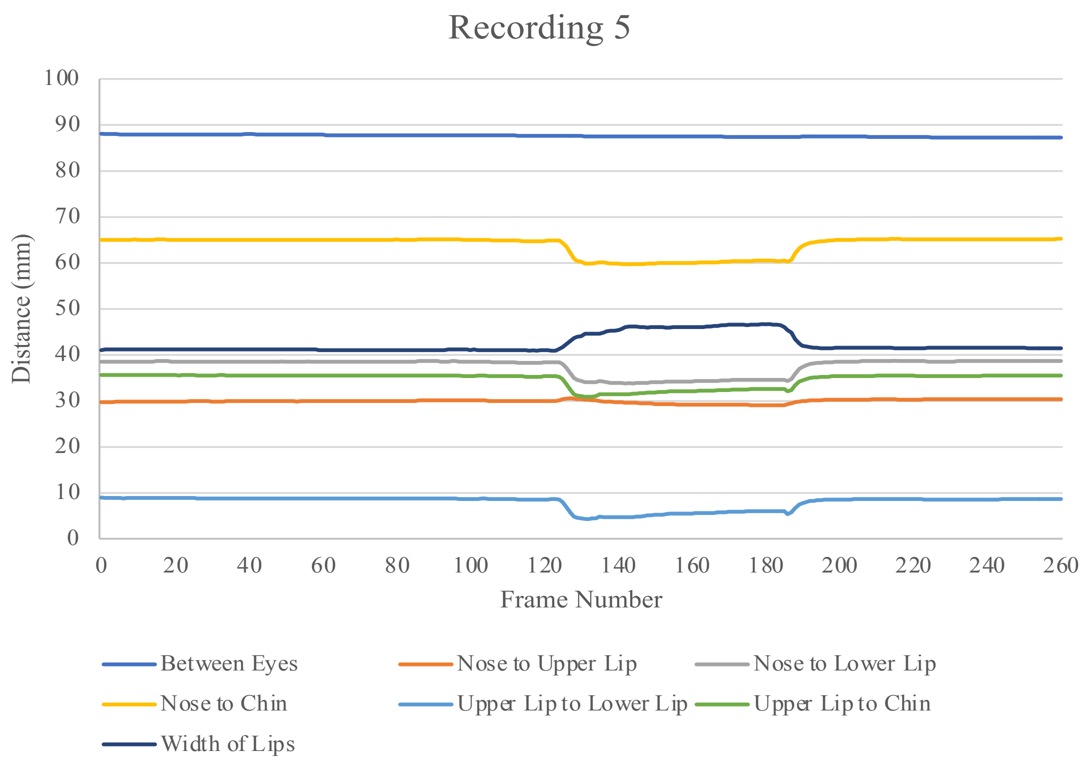


Recording 2




Recording 3
Distance (mm)
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips


Recording 4
Distance (mm)
Frame Number
Between Eyes
Nose to Upper Lip
Nose to Lower Lip
Nose to Chin
Upper Lip to Lower Lip
Upper Lip to Chin
Width of Lips



Note. Movement trajectories for all tracked points across all stimuli in Chapter 2, Question 1.