

SCHLIEREN IMAGING: VISUALIZATION OF AIRFLOW IN SPEECH

by

Jeffrey Rowell

B.A., Simon Fraser University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2015

© Jeffrey Rowell, 2015

Abstract

Schlieren imaging is a non-invasive research tool that enables real-time visualization of airflow through refraction of light. Used predominantly in aerospace and ballistics research, its suitability for observing airflow in speech was proposed nearly 40 years ago. To date, this potential has been virtually unexplored. The following proof-of-concept study investigates the visual correlates of nasal versus non-nasal airflow to provide a preliminary demonstration of the tool's ability to visualize aerodynamic events in speech. Simultaneous schlieren and audio recordings were made of three French nasal/non-nasal minimal pairs spoken by eight native-French speakers. These stimuli were presented to 10 raters in Video-only, Audio-only and Combined Audio-Visual formats. The raters coded each stimulus as either "nasal" or "not nasal". Accurate designation of Video-only stimuli was significantly above chance response ($p < .05$), indicating that the difference between airflow for nasal and non-nasal sounds can be visualized and perceived through schlieren imaging alone. Non-significant improvements were observed over time in the Video-only condition. Differences between Combined Audio-Visual and Audio-only stimuli were non-significant and likely influenced by a ceiling effect for the auditory information presented in both conditions. Further research is needed with more difficult auditory-perceptual tasks to explore potential supplementary advantages of schlieren visual feedback alongside auditory ratings of resonance. Future research may also benefit from improved training procedures for schlieren imaging. Nonetheless, schlieren imaging has promising potential for future implementation in both speech research and clinical applications, particularly for speech resonance disorders.

Preface

The idea for the research presented in this thesis was conceived by the author, who also chose the experimental design and performed all data collection, analysis and interpretation of the results.

Dr. B. May Bernhardt provided expertise in resonance disorders and articulation. Dr. Anthony Herdman contributed to the design, programming of stimulus presentation, and calculation of the d-prime statistic.

This study was reviewed and approved by the Behavioural Research Ethics Board of the University of British Columbia. The ethics certificate number is H14-03473.

Table of Contents

Abstract	ii
Preface	iii
Table of Contents.....	iv
List of Tables	vii
List of Figures	viii
Acknowledgements	ix
Dedication	x
Chapter 1: Introduction	1
1.1 Airflow in Speech	1
1.2 Schlieren Imaging	2
1.3 Nasality	7
1.4 Research Questions	9
Chapter 2: Methods	11
2.1 Speakers	11
2.2 Stimuli	11
2.3 Schlieren Setup.....	13
2.4 Rating Task	15
2.4.1 Raters	15
2.4.2 Rating Task Procedures	16
2.5 Data Analysis	18
2.5.1 Operational Definitions	20

Chapter 3: Results	22
3.1 Hypothesis 1: Visual Indications of Nasality	22
3.2 Hypothesis 2: Learning Effects	23
3.3 Hypothesis 3: Advantage of Combined Stimuli over Audio-only Stimuli.....	24
Chapter 4: Discussion.....	25
4.1 Visual Indications of Nasality	25
4.2 Hypothesis 2: Learning Effects	28
4.3 Hypothesis 3: Advantage of Combined Stimuli over Audio-only Stimuli.....	29
4.3.1 Ceiling Effects.....	29
4.3.2 Sensitivity and Specificity	31
4.4 Speaker Characteristics.....	34
4.5 Rater Characteristics	35
4.6 Limitations of the Study and Directions for Future Research.....	38
4.6.1 Confounds in Visual Stimuli	38
4.6.2 Challenges of Airflow	40
4.6.3 Issues in Quantification	42
4.6.4 Future Applications of Schlieren Imaging	43
Chapter 5: Conclusion	46
References	48
Appendices	53
Appendix A Raw Data	53
A.1 List of Hits, Correct Rejections, Misses and False Alarms.....	53
A.2 Rates for Hits, Correct Rejections, Misses and False Alarms	54

A.3	Values for d-Prime and Bias	55
	Appendix B Practice Session Instructions	56

List of Tables

Table 1 List of words presented to speakers	12
Table 2 Values for d-prime in “Video-only” conditions	22
Table 3 Paired sample t-tests between "Video-only" blocks 1, 2 and 3	23
Table 4 Mean d-primes for blocks 1, 2 and 3 by rater and condition	24
Table 5 Paired t-test between “Combined” and “Audio-only” conditions	24

List of Figures

Figure 1 Schlieren photographs	3
Figure 2 Single-mirror schlieren setup.....	14
Figure 3 Distribution of each rater's mean d-prime scores for Video-only condition	26
Figure 4 Sum of incorrect responses by trial order in the Video-only practice session.....	28
Figure 5 Ceiling effects for both Audio and Combined modalities	30
Figure 6 Histogram of response bias means for each condition	32

Acknowledgements

I wish to acknowledge my University of British Columbia thesis committee members, Dr. B. May Bernhardt, supervisor, and Dr. Anthony Herdman, for believing in my project. Their patience, guidance and dedication enabled me to pursue a thesis that I felt was rather unique, and a lot of fun.

I would also like to acknowledge and deeply thank Dr. Steve Rogak from UBC's Department of Mechanical Engineering for graciously lending me his schlieren equipment, and to his research assistant Lee Sutton who helped explain the use of the equipment. Without their contributions, this research would not have been possible.

Thank you to Dr. Stefka Marinova-Todd whose guidance and support during my first year of studies helped spark my interest in pursuing a thesis and enabled me to take on the current project.

A special thank you to those in UBC's Department of Linguistics, in particular Masaki Noguchi, whose shared interest, feedback, and generosity in lending recording equipment helped encourage me to keep going.

Thank you to the Social Sciences and Humanities Research Council (SSHRC) for the generous graduate scholarship that made this endeavour financially viable.

Finally, I would like to acknowledge my family, friends and classmates for all their encouragement, support and assistance throughout this project.

To my wife Liana

Chapter 1: Introduction

1.1 Airflow in Speech

Speech production is a multi-faceted human behaviour. Speech has been investigated with many lenses ranging from articulatory to acoustic, and as in this research, aerodynamic. Airflow, and therefore aerodynamics, is the underlying source of energy for all speech sounds (Miller & Daniloff, 1993). The Myoelastic and Aerodynamic Theory of Phonation describes sound production in terms of Bernoulli forces, where moving air in contact with the vocal folds causes self-sustained oscillations, resulting in the production of sound (van den Berg, 1958; Titze, 2006). Constrictions above the glottis may result in turbulent airflow (Stevens, 1971); this turbulence acts as a source of sound, such as in voiceless fricatives, and demonstrates the importance of airflow dynamics in speech beyond the vocal folds. Examples can also be found from disordered speech, where an incomplete closure of the velopharyngeal port may cause an uncontrolled escape of air from the nasal cavity. The resulting audible nasal turbulence is known as nasal emission, which is associated with hypernasality and is a common speech distortion in persons with cleft palate (Baylis, Munson & Moller, 2011). These examples demonstrate that research on the aerodynamics is vital to understanding how speech functions.

Several tools have been used to investigate airflow of speech: (1) *in vivo* observations with pneumatochographs and electroaerometers use face masks with gauges to measure air pressure; (2) hot-wire anemometers measure breath as it cools along a wire; (3) body plethysmographs calculate air pressure changes as the torso moves through an airtight chamber; and (4) distributed vocal-tract pressure transducers estimate airflow from intrapharyngeal air pressures measured by an array of small microphones (Miller & Daniloff, 1993). Bettens, Wuyts and Van Lierde (2014) also provide a thorough review of instrumental assessment of the

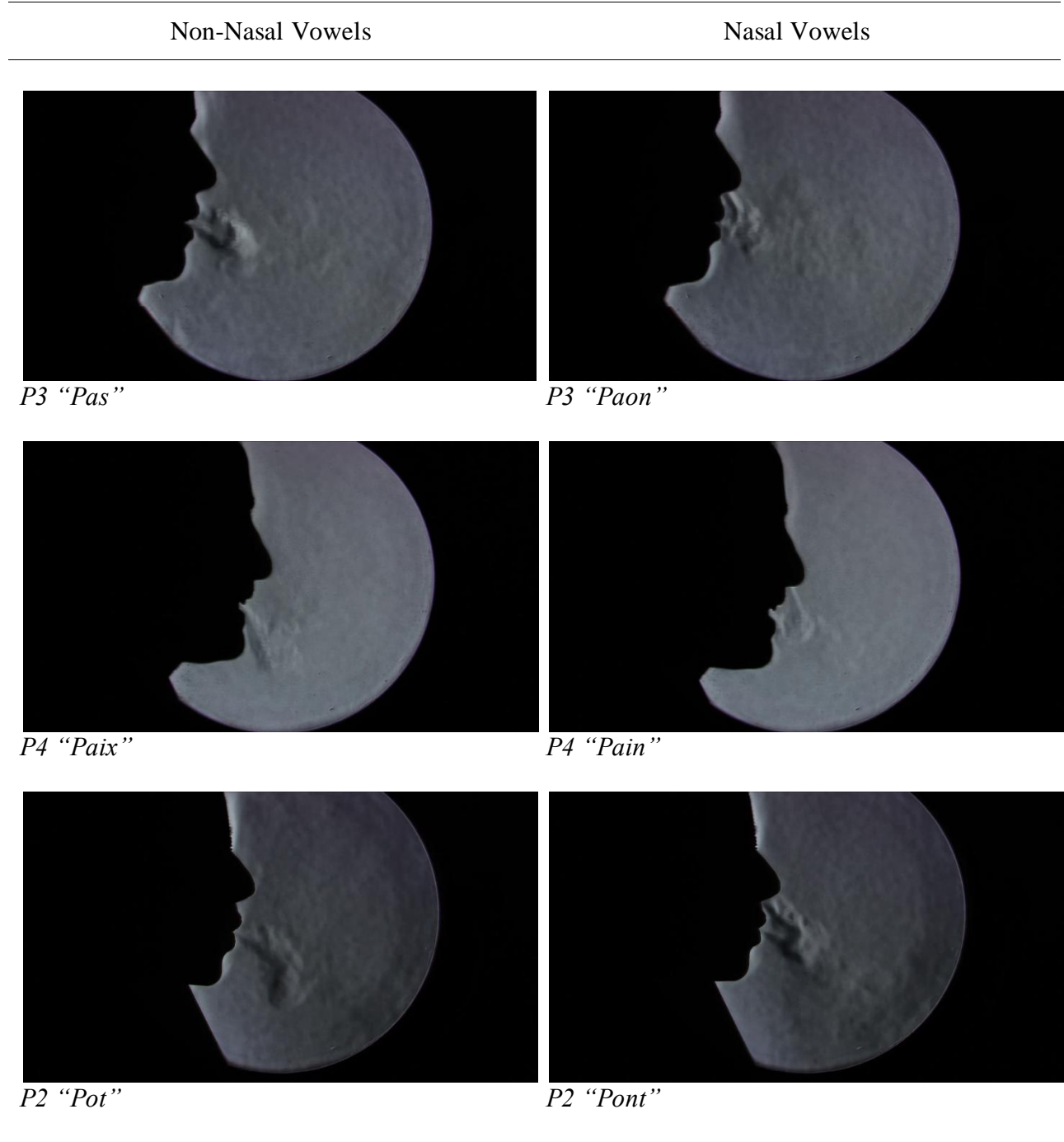
velopharynx, including aerodynamic measures. In their review they mention several additional tools such as the *aerophonoscope* which uses sensors placed on the nostrils and mouth to measure airflow. Thus, the toolbox of aerodynamic measures for speech in human participants is already well-established in both the research and clinical domains, yet there are still properties, such as the visible aspects of airflow in speech, that remain largely explored.

Flow visualization is the method of making invisible movement visible, and is typically accomplished by adding traceable substances into the air. Several studies have used this method to visualize airflow in models of the vocal tract, primarily in order to investigate aerodynamic events such as turbulence in relation to the glottis and larynx (Drechsel & Thomson, 2008; Khosla, Muruguppan, Gutmark & Scherer, 2007; Kucinski, Scherer, DeWitt & Ng, 2006; Neubauer, Miraghaie & Berry, 2007; Shinwari, Scherer, DeWitt & Afjeh, 2003). Despite its apparent utility, flow visualization of airflow in speech appears to be limited to models of the vocal tract, with little or no observation in human participants. There is to our knowledge only one such study which used white smoke and high speed imaging to visualize and measure turbulent flow of the sound /pa/ (Derrick, Anderson, Gick, & Green, 2009). Essentially, flow visualization has only scraped the surface of *in vivo* human speech, and further implementation of the technique may provide new insights for both research and clinical applications.

1.2 Schlieren Imaging

The current study used a visualization technique called schlieren flow visualization, or schlieren imaging, to investigate nasal versus oral airflow in speech. Schlieren imaging enables transparent media, such as air, to become visible. Figure 1 provides examples taken from this study's data demonstrating schlieren visualization of nasal versus oral airflow.

Figure 1 Schlieren photographs demonstrating the visual difference in airflow between non-nasal (oral) and nasal vowels. Only a single stream of airflow from the oral cavity is apparent in non-nasal vowels, whereas two streams are apparent in nasal vowels, coming from both the oral and nasal cavities.



In the past, the technique has been used with human participants to measure sneeze and cough trajectories in order to study the mixing of airborne pathogens (Tang et al., 2011). The technique has also been used in a similar fashion to observe airflow patterns in canine sniffing (Settles, Kester, & Dodson-Dreibelbis, 2003). Although these two studies come from different bodies of literature and are seemingly unrelated, it is worth noting that they both successfully made observations of airflow from the nasal cavity. In an admittedly bizarre fashion, it was sneezing and dog sniffing that first provided evidence to suggest that the current study was worth pursuing. Specifically, it was through watching a YouTube clip of Tang et al.'s (2011) recordings of sneezes that prompted the question: *but what does airflow look like in speech?*

Schlieren imaging works by capitalizing on the refraction of light; light bends as it travels through media of different densities or heat and is commonly observed as a visible rippling effect in the air above hot objects, such as a barbeque or the hood of a car. Mirages are also a result of this effect. Although its origins trace back to 1685, August Toepler is given credit for the invention and naming of the schlieren technique sometime between 1859 and 1864 (Settles, 2001). Toepler's term *schlieren* is the plural of *schliere*, from Old German and means "bits or pieces", but is now used in optics to refer to an object or area whose refractive index is different from its surroundings (Settles, 2001). Use of the term "schlieren imaging" in this study (including its dropped capitalization) is in emulation of Settles' own use of the term throughout his work, but the technique may be alternately referred to as schlieren flow visualization, schlieren photography or simply schlieren. Considering the age of the technique, it is easy to brush off schlieren imaging as antiquated. Indeed, its creation nearly coincides with the invention of photography, and just as traditional photography has been eclipsed through digital cameras in the age of computers, technological advances in flow visualization, such as laser interferometry

and particle image velocimetry, make schlieren imaging seem similarly outdated. These alternate techniques, however, may require the use of eye-damaging lasers or airflow laced with smoke or other visible particles. Schlieren imaging involves no such hazards, making its apparent outdatedness and simplicity a relative strength.

The schlieren effect also occurs in exhaled human airflow, albeit on a scale that is not visible to the naked eye. Because exhalations are typically warmer than ambient temperatures (Tang et al., 2011), light will bend differently as it travels through this slightly warmer plume of air. Schlieren imaging takes advantage of this by blocking out some of the refracted light, resulting in visualization of airflow in the form of shadows. Simply put, schlieren imaging is essentially a way to visualize shadows of airflow. Settles (2001) provides an in-depth description of the history, setup, and application of schlieren imaging and is an excellent resource for anyone interested in learning about this technique.

There are many advantages to schlieren imaging compared with its alternatives. As previously mentioned, a practical benefit of using schlieren imaging as a speech research tool is that it is completely non-invasive (Settles, 2001); participants have no physical contact with any equipment and are exposed to a small LED light source. This is in contrast to measurement tools that use surgical masks or contact microphones that may not only be uncomfortable for participants to wear, but may interfere with the movement of air. As previously mentioned, recent efforts to apply flow visualization to speech have used devices such as particle image velocimetry that require a tracing substance to make airflow visible. For instance, Neubauer et al. (2007) used a theatrical fog machine to permit visualization of flow through a model vocal tract, and Derrick et al. (2009) used white smoke to visualize airflow of /pa/ from a human participant. Seeding airflow with smoke may not always be practical in human research. Moreover, it may be

unethical depending on the type of smoke used. In contrast, schlieren imaging requires no such tracing substance and therefore does not suffer the same drawbacks. Lastly, schlieren imaging has the additional advantage of producing a "live image," which could be used to provide on-line biofeedback to participants. A useful comparison is nasopharyngoscopic biofeedback, where a flexible fiberoptic camera is inserted through the nose, permitting visualization of velopharyngeal valving. Use of this biofeedback tool has demonstrated excellent results in improving velopharyngeal dysfunction (Brunner, Stellzig-Eisenhauer, Proschel, Verres & Komposch, 2005). Schlieren imaging may provide feedback for virtually the same physiological event (i.e. closing of the velopharyngeal port) without the need for any equipment to be inserted into the patient's body. Brunner et al. also argued that biofeedback was beneficial because patients gained improved self-perception of their articulation; it is reasonable to expect that schlieren imaging may similarly increase patients' self-perception and control of articulation, albeit in a much less invasive manner.

The current thesis was designed to investigate schlieren imaging in speech research, although it is not the first study on this topic. In fact, it was originally argued nearly 40 years ago that the technique is well-suited for speech research and clinical use (Davies, 1979). It appears that there has been some effort to apply schlieren imaging in this way; a meeting abstract by Krane and Settles (2004) describes its use to investigate the interaction of airflow with the teeth and lips during articulation of /s/ and /z/. Unfortunately a full article was never published. The previously mentioned study by Tang et al. (2011) only references airflow of speech incidentally, making the observation that visible exhaled puffs of air from the nose and mouth differed between subjects while talking. It goes without saying that speech scientists are probably interested in a much deeper analysis than this. Through personal correspondence, Davies

explained that speech therapists have enthusiastically expressed interest in using schlieren imaging in the past, especially as a biofeedback technique; however, nothing has ever come of it (personal communication, September 21, 2014). As was the case for this study, it is possible that many universities already possess the essential components to conduct schlieren-based research, but a lack of inter-disciplinary knowledge (e.g. between linguists and mechanical engineers and their respective research tools) has prevented such an undertaking. This is unfortunate considering one can successfully set up and operate a schlieren system with minimal knowledge or experience with physics. Despite its potential as a speech research tool and apparent ease-of-use, a leading expert in schlieren imaging concluded that the research and clinical application of the technique has remained “entirely unexplored” (Settles, 2001). The current study seeks to take up this challenge by providing a preliminary investigation into the airflow patterns of nasality, and in doing so, introduce schlieren imaging to the speech science community.

1.3 Nasality

Nasality, as the term suggests, is the resonance of sound through the nasal cavity. It is typically produced by opening the velopharyngeal port, thereby coupling the vocal tract to the nasal resonance chamber; acoustic energy is transmitted into the nose, causing sounds to be perceived as nasal or nasalized (Maeda, 1993). Normally, this opening of the velopharyngeal port results in air flowing from the nose, meaning nasal airflow may indicate velopharyngeal function (Krakow & Huffman, 1993). Velopharyngeal function is important due to its ability to indicate contrastive meaning in certain languages and due to its role in speech disorders. In French, for example, the presence of nasality helps to distinguish the words *pain* (/pɛ̃/) from *paix* (/pɛ/) (Maeda, 1993; Carignan, 2014). Similarly, speakers with velopharyngeal dysfunction may

demonstrate different nasal airflow patterns compared with speakers with typical resonance; therefore clinical assessment of resonance disorders are often based on airflow (Krakow & Huffman, 1993).

The relationship between nasality and articulation is not perfectly linear. Carignan (2014) demonstrated that production of French nasal vowels involves idiosyncratic oral articulations, which ultimately result in similar acoustic signals despite variance between individual speakers. Carignan explains that this is because the goal of speech is to produce an accurate acoustic signal and not simply to perform articulatory gestures; in other words, when it comes to articulation of nasal sounds in speech, the ends justify the means. Evidently, if the relationship between velopharyngeal function and nasality is not so straightforward, then judging nasality based solely on nasal airflow may have its limitations. Krakow and Huffman (1993) argue this point stating that velopharyngeal opening and airflow are only related to a certain degree, beyond which the effects of velum shape have a nonlinear effect on airflow and cause minimal changes in the acoustic signal. Indeed, nasalization may even be possible in the absence of velopharyngeal opening if acoustic energy is able to travel through the tissue of the soft palate (Gildersleeve-Neumann & Dalston, 2001). Nonetheless, Krakow and Huffman (1993) go on to describe several tools that have been used successfully to study nasalization through airflow, indicating that the relationship between nasality, velopharyngeal function and airflow is sufficient to merit empirical research despite the relationship's nonlinearity. Moreover, for the purposes of the current study, it is not essential that nasality and nasal airflow correspond perfectly in order to investigate the presence or absence of nasal airflow. Nasal vowels are characterized by "some degree" of velopharyngeal opening (Carignan, 2014), and therefore the expectation that nasal airflow will be present during nasal vowels, at least to some degree, is reasonable.

1.4 Research Questions

Despite being proposed as a speech research tool nearly 40 years ago, the use of schlieren imaging in speech research has remained virtually unexplored. This lack of research is particularly surprising considering schlieren's potential utility as a non-invasive visualization tool for *in vivo* airflow of speech. The major purpose of this study, therefore, was to help fill this gap by investigating schlieren imaging's potential as a speech research tool. To demonstrate this potential, any number of the aerodynamic aspects of speech could have been selected, e.g., observations of the difference in turbulent/laminar airflow across manners of articulation, differences in airflow vectors according to place of articulation, or individual differences in the aerodynamics of any phoneme. After preliminary observations, nasality was chosen as the topic to demonstrate proof-of-concept based on the following reasons: (a) there is evidence that suggests nasal airflow is associated with nasality, meaning the phenomenon of interest was confidently expected to exist; (b) the presence or absence of nasal airflow was deemed simpler to investigate than airflow properties *within* a stream of air; and (c) successful proof-of-concept using nasality appeared to have stronger potential for direct clinical applications (e.g. evaluation or biofeedback for resonance disorders) compared with other aerodynamic speech events. Essentially, nasality seemed the least complicated in terms of aerodynamic activity, had support in the literature, and could potentially provide a more immediate contribution to clinical populations.

During the preliminary investigation, airflow appeared to follow the pattern of oral-only airflow for non-nasal sounds, and oral *and* nasal airflow for nasal sounds; thus an experiment was conducted to test whether qualified individuals, namely speech-language pathologists,

agreed with our preliminary observation. If nasal airflow is truly associated with nasality, then at the most basic level raters should be able to distinguish between nasal and non-nasal sounds greater than chance, based solely on schlieren imaging of visual airflow. Above-chance correct ratings would essentially be enough to demonstrate that schlieren imaging has *potential* for further speech research by indicating that aerodynamic correlates of acoustic phenomena can be visualized in airflow. Second, because raters would most likely be inexperienced with visual perception of speech airflow, a learning curve should be expected. Lastly, visual feedback could potentially strengthen auditory perceptual ratings of speech events. That is, raters should score higher when given stimuli that contain both audio and visual information, compared with audio alone. If true, this may demonstrate that schlieren imaging has potential to help clinicians perform more accurate assessments of resonance and hypernasality disorders.

The questions led to the following general research hypotheses:

- 1) Qualified raters' accurate discrimination of schlieren images as nasal versus non-nasal would be greater than chance.
- 2) Learning effects would be observed in raters' discrimination of schlieren images as nasal versus non-nasal.
- 3) Simultaneous ratings of combined audio and visual stimuli would be more accurate than ratings of audio-only stimuli.

Chapter 2: Methods

To investigate the visual aspect of nasal airflow, eight native French speakers were recorded using schlieren imaging while reading nasal/non-nasal minimal pairs aloud. These recordings were later edited into Audio-only files, Video-only files, and Combined (audio-visual) files, which were then rated by entry-level speech-language pathologists (SLPs) as nasal versus non-nasal. All participants were recruited through word of mouth and through advertisements distributed to faculty in the local speech-language pathology program. Speakers were compensated by either choosing \$10, or the chance to make a personal recording using schlieren imaging. Raters were compensated with the option of receiving \$10. Details of the research design follow below for speakers, stimuli, schlieren setup, rating methodology and analysis.

2.1 Speakers

The experimental stimuli were recorded from eight participants who were all native speakers of French (three female, five male); age-ranges 19-29 ($n = 2$), 30-39 ($n = 3$) and 40-49 ($n = 3$). Six speakers identified their French dialect as “Quebecois”, one as “Vosgien”, and one did not identify a dialect. Nine speakers were originally recorded, but one speaker’s data were eliminated from the dataset due to technical issues during recording, resulting in a total of eight speakers.

2.2 Stimuli

Stimuli were a list of French minimal pair words. Selection of minimal pairs was adapted from Carignan (2014). Table 1 shows the complete list of target words. The speakers read the list

of words three times at a natural pace with a pause between each word. The participants also recorded a list of practice words beforehand to become acquainted with the process.

Table 1 List of words presented to speakers adapted from Carignan (2014)

Vowel	/p/	/t/	/k/
/a/	<i>pas</i> /pa/ ‘step’ <i>papa</i> /papa/ ‘daddy’	<i>ta</i> /ta/ ‘your’	<i>cas</i> /ka/ ‘case’ <i>caca</i> /kaka/ ‘poop’
/ã/	<i>paon</i> /pã/ ‘peacock’	<i>temps</i> /tã/ ‘weather’	<i>quand</i> /kã/ ‘when’
/ɛ/	<i>paix</i> /pɛ/ ‘peace’	<i>tait</i> /tɛ/ ‘keep quiet’ <i>taie</i> /tɛ/ ‘cover’	<i>quai</i> /kɛ/ ‘platform’ <i>paquet</i> /pakɛ/ ‘package’
/ẽ/	<i>pain</i> /pẽ/ ‘bread’	<i>teint</i> /tẽ/ ‘complexion’	<i>coquin</i> /kokẽ/ ‘scoundrel’
/o/	<i>pot</i> /po/ ‘jar’	<i>tôt</i> /to/ ‘early’	<i>coco</i> /koko/ ‘coconut’
/õ/	<i>pont</i> /põ/ ‘bridge’	<i>thon</i> /tõ/ ‘tuna’	<i>con</i> /kõ/ ‘idiot’

Stimuli consisted of individual words in French, and were therefore categorized and defined as nasal or non-nasal according to which word was used. The words *Pain*, *Paon*, and *Pont* were defined as nasal, and the words *Paix*, *Pas*, and *Pot* were defined as non-nasal/oral. Use of these stimuli was based on Carignan (2014) who argued that *Paix/Pain*, *Pas/Paon* and *Pot/Pont* were each valid minimal pairs. Because all participants in this study were native French speakers with

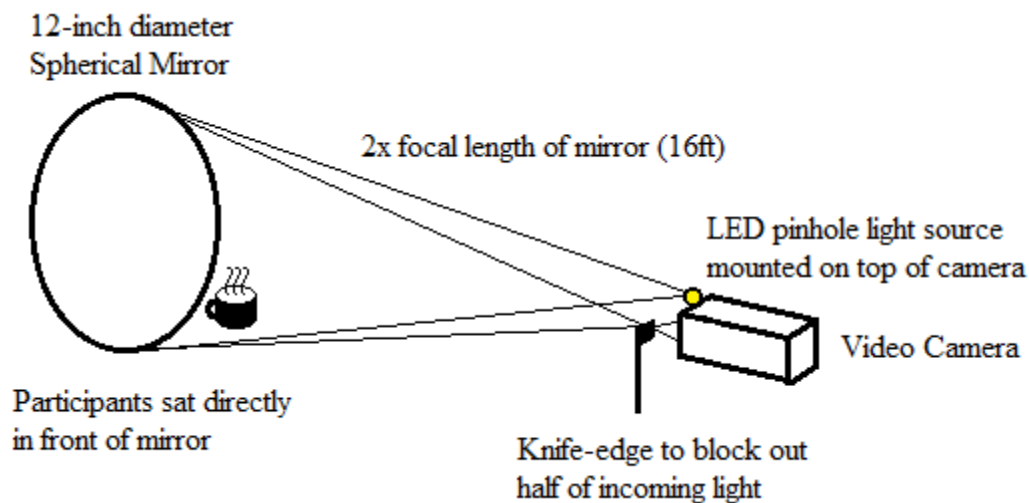
no reported speech disorder, these stimuli could serve as typical exemplars of nasal/non-nasal contrasts.

2.3 Schlieren Setup

The schlieren system used in this study was similar to the single mirror setup described in Settles (2001), with slight adjustments emulating Tang et al. (2011). Although two mirror arrangements exist, it was decided to use a single mirror due to its simplicity, as well as space constraints. A rough schematic of the current study's schlieren setup is provided in Figure 2. The system consisted of: (1) a 12-inch diameter spherical mirror with a focal length of 8 feet attached to an optical mirror mount which allowed fine adjustments in both the horizontal and vertical planes; (2) a battery-powered pinhole light source created by poking a small hole in aluminum foil covering a bicycle LED light which was attached to the top of the camera; (3) a video camera capable of recording in high definition at 60 frames per second; and (4) one razor blade attached to an adjustable optics mount which permitted fine tuning of the blade's position. The light source was mounted separately from the camera throughout the preliminary stages of the experiment. After difficulty found aligning the light source with the camera, it was discovered that simply attaching the light to the camera (whether on top or to the side) simplified the calibration process. The mounted mirror was securely attached to a table and elevated roughly 4 feet from the floor. An adjustable chair was used to position each participant's face appropriately in front of the mirror. Participants were positioned left-of-centre of the mirror in order to capture the maximum amount airflow while ensuring the participant's mouth remained on screen. The camera was zoomed to where the mirror's top and bottom touched the edges of the camera's viewfinder. The decision to have speakers facing right rather than left was arbitrary. Participants

were also positioned as close to the mirror as possible while keeping sufficient distance so as not to accidentally touch the mirror, thereby throwing the system out of calibration. Lastly, a microphone was placed 12 inches from the right side of the mirror; this distance was required due to available countertop space.

Figure 2 Single-mirror schlieren setup. LED light source is positioned on top of the camera.



Despite best efforts to maintain a perfectly consistent arrangement of the equipment, data collection over several weeks made it impossible to leave the equipment in place. Positions for every component of the device were marked with each piece being replaced as closely as possible to its original location. Camera zoom and brightness for recordings of new participants were compared with previous recordings in order to ensure a level of consistency in how large and bright the mirror appeared on the camera's viewfinder. Unfortunately, small changes over time due to vibration and movement are unavoidable, meaning the schlieren system would still need recalibration even if the system was not dismantled between participants. By the end of

stimuli collection, however, the study's primary researcher was able to set up and calibrate the entire system in less than five minutes.

Preliminary observations made it quite clear that disturbances caused by other sources of air movement had to be controlled. For example, schlieren imaging also allows for the visualization of body heat, which became an issue during trials as the heat entered the observation area of speech airflow. Several methods of redirecting or containing this body heat were attempted during trials, but the most effective appeared to be a combination of wearing a jacket, placing a yoga mat over the participant's legs, and having the participant hold a piece of cardboard over their chest. Although not elegant, this strategy worked well. The researcher was also unaware of an air vent situated directly above the mirror until background airflow was detected after the schlieren system was set up; the flow of air was so minute that it was not noticeable beforehand, but the schlieren system was sensitive enough to detect it. The researcher attempted to stop the flow of air by sealing the vent with duct tape; however, the vent began to create a high-pitched noise as air escaped from a small crack in the duct. Cardboard was again used as a funnel to redirect the flow of air away from the observation area. Absence of background airflow was established using a cup of hot water placed in front of the mirror; the water's rising heat was then viewed using schlieren imaging and verified as undisturbed.

2.4 Rating Task

2.4.1 Raters

A total of 10 raters were recruited from the local speech-language pathology program. The inclusion criteria were that participants were either current students or recent graduates from the speech-language pathology program with normal hearing and vision. Research suggests that

entry-level clinicians' ratings of nasality may benefit from additional means of nasality assessment, such as nasometry, whereas experienced clinicians show no advantage beyond auditory perceptual ratings (Brunnegard, Lohmander & van Doorn, 2012). Following this reasoning, students and recent graduates may not yet have reached the stage of expertise where auditory perception alone is adequate for high accuracy in rating nasality, thereby decreasing the possibility of ceiling effects for Audio-only versus Combined Auditory-Visual modalities. Conversely, the phonetics background of speech-language pathology students and recent graduates could also reduce possible floor effects that may occur due to lack of knowledge of nasal/oral contrasts. A prerequisite for entry into the program is an upper-level phonetics course with transcription; therefore students have received some degree of professional training in auditory perceptual assessment and were unlikely to demonstrate floor effects.

2.4.2 Rating Task Procedures

There were three rating conditions: Video-only, Audio-only and Combined (audio-visual). Each was segmented into three blocks, with all speakers' data being presented once per block (see below for details on the set of stimuli). All stimuli during the rating task were presented in random order using a custom-made MatLab script, with a brief washout period of six seconds between ratings. Raters were given the opportunity to take a break at the end of each block, which took roughly seven minutes to complete. All raters began with the Video-only condition. The order of presentation for the Audio-only and Combined conditions was counterbalanced, with five raters randomly assigned to rate Audio-only first, and five randomly assigned to rate Combined first. Counterbalancing was necessary to control for learning and practice effects, thereby enabling a valid comparison between the two modalities for hypothesis

2 (learning effect). Conversely, the Video-only condition was always presented first so hypothesis 1 (visual indications of nasality) could be tested without worrying about the influence of ordering effects. Raters were required to wear headphones for the Combined and Audio-only tasks, and sound volume was adjusted to a comfortable level.

For each condition, only one recording of each word was selected, the recording with the least amount of visible airflow preceding the initiation of each word. When a speaker breathes out immediately before speaking, non-speech airflow masks the airflow of the subsequent vocalization. Only recordings of monosyllabic words were included in the rating task in order to maximize the level of visual similarity between minimal pairs. For example, the similarity of facial movements between the minimal pair *Pas* and *Paon* were likely to be less noticeable than the facial movements between *Pas* and *Papa*; in this way, disyllabic words were eliminated in order to decrease the likelihood that ratings of the visual recordings would be based on visible characteristics of speech production that were unrelated to nasal airflow. One participant's recordings for minimal pair *Paix/Paon* were excluded from the rating task due to a disyllabic articulation of *Paon*. Finally, only the list of minimal pairs starting with /p/ was chosen to be included in the rating task in order to keep completion time within reasonable limits (approximately 90 minutes). The final number of tokens rated by each person was 414: 46 words per block (46 words/block x 3 blocks x 3 conditions).

Raters began their task with a training session. A PowerPoint presentation provided a simple description of the physiology behind velopharyngeal opening and nasal airflow, and a brief description of the acoustic properties of nasality. Raters were able to see and hear schlieren recordings of *et* and *en* taken from speaker 3's practice words (this speaker's data were chosen for the instruction section due to clarity of airflow and audio recording of nasality). Instructions

were given to raters to pay attention to airflow at the beginning of words, which was deemed necessary due to their complete lack of experience. Raters then performed two practice sessions; 10 practice ratings of the Video-only stimuli showing the words *et* and *en* taken randomly from all speakers, and 10 practice ratings of the Audio-only versions. The Video-only practice session was considered essential due to the fact that none of the raters had any previous experience rating schlieren images, and the Audio-only practice session was included to avoid bias in exposure time between modalities.

2.5 Data Analysis

The expected airflow patterns of nasal and non-nasal stimuli can be interpreted in terms of signal and noise. Non-nasal sounds are expected to exhibit only oral airflow, whereas nasal sounds are expected to exhibit both oral and nasal airflow (illustrated in Figure 1). In this way, oral airflow can be considered “background noise” because it mostly held constant in both conditions, while nasal airflow is the “active signal” that helps to differentiate between the two sounds. This experiment involves perceptual differentiation between a “noise” condition and a “noise + signal” condition, and is therefore amenable to analysis using Signal Detection Theory (SDT) with a Yes-No design. This method was originally used by Tanner and Swets (1954), and involves presenting stimuli from both the signal and noise conditions. The following summary is based on Stanislaw and Todorov (1999) who provide a detailed explanation of SDT. Raters respond “yes” when they perceive a signal and “no” when they perceive noise. It is expected that the rater will occasionally respond to these stimuli incorrectly, resulting in four possible outcomes: hit, miss, correct rejection and false alarm. These data enable the calculation of two separate distributions; the distribution of signal condition responses (hits and misses) and the

distribution of noise condition responses (correct rejections and false alarms). The mean difference between these two distributions is measured in standard deviation units, and the resulting value is called d-prime. Lower d-prime values indicate overlap of both conditions' distributions with a d-prime value of zero suggesting that raters were unable to differentiate between signal and noise. Likewise, higher d-prime values indicate better discrimination between the two conditions.

An advantage of d-prime is that it enables calculation of the rater's response bias, which is essentially a calculation of sensitivity and specificity. SDT theorizes that raters have individual perceptual thresholds, a benchmark around which their system decides whether a stimulus is a signal or simply noise. Negative scores indicate a *liberal* bias, where a rater tends to respond more often with "yes," leading to a relatively larger number of false alarms, whereas positive scores indicates a *conservative* bias, where a rater tends to respond "no," with relatively more misses. It is therefore possible for two raters to have exactly the same d-prime values, but different response bias values. Such an occurrence simply indicates that the difference between their signal and noise distributions was the same, although one rater has a relatively higher and conservative perceptual threshold, resulting in more misses, while the other's threshold is lower and more liberal, resulting in more false alarms. Calculation of response bias is not necessary to determine if two distributions are significantly different and therefore a prediction of response bias is not included in this study's primary hypothesis (presence/absence of nasal airflow). Nonetheless, it may provide further insight into rating differences and will be included in discussion.

One disadvantage is that calculations of d-prime may require corrections due to "perfect performance"; the formula for d-prime, which relies on calculation of z-scores, is unable to

handle response rates of 1 or 0. Stanislaw and Todorov (1999) suggest methods for working around this problem. The study's MatLab script, created by one of the study's committee members, automatically calculated d-prime and response biases for each participant, condition, and trial. Correction of extreme rates was only necessary in the Audio-only and Combined conditions because the Video-only condition had no perfect response rates.

All three of this study's hypotheses predict a significant difference between specified group means; therefore the use of a single-sample t-test and paired-sample t-tests are appropriate because they enable statistical comparisons between means. For the first hypothesis (visual indications of nasality), a single-sample t-test with a test value of 0 will permit calculations of whether the observed average d-prime value is significantly greater than chance, in other words a d-prime of 0. Hypothesis 2 (learning effects) will be tested by comparing mean d-primes between the three blocks of Video-only stimuli in paired-sample t-tests. Finally, a paired-sample t-test will appropriately compare averages between conditions for hypothesis 3 (advantage of Combined stimuli over Audio-only stimuli). In this way, d-prime is only the first step in analyzing the data; the true importance of this study's results and implications will be drawn from significant differences between group means as calculated by t-tests.

2.5.1 Operational Definitions

Visual indications of nasality were defined as the simultaneous presence of airflow from both the nasal and oral cavities near the beginning of a word. Such a constrained definition was necessary due to the task being completely novel for all raters, and due to the pervasive interference of background airflow caused by breathing, particularly from exhalation at the end

of words. Likewise, visual indications of non-nasality were defined as the presence of airflow from only the oral cavity near the beginning of a word.

Chapter 3: Results

3.1 Hypothesis 1: Visual Indications of Nasality

First, it was predicted that accuracy in discriminating between nasal/non-nasal sounds would be greater than chance performance, basely solely on visual indications of airflow. To test this, d-prime values from Video-only blocks one, two and three were averaged for each rater (Table 2). The resulting means were analyzed using a one-sample t-test with a test value of 0, which was the expected d-prime value if responses were random. On average, d-primers ($M = 1.44$, $SE = .10$) were greater than 0. This difference is significant $t(9) = 14.61$, $p < .05$, 95% CI [1.22, 1.67]. Therefore the null hypothesis that d-primers would not be not greater than chance response (i.e. d-prime = 0) is rejected. Furthermore, p (.000) was smaller than the Bonferroni adjusted alpha level of .01, suggesting that this result was unlikely a chance occurrence despite multiple comparisons.

Table 2 Values for d-prime in “Video-only” conditions

Rater	Blocks			Mean	SD
	1	2	3		
R1	1.29	1.42	2.06	1.59	0.41
R2	1.29	1.52	1.17	1.33	0.17
R3	1.75	1.10	1.64	1.50	0.35
R4	1.29	1.91	1.18	1.46	0.39
R5	2.35	1.45	1.45	1.75	0.52
R6	1.42	1.52	1.75	1.56	0.17
R7	1.29	2.10	1.66	1.68	0.41
R8	1.31	1.29	1.02	1.21	0.16
R9	1.52	2.06	1.42	1.67	0.35
R10	0.43	0.59	1.06	0.69	0.33
Mean	1.39	1.50	1.44	<u>Tot mean</u>	<u>SD means</u>
SD	0.48	0.46	0.34	1.44	0.31

3.2 Hypothesis 2: Learning Effects

Second, due to raters' lack of experience with schlieren imaging, it was hypothesized that a learning effect would occur between all three blocks of the Video-only condition. Three paired-sample t-tests were performed (Table 3). On average, d-primes were higher in Video Block 3 ($M = 1.44$, $SE = .11$) than in Video Block 1 ($M = 1.40$, $SE = .15$), although not significantly $t(9) = .31$, $p > .05$, 95% CI $[-.30, .40]$. The average d-prime of Video Block 2 was also higher than that of Video Block 1, but this difference was also not significant $t(9) = .61$, $p > .05$, 95% CI $[-.28, .48]$. Demonstrating an opposite direction than would be expected for a learning effect, d-primes were lower in Video Block 3 than in Video Block 2 ($M = 1.50$, $SE = .14$), but this difference was not significant $t(9) = -.34$, $p > .05$, 95% CI $[-.41, .30]$. Due to non-significant results in all three pairs, the research hypothesis was not supported. All p values were greater than the Bonferroni adjusted alpha level of .01, supporting the failure to reject the null hypothesis.

Table 3 Paired sample t-tests between “Video-only” blocks 1, 2 and 3

Pair	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	p (1-tailed)
				Lower	Upper			
1 V3 - V1	.05	.49	.15	-.30	.40	.31	9	.383
2 V3 - V2	-.05	.50	.16	-.41	.30	-.34	9	.370
3 V2 - V1	.10	.53	.17	-.28	.48	.61	9	.280

3.3 Hypothesis 3: Advantage of Combined Stimuli over Audio-only Stimuli

Third, it was predicted that ratings of the Audio-Visual Combined condition would be greater than the Audio-only condition. Data from blocks one, two and three in both respective conditions were averaged by rater (Table 4). On average, d-primes were greater for Combined ($M = 3.46$, $SE = 0.15$) than for Audio-only ($M = 3.36$, $SE = 0.21$); however, this difference was not significant $t(9) = 0.91$, $p > .05$, 95% CI [-.16, .36] (Table 5). Therefore, the research hypothesis is rejected. Lastly, p (.19) is greater than the Bonferroni adjusted alpha level of .01, again supporting rejection of the research hypothesis.

Table 4 Mean d-primes for blocks 1, 2 and 3 by rater and condition

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	Total Mean
“Combined”	2.76	3.52	3.20	3.87	3.97	3.79	2.67	3.45	3.43	3.97	3.46
“Audio-only”	2.64	3.97	2.90	3.97	3.55	3.79	2.23	3.70	2.70	4.14	3.36

Table 5 Paired t-test between “Combined” and “Audio-only” conditions

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	p (1-tailed)
				Lower	Upper			
C-A	.104	.36	.11	-.16	.36	.91	9	.19

Chapter 4: Discussion

This proof-of-concept study sought to demonstrate schlieren imaging's potential as a speech research tool by testing whether visual correlates of nasality are observable using schlieren imaging. First, it was predicted that raters would accurately judge nasal and non-nasal words based solely on visual indications of nasal airflow. If nasality has no visual correlate, as captured by schlieren imaging, then it would be very unlikely that raters would perform at greater than chance levels because they would have no information on which to base their ratings. In contrast, performance above chance would indicate that nasality is detectable in schlieren imaging of airflow, at least to some degree. Second, it was predicted that there would be a learning effect in the Video-only modality. Raters were completely inexperienced with schlieren imaging, and therefore likely to show improved scores as they became more familiar with the tool. Third, it was predicted that raters would respond more accurately when stimuli included both video and audio information compared with audio information alone. Such an observation would help demonstrate schlieren's usefulness in facilitating perceptual ratings of nasality above and beyond traditional auditory-only methods. These three predictions will be discussed in detail below.

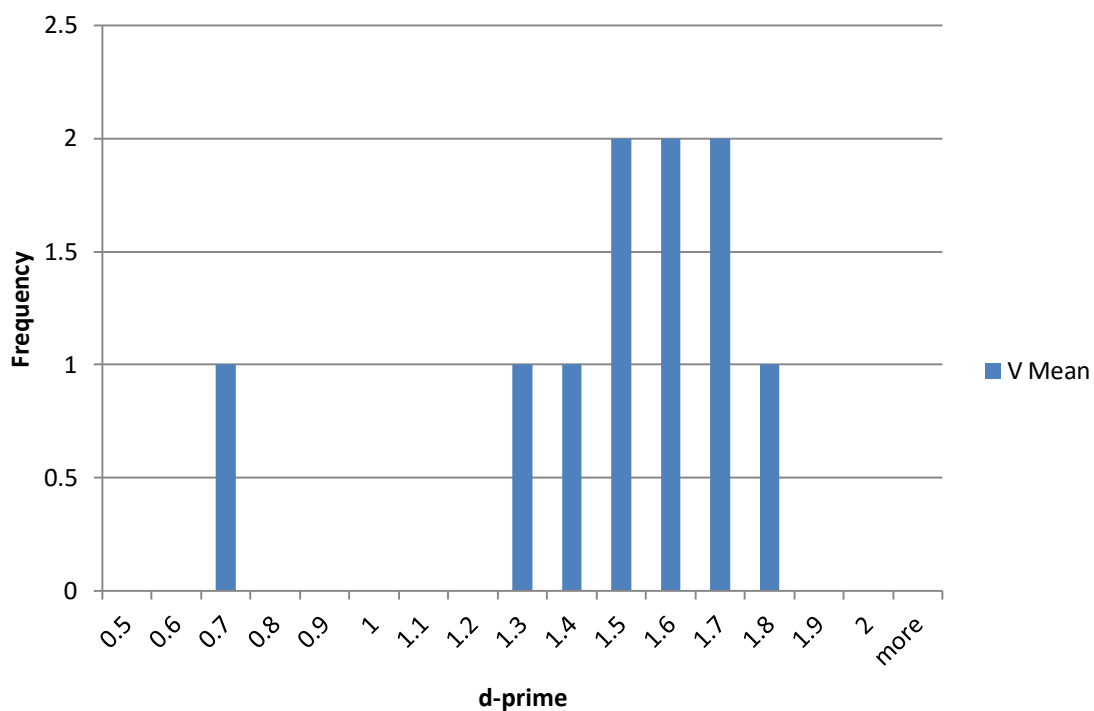
4.1 Visual Indications of Nasality

Findings from the study strongly indicate that nasality is detectable in schlieren imaging of airflow of speech. Accuracy in binary ratings of visual stimuli for nasal and non-nasal words were significantly above chance, suggesting that raters visually perceived some degree of difference between the patterns of airflow for nasal and non-nasal sounds. Although the analysis cannot confirm exactly what this pattern of airflow was, previous research shows that nasality is

associated with nasal airflow. Similarly, raters were prompted to watch for nasal versus oral airflow, and therefore it is likely that the predicted pattern of oral-only airflow for non-nasal sounds and oral + nasal airflow for nasal sounds is more or less accurate. This proof-of-concept exploration is one of the first studies to empirically demonstrate that schlieren imaging has potential to be used as a speech research tool.

Despite apparent significance, there is reason to be cautiously conservative in the interpretation of this test's statistical results. A single-sample t-test was used to test the first hypothesis; however, slight divergence from the t-test's underlying assumptions may render the statistic inaccurate. Among other assumptions, t-tests require raw data to be approximately normally distributed, and contain no outliers. Data from averaged V1, V2 and V3 appeared to be approximately normally distributed (Figure 3) except for one outlier which caused the data to be negatively skewed.

Figure 3 Distribution of each rater's mean d-prime scores for Video-only condition

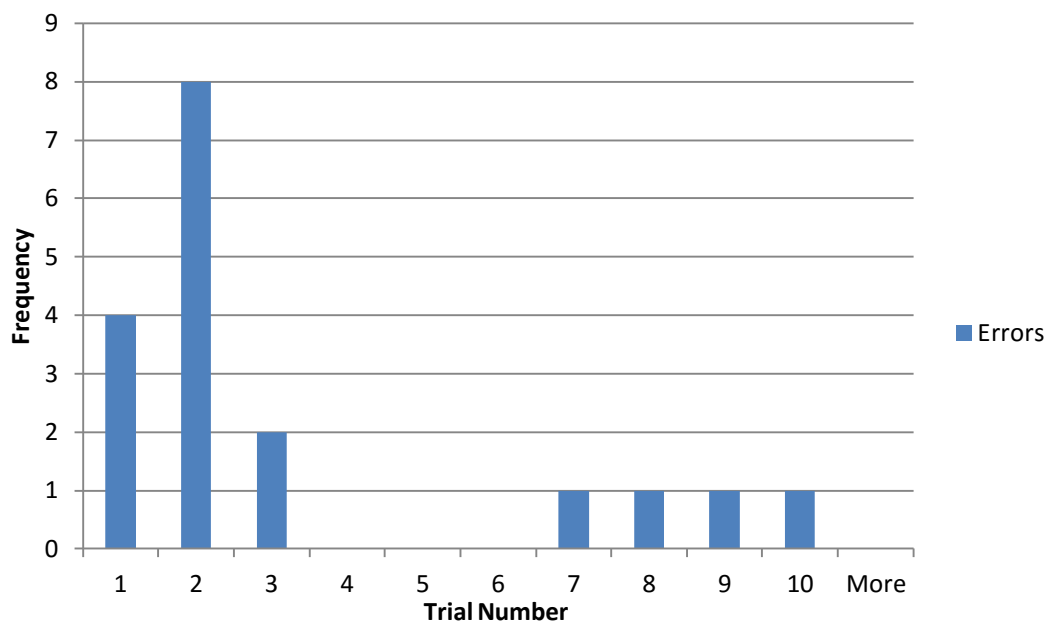


This outlier has the effect of decreasing the overall mean ($M = 1.44$ with the outlier included versus $M = 1.53$ with the outlier removed), which increases the risk of a Type II error as the sample value's mean draws closer to the test value of 0. Similarly, concentration of the scores around the mean suggests that the data's distribution has high kurtosis which serves to further decrease the t-test's statistical power compared with analogous nonparametric tests (Reineke, Bagget, & Elfessi, 2003). Taken together, these two drawbacks would likely render the single-sample t-test to be non-significant as they both contribute to the increased likelihood of a type II error. In spite of this, the t-value for hypothesis 1 was nonetheless very high ($t = 14.61, p = .000$), keeping in mind that this likely *underestimates* the true t-value. Realistically, one would be justified in rejecting the null hypothesis simply by looking at the raw d-prime scores; the lowest score from all blocks of all raters (constituting 30 repetitions) was d-prime = 0.43, with virtually all other d-prime values being greater than 1.00. This certainly gives the strong appearance of greater-than-chance performance, and suggests that the reported t-value is sufficient at face value. It must be kept in mind that the statistic has nothing to say about the *degree* of nasal airflow, but simply whether raters were successful in using nasal airflow to differentiate between stimuli. Considering the simplicity of the task for hypothesis 1, the strength of the reported statistic, and the face-value agreement of the statistic with raw data, rejection of the finding simply because it could potentially be even stronger would serve no useful purpose. The data show that raters scored above chance.

4.2 Hypothesis 2: Learning Effects

Tests for hypotheses 2 failed to demonstrate a learning effect between Video-only blocks, although a non-significant improvement between the first and third blocks was observed. One possibility for the lack of significant improvement is the relatively short period of training time in which raters were exposed to Video-only stimuli. The practice session consisted of only 10 stimuli presentations, and raters were notified after each trial if they scored the video correctly or incorrectly. Feedback was not provided during the main experiment. It was difficult to anticipate the necessary degree of training, and the practice session was designed to find a balance between providing raters *some* indication of what to look for (which was essential considering raters' lack of experience with schlieren imaging), and biasing their responses by essentially 'giving them the answer.' The question remains whether scores would reach ceiling with more practice. Figure 4 demonstrates that the majority of errors occurred in the first three turns of the practice session, suggesting that feedback was effective in improving accuracy.

Figure 4 Sum of incorrect responses by trial order in the Video-only practice session.



It is unlikely that this pattern was simply the result of practice effects; otherwise one would have expected similar improvements over the course of the main experiment. The data for hypothesis 2 suggest that this was not observed. It is more reasonable to assume that the feedback itself resulted in improved scores in the practice session. Schlieren imaging is a novel technique for speech research, and these findings suggest that detailed training and sufficient experience in visual perceptual rating may need to be provided for future users.

4.3 Hypothesis 3: Advantage of Combined Stimuli over Audio-only Stimuli

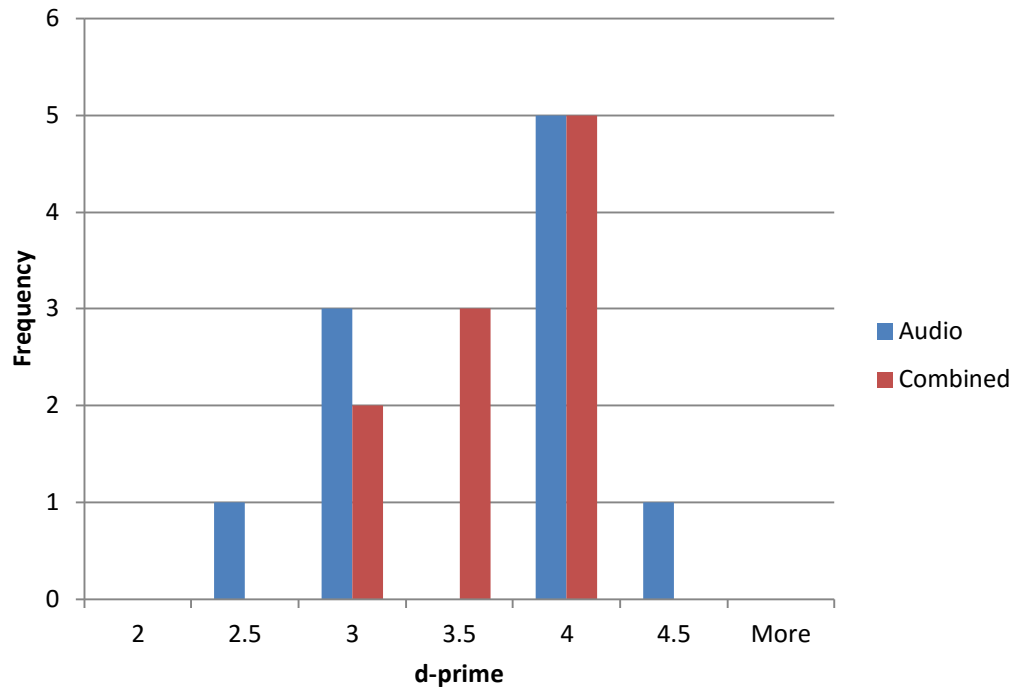
Hypothesis 3 stated that the Combined Audio-Visual condition would demonstrate higher d-prime values than Audio-only, thus indicating that schlieren imaging could play a facilitative role in the perceptual evaluation of nasality. A paired t-test between the averaged d-primes for the Combined and Audio-only conditions respectively revealed a slight advantage for Combined; however, this difference was not significant. Although this result does not provide evidence for schlieren's facilitative role in perceptual evaluation, neither does it disprove its potential. Further examination of the available data seems to indicate promising signs for this application of schlieren imaging, demonstrating that additional and improved research is warranted.

4.3.1 Ceiling Effects

A close look at the data reveals negatively skewed distributions of scores in both conditions (Figure 5), which may be suggestive of ceiling effects. Indeed, the formula used in calculations of d-prime resulted in the highest possible value being 4.14, which coincides with perfect response (i.e. 23 hits, 0 false alarms). The d-prime values from the entire dataset of all

blocks for all raters reveal a mode of 4.14 in both the Combined and Audio-only conditions respectively (a full data table is available in Appendix A).

Figure 5 Ceiling effects for both Audio and Combined modalities as indicated by negative skew.



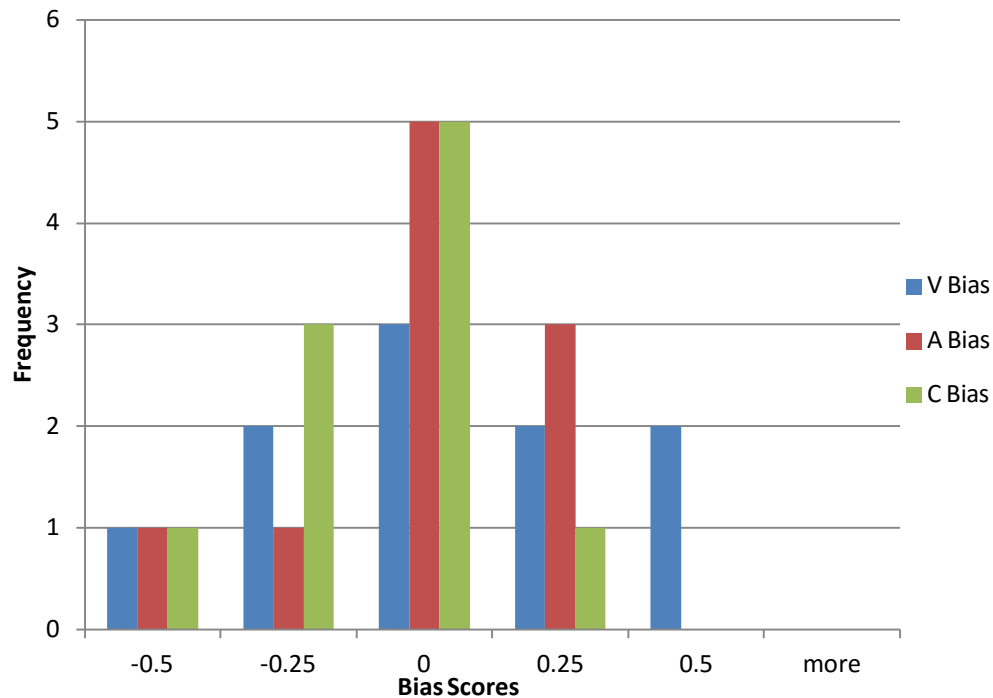
As is the case with ceiling effects, these data suggest that raters' true discrimination ability may actually be greater than the reported d-primes, which were artificially restrained by the fact that it was impossible for raters to perform any better. Evidently, binary auditory rating of typically produced nasal/non-nasal sounds is likely too easy a task for entry-level SLPs. In previous research, inexperienced clinicians benefitted from using nasometry to identify and rate hypernasal and hyponasal speech (Brunnegard et al., 2012). Discrimination between these two resonance disorders is quite a different task than that which was tested in the current study, where raters only had to decide whether a stimulus was nasal or not nasal. After completion of the rating task, one of the raters even commented that they hardly focused on the visual stimuli at

all, relying mostly on auditory stimuli during the Combined condition. Again, this is not surprising considering all raters' relatively high level of experience with auditory perception, and virtually nonexistent experience with schlieren visual perception. A discrimination task between normal and disordered resonance or discrimination between hypernasal and hyponasal resonance disorders may prove to be more appropriately challenging, not to mention more clinically insightful.

4.3.2 Sensitivity and Specificity

An alternate way to investigate whether multi-modal stimuli provide an advantage over auditory alone is in the dimension of *sensitivity* and *specificity* which can be observed using d-prime's complementary statistic, response bias. As previously mentioned, a negative (liberal) response bias indicates a higher rate of false positives (referred to as 'false alarms' in SDT) whereas a positive (conservative) response bias indicates a relatively higher rate of false negatives ('misses' in SDT). In this way, a liberal response bias has relatively high sensitivity at the cost of low specificity, and a conservative bias has low sensitivity with high specificity. Although no predictions were made concerning the degree or direction of response bias, scores may shed more light on the differences between ratings of each condition. Figure 6 illustrates the distribution of mean bias scores, averaged from blocks 1, 2 and 3 for each condition.

Figure 6 Histogram of response bias means for each condition.



Although the data look approximately equal between conditions, three post-hoc 2-tailed single-sample t-tests reveal subtle differences; Video-only bias ($M = -.09$, $SE = .10$) was not significantly different than the response bias for a hypothetical population mean of 0 ($p = .44$), although Audio-only bias ($M = -.15$, $SE = .07$) approached significance ($p = .07$). Conversely, Combined bias ($M = -.24$, $SE = .08$) was significantly different than chance response ($p = .01$), although a Bonferroni correction of $.01(.5/3)$ suggests that bias scores for Combined are close to non-significant. Although differences appear to be minute, the data are suggestive of a pattern of differing degrees of sensitivity and specificity between conditions. Indeed, although the Video-only modality received the lowest d-prime scores, a response bias close to 0 indicated a balance between false positive and false negative rates. Conversely, the Audio-only and Combined

conditions demonstrated more liberal biases, indicating relatively higher false alarm rates and poorer specificity compared with Video-only's data.

Although not significant, this pattern may suggest that ratings based solely on aerodynamic characteristics, as was the case in the Video-only condition, may result in a stronger balance between rates of false positives and false negatives. Indeed, when it comes to velopharyngeal function, "aerodynamic parameters may be considered as 'perception-neutral', and therefore constitute a valuable insight into the production of nasal vs. oral speech sounds" (Delvaux, Demolin, Harmegnies, & Soquet, 2002, p. 579). A lower rate of false positives were observed in the Video-only condition, suggesting that perceptual ratings based on airflow alone may be more "neutral" than auditory-perceptual ratings. Therefore visual ratings may have some advantage in terms of specificity over acoustic ratings when it comes to indications of velopharyngeal status. It would be worthwhile to test if this apparent advantage holds for more experienced raters.

As previously mentioned, assumption violations for t-tests may render these calculations less valid; therefore it is important to view these statistics in context of the raw data. Indeed, the miss-to-false alarm ratio for Video-only is 161-to-191, but only 17-to-56 for Audio-only and 8-to-53 for Combined. Even though Video-only's ratio is technically more even, it has impressively more mistakes than the other conditions, which of course is reflected in the difference of d-prime values. In light of this, Video-only's relatively better bias score may simply be the result of many more instances of blind guessing and ultimately may not suggest an advantage in true positive or true negative response rates. Even if there were an advantage, the Combined condition's bias score suggests that schlieren stimuli either detract or have no effect on multi-modal evaluations. Again, it is reasonable to think that these negative results are largely

related to the raters' lack of experience with schlieren imaging. Furthermore, these suggestions are speculative at best because they are based on observations drawn mostly from non-significant results. On the other hand, positive findings from hypothesis 1 indicate that raters are able to detect nasality using schlieren imaging, and therefore it is expected that further experimentation using tasks of more appropriate difficulty to eliminate ceiling effects, in addition to considerably more training to raters, will reveal the hypothesized advantage. Further consideration of the experimentation below will suggest other research approaches for the future use of schlieren imaging.

4.4 Speaker Characteristics

Qualitative observations of speaker differences can be made by combining each speaker's hit and correct rejection rates. The resulting percentage is an indication of how successful raters were overall in accurately discriminating between a speaker's nasal and non-nasal stimuli. As would be expected, some speakers received more accurate ratings than others. For example, in the Video-only condition, speaker 5's data appeared to be rated the least accurately (62% correct), whereas speaker 7 received the most accurate ratings (89% correct). However, speaker 7's dataset is somewhat smaller due to the exclusion of *Pas/Paon*; therefore a better comparison may be with speaker 4 (86% correct). Qualitative differences do not appear to be related to age or gender because speaker 5 is within the same age range as speaker 7, and the same gender as speaker 4. Unfortunately, rater 4 did not identify a specific dialect so it is uncertain whether this speaker's ratings were greater due to this difference. However, speaker 7's dialect was self-identified as "Vosgien" whereas speaker 5 identified as Quebecois, indicating that dialect might play a role in aerodynamic differences. Carignan (2014) explains that finding 'accurate' minimal

pairs for nasal and non-nasal vowels is difficult in French because previous research has used different vowel pairings for different dialects. It may be the case that, due to a difference in vowel use, speaker 7 produced the nasal/non-nasal minimal pairs in a way that created a clearer contrast. Of course, patterns cannot be based on one participant alone, and further research should investigate potential dialect differences in nasal airflow.

One non-linguistic characteristic that might explain differences in ratings may simply be the degree of background airflow (indeed, this is a strong limitation of schlieren imaging in general, and will be elaborated on more in depth in the limitation section). A qualitative observation of recordings suggests that there was noticeably more background airflow for speaker 5 compared with speaker 4 and 7. All sources of background airflow were eliminated and verified to be absent prior to each recording; therefore the only likely remaining source of background airflow is simply body heat. Unfortunately this study's methodology provides no way of quantifying airflow, including background airflow, and therefore this observation is largely made as a warning for future research. On the other hand, this background airflow likely did not affect overall ratings, and seemed to only be present for some individual speakers. Signal Detection Theory works under the premise that background 'noise' will always be present, and raters must try to detect signals within this noise. There is no indication that body heat fluctuated systematically between nasal and non-nasal sounds, and therefore this 'noise' was held constant, at least at the level of individual speakers.

4.5 Rater Characteristics

Although d-prime values for all raters were greater than chance response, variance in scores demonstrated that raters did not perform identically. An interesting case is rater 10 whose

d-prime values presented as outliers in blocks 1 (d-prime = 0.43) and 2 (d-prime = 0.59) in the Video-only condition. This rater showed improved scores between each block, but on average scored well below the mean d-prime for all raters (d-prime = 0.69 for rater 10 versus 1.44 for all raters). What makes this case particularly interesting is rater 10's d-prime scores for the Audio-only condition were the highest of all raters (d-prime = 4.14 for rater 10 versus 3.36 for all raters), and tied for highest in for the Combined condition (d-prime = 3.97). This rater provided feedback after the experiment, stating that the auditory ratings were much easier than the visual ratings, so much so that it was possible to rely solely on auditory stimuli during the Combined condition, with no need to pay attention to visual stimuli. Although other raters expressed that the auditory stimuli were easier to perceptually discriminate compared with the visual stimuli, they generally reported that they attempted to make use of both stimuli in the Combined modality. Perhaps a pattern that can be drawn from rater 10's data is that raters will tend to rely on their stronger suit when it comes to multi-modal perceptual ratings, especially if one modality is particularly stronger than the other.

In addition to raters' professional training in speech-language pathology, perceptual ratings of nasal/non-nasal contrasts might have been facilitated depending on a rater's proficiency in languages that use nasality contrastively. For such cases, one might expect relatively higher scores for Audio-only discrimination, and potentially higher scores for Video-only. Rater 5, who was proficient in French, presented as an outlier in Video-only block 1 with a d-prime of 2.35, well above the average d-prime of 1.39 for block 1. Rater 5's average d-prime across the condition was also the highest compared with the other raters. Although rater 5 spoke the stimulus language French, it is unlikely that lexical information or familiarity with words provided an advantage in the Video-only condition because stimuli were presented in the visual

modality alone with no audio and the rater did not yet know which words were being spoken. It is therefore possible that experience with auditory-perceptual discrimination between nasal and non-nasal sounds transfers to discrimination of these sounds with visual perception. Further experimentation with a variety of languages that use nasal vowels contrastively is needed to explore this hypothesis.

Interestingly, rater 5's average d-prime for the Audio-only condition was 3.55, only slightly above the group average of 3.36. Furthermore, despite the rater's high level of proficiency in French, a d-prime of 4.14 (indicating no errors in responses) was not observed. This may call into question the construct validity of the study's stimuli. Indeed, it is logical that a French speaker would be able to discriminate perfectly between French nasal and non-nasal words, indicating that some of the study's pre-defined nasal and non-nasal words could have been erroneous. However, rater 10 (whose d-primes were the lowest in the Video-only condition) demonstrated an average d-prime of 4.14 in the Audio-only condition, indicating that this rater made no errors. This suggests that construct validity may be intact; otherwise one would expect no perfect response rates. An alternate explanation to rater 5's data may simply be that perceptual evaluation of typical nasality is difficult, even for individuals who use nasal contrasts on a daily basis. Therefore, 'imperfect' response rates may coincide with typical nasality variation between speakers. This leaves room for schlieren imaging as a complementary tool in perceptual evaluation of nasality; perhaps where auditory-perception fails, visual-perception of aerodynamic events can pick up the slack.

4.6 Limitations of the Study and Directions for Future Research

4.6.1 Confounds in Visual Stimuli

A possible confound for visual rating of French vowels is lip rounding. Carignan (2014) found that the nasal vowel /*ɑ̃*/ appeared to be articulated with more lip protrusion/rounding compared to its oral counterpart /*a*/, whereas differences between /*ɔ̃*/ and /*o*/ were variable between speakers, and no consistent differences in labial articulation were observed between /*ẽ*/ and /*ɛ*/. If lip rounding provided an additional cue for discrimination, one would expect a higher level of accuracy in ratings. To test this, a percentage of average correct responses according to each vowel pair were calculated based on each rater's number of hits and correct rejections. The resulting scores were *Pas/Paon* = 76%, *Pot/Pont* = 74%, and *Paix/Pain* = 74%. 95% confidence intervals were respectively 6%, 7% and 6%. Correct responses for the word pair *Pas/Paon* (vowels /*ɑ̃*/, /*a*/) are minimally higher than the other pairs; however, overlapping confidence intervals suggests this difference is not significant. Paired t-tests revealed differences were indeed non-significant ($p > .05$) for all three combinations. Lip rounding, therefore, likely did not act as an extra clue to help discriminate between sounds.

A second possible confound is the airflow preceding phonation, primarily in the form of breath exhalation from the nose. This might have acted as a visual confound in the Video-only condition because raters likely lacked the necessary experience to distinguish this form of airflow from the onset of nasal airflow associated with nasal sounds. Although stimuli were carefully selected in order to minimize the amount of preceding airflow, this was impossible in some cases. For instance, speaker 8 exhaled air immediately prior to producing the word *Pot* in all three recording attempts, and consequently a “clean” recording could not be selected. Raters misidentified this word as nasal on 26/30 opportunities. This, of course, is a red flag; if airflow

associated with breathing can cause raters to perceive a video as “nasal”, then how can one be sure this confound was not the basis for all ratings of “nasal”? The answer lies within the response bias score. Such pervasive confounding airflow would lead to a greater tendency towards false positive ratings, and would present as a negative (liberal) response bias. As previously discussed, Video-only’s average response bias was -.09, and not significantly different than 0 ($p > .05$). Although the presence of a slight negative response bias suggests that breath airflow played some role in influencing raters’ perceptions, this effect was statistically non-significant and indicates that breath airflow was not a significant confounding variable.

With these limitations in mind, future research may move beyond the use of nasal/non-nasal minimal pairs. Indeed, it appears that research may be largely constrained to the use of single syllable utterances, and therefore use of words beyond minimal pairs alone would greatly expand the size of possible datasets. Similarly, nasality could be investigated in languages other than French. For example, nasal airflow measures in English could be compared to nasalance scores as measured by nasometry. This would provide insight on how non-contrastive nasality typically functions in English, and may further assist in laying down the essential groundwork for schlieren imaging to be applied to clinical nasality measures. Of course, schlieren imaging only *appears* to be limited to single syllable utterances; further research should investigate the use of carrier phrases, as it may still be possible that nasality is distinguishable amid masking airflow. Such findings would undoubtedly strengthen schlieren imaging’s potential, and therefore a deeper investigation with more sophisticated measures is highly encouraged.

4.6.2 Challenges of Airflow

As previously mentioned, schlieren imaging entailed several unanticipated challenges with background airflow. First, controlling body heat became an apparent necessity as heat emanating from the chest and legs often intermingled with speech airflow. The solution was to simply cover the participants, thus either containing or re-directing body heat from the observation area. This obviously detracts from the aforementioned advantage of schlieren imaging being unobtrusive; indeed, wearing a winter jacket to hold in body heat is probably more uncomfortable over extended periods of time compared with attachment of microphones to one's nose. However, as I gained experience with schlieren imaging, I found that the use of jackets in the end was not necessary because re-directing body heat with a piece of cardboard seemed to solve the issue sufficiently well. Alternately, a remarkably easy solution may simply be to have participants stand rather than sit. In doing so, participants' legs would no longer be positioned directly under the area relevant to speech airflow. Due to difficulties adjusting the mirror to a sufficient height for multiple participants, this option was unfortunately never explored. Similarly, a stationary barrier to re-direct heat (like a piece of cardboard) could be permanently positioned beneath the mirror.

Second, ambient airflow and temperature must also be controlled. Even the best attempts to find a room with minimal background airflow were foiled due to a small air vent. Indeed, pilot recordings of schlieren imaging gave the researcher a better appreciation of the subtle and invisible aerodynamic events surrounding us at all times; nonetheless these aerodynamic events had to be corrected before conducting the study. With the air vent controlled, a medium-sized storage room provided decent protection from airflow fluctuations and maintained a relatively cool ambient temperature. Warm temperatures decrease the difference between the refractive

indices of ambient air and breath and should be avoided in order to maintain an adequate schlieren effect. Of course, as the strength of the schlieren effect increases, so too does visualization of background airflow. It is therefore preferable to find a balance, and typical room temperature is likely adequate for the purpose of speech measurement.

Third, the shape and size of a speaker's face may affect how clearly airflow from the nose and mouth can be differentiated. A shorter distance between the nostrils and mouth appeared to cause the streams of nasal and oral airflow to collide relatively earlier than noses and mouths positioned farther apart. Similarly, the angle at which the nostrils project airflow may influence how easily the nasal jet can be differentiated from the oral jet. Evidently, measuring physical differences in anatomy and their effect on the shape of airflow may prove to be an insightful endeavour. Additionally, humans also exhibit a "nasal cycle", where the nasal airways take turns alternating between congestion and decongestion (Hasegawa & Kern, 1977); however, total nasal resistance reportedly remains fairly constant. The nasal cycle likely plays a minimal role in nasal airflow variance because the total amount of airflow remains relatively the same. Nonetheless, this phenomenon should at least be taken into consideration as a potential confound in future research and could be investigated by positioning the schlieren equipment vertically around the participant rather than horizontally (with careful attention to stabilization of heavy equipment so it does not fall on the participant!).

Lastly, a major limitation of schlieren imaging is its relative inability to visualize airflow *within* a stream of air. The schlieren effect only takes place when there is a change in refraction of light; therefore a stream of perfectly laminar flow, for example, would only reveal the stream's leading edge, with no visible change within the stream itself. This indicates that visible aerodynamic events in speech may be best observed during the initial release of air, hence this

study's use of monosyllabic words in isolation. Carrier phrases may still be informative and worth using with schlieren imaging due to the likelihood of turbulent airflow, thus enabling observation of movement within a stream of air, at least to some degree. However, such flow characteristics are likely more suitably observed using high speed recording due to their rapid and likely minute changes, but this may serve no practical role in real-time evaluation of speech. In any case, further research using more sophisticated equipment is likely to continue shedding light on this topic.

4.6.3 Issues in Quantification

It is worth noting that quantification of flow using schlieren imaging is certainly possible, however its implementation requires expertise in computer-based modeling and fluid dynamics, and therefore went beyond the scope of this study based on human perception. Those interested and knowledgeable in flow analysis should refer to chapter 10 in Settles (2001), who provides an extensive overview of applicable techniques. Additionally, a mechanical engineer colleague suggested the use of optic flow analysis at the onset of our study. Optic flow essentially works by observing the overlap of pixels of an image between frames, thus inferring movement. This particular engineer wrote a custom-made optic flow script using JavaScript to analyze recordings from our study's initial trial; despite some indications of promise, this approach was ultimately abandoned due to perceptual ratings being more straightforward. Nonetheless, quantification of schlieren imaging is possible as a future endeavour.

4.6.4 Future Applications of Schlieren Imaging

A motivating factor for investigating nasality with schlieren imaging was to provide insight on schlieren's practical clinical applications. In addition to understanding the aerodynamic properties of typical nasality, much of the current research in the aerodynamics of nasality comes from the evaluation of velopharyngeal function in the cleft palate population. Dotevall, Lohmander-Agerskov, Ejnell and Bake (2002) provide an overview of this area of research; in reviewing 16 studies investigating the relationship between perceptual evaluation and aerodynamics associated with velopharyngeal function of the cleft palate population, the authors observed a moderate correlation between perceptual ratings of velopharyngeal competence, hypernasal resonance, and aerodynamics of speech sounds. A specific example is provided by Warren, Dalston, and Mayo (1994) who found moderate correlations between perceived hypernasality and velopharyngeal opening, nasal airflow and nasal airflow duration. These observations suggest that disordered nasality is likely represented in nasal airflow, at least to some degree.

Schlieren imaging may in fact be a particularly useful non-auditory measure to help discriminate between hyper- and hyponasality. As the prefixes 'hyper' and 'hypo' suggest, these two conditions may produce an exaggerated pattern of airflow in relation to the nasal/oral airflow pattern observed in the current study, resulting in a hypothetically clearer distinction in schlieren images. A worthwhile starting point would be to look at timing and duration of nasal airflow, which previous research suggests corresponds most strongly to hypernasality (Warren, Dalston, & Mayo, 1993). A comparison between resonance disorders was indeed considered at the inception of the current study, but it was felt to be more appropriate to establish proof-of-concept before venturing forth with clinical populations. Indeed, schlieren imaging is safe and

demonstrates potential as an aerodynamic measure; therefore it is strongly suggested that clinical research be pursued.

Schlieren may be compared with some of the current tools in speech-language pathology's clinical repertoire. First, nasometry is one of the key devices for assessment and treatment of resonance disorders and functions by acoustically measuring and comparing sound signals from the nose and mouth, and has been shown to provide an advantage over perceptual ratings of nasality, specifically for inexperienced clinicians (Brunnegard et al, 2012).

Unfortunately, its potential effectiveness as a biofeedback tool has limited evidence (Howard & Lohmande, 2011). Considering schlieren's similar ability to detect speech signals from the nose and mouth, albeit aerodynamic rather than acoustic, it would be worthwhile to compare and contrast these tools' functions and potentials. Schlieren imaging's "real life" depiction of nasal airflow may ultimately be more easily understood by patients compared with nasometry's graphical or numerical representation of nasality. Second, non-surgical interventions such as speech prostheses help to decrease the size of the velopharyngeal opening, thereby facilitating closure and reduction of hypernasality or nasal emission. Treatment with speech bulbs, a type of prosthesis, has demonstrated reduced hypernasality and nasal emission, especially when surgical interventions were not an option (Sell, Mars, & Worrell, 2006). It is possible that schlieren imaging could provide biofeedback to help facilitate improvements in velopharyngeal function alongside these tools. Similarly, schlieren imaging may also be used to assess the degree of velopharyngeal closure, as measured by airflow, following implementation of these interventions.

Speech visualization tools have been demonstrated to provide an additional advantage over traditional methods of perceptual judgment or feedback, particularly when applied to

clinical conditions, suggesting that schlieren imaging may yet hold potential as a facilitative tool. For instance, research showed that visual perception of speech through the use of spectrograms led to better inter-rater reliability for ratings of voice disorders (Martens, Versnel, & Dejonckere, 2007); similarly, spectrograms have been used as biofeedback to improve therapy outcomes (Byun & Hitchcock, 2012). Ultrasound has been used for a variety of speech disorders and has shown promise in speech rehabilitation (Bernhardt et al., 2008). Its basis as a biofeedback tool is to provide a visible image of otherwise unobservable phenomena, and has contributed to faster gains in speech therapy compared to traditional approaches (Adler-Bock, Bernhardt, Gick, & Bacsfalvi, 2007; Bacsfalvi, Bernhardt, & Gick, 2007; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003; Gick et al., 2008; Shawker & Sonies, 1985). Like these examples, schlieren imaging is a speech visualization tool, and it is therefore reasonable to predict that similarly positive outcomes in speech therapy may arise from its use.

Although already touched upon, one of the more enticing potential uses of the schlieren technique is biofeedback. If schlieren's utility as a research measure is ultimately demonstrated to be inferior to pre-existing technologies, it may yet prove to be an effective biofeedback device thanks to its ability to visualize real-life physical events rather than provide graphical or numerical depictions, as is the case in spectrometry or nasometry. Brunner et al. (2005) demonstrated the effectiveness of visual biofeedback for velopharyngeal dysfunction through visualization of the velopharyngeal port, which incidentally had the additional positive effect of increasing patients' self-perception of articulation. This coincides with the previously discussed effectiveness of ultrasound biofeedback. Biofeedback appears to strengthen outcomes of speech therapy, and it will be worthwhile to explore whether schlieren imaging can serve as an additional tool in the clinician's toolbox.

Chapter 5: Conclusion

The results of this study demonstrate that the acoustic property of nasality has an aerodynamic correlate that distinguishes it from non-nasal sounds, and this difference can be detected visually. Although these results cannot demonstrate exactly what this aerodynamic correlate looks like, its existence is highly probable because airflow was virtually the only visible difference between video-only recordings of “nasal” and “not nasal” sounds. It is reasonable to think, however, that nasality is aerodynamically represented by airflow leaving the nasal cavity. Similarly, confounding visual variables such as lip rounding were found to be non-significant, further suggesting that correct discrimination was based on airflow differences alone.

Despite this above-chance discrimination between visual stimuli, the observed relationship between nasality and its aerodynamic correlate may not be perfect, as indicated by relatively lower rating scores for visual-only stimuli than for auditory stimuli. It is uncertain whether this relationship reflects observations from past research that suggest aerodynamics are nonlinearly related to nasality, or if lower scores are the consequence of raters’ lack of experience with schlieren imaging. Indeed, ratings of visual stimuli did not significantly improve over time. Although raters still accurately judged stimuli at a well above chance response rate, this lack of improvement may in fact indicate a relatively low correlation between aerodynamics and nasality. Conversely, it may simply suggest that perceptual evaluation with schlieren imaging has a slow learning curve. This same dichotomy makes it difficult to interpret whether schlieren imaging coupled with auditory stimuli has an advantage over auditory stimuli alone. The former demonstrated slightly higher scores, but the results were not significant. Further research using raters who are more experienced with schlieren imaging is needed to determine if visual feedback indeed provides the predicted advantage in perceptual rating.

Overall, this study indicates that schlieren is a promising speech research tool. Its use in speech science has been suggested in the past, but this study appears to be the first to empirically investigate schlieren's usefulness as a research tool. Its potential utility in the assessment and treatment of resonance disorders is a similarly exciting prospect. Flow visualization of human participants has indeed only scratched the surface of the aerodynamics of speech, and it is hoped that this study's efforts to demonstrate proof-of-concept of schlieren imaging's capabilities will spur further exploration.

References

- Adler-Bock, M., Bernhardt, B. M., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of North American English /r/ in 2 adolescents. *American Journal of Speech-Language Pathology*, 16(2), 128-139.
- Bacsfalvi, P., Bernhardt, B. M., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *International Journal of Speech-Language Pathology*, 9(1), 36-45.
- Baylis, A. L., Munson, B., & Moller, K. T. (2011). Perceptions of audible nasal emission in speakers with cleft palate: a comparative study of listener judgments. *The Cleft Palate-Craniofacial journal*, 48(4), 399-411.
- Bernhardt, B., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*, 17(3), 199-216.
- Bernhardt, M. B., Bacsfalvi, P., Adler-Bock, M., Shimizu, R., Cheney, A., Giesbrecht, N., ... & Radanov, B. (2008). Ultrasound as visual feedback in speech habilitation: Exploring consultative use in rural British Columbia, Canada. *Clinical Linguistics & Phonetics*, 22(2), 149-162.
- Bettens, K., Wuyts, F. L., & Van Lierde, K. M. (2014). Instrumental assessment of velopharyngeal function and resonance: A review. *Journal of Communication Disorders*, 52, 170-183.
- Brunnegard, K., Lohmander, A., & van Doorn, J. (2012). Comparison between perceptual assessments of nasality and nasalance scores. *International Journal of Language & Communication Disorders*, 47(5), 556-566.

- Brunner, M., Stellzig-Eisenhauer, A., Pröschel, U., Verres, R., & Komposch, G. (2005). The effect of nasopharyngoscopic biofeedback in patients with cleft palate and velopharyngeal dysfunction. *The Cleft Palate-Craniofacial Journal*, 42(6), 649-657.
- Byun, T. M., & Hitchcock, E. R. (2012). Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *American Journal of Speech-Language Pathology*, 21(3), 207-221.
- Carignan, C. (2014). An acoustic and articulatory examination of the “oral” in “nasal”: The oral articulations of French nasal vowels are not arbitrary. *Journal of Phonetics*, 46, 23-33.
- Davies, T. P. (1979). Schlieren photography- a tool for speech research. *Acoustics Letters* 3(3), 73-75.
- Delvaux, V., Demolin, D., Harmegnies, B., & Soquet, A. (2008). The aerodynamics of nasalization in French. *Journal of Phonetics*, 36(4), 578-606.
- Derrick, D., Anderson, P., Gick, B., & Green, S. (2009). Characteristics of air puffs produced in English “pa”: Experiments and simulations. *The Journal of the Acoustical Society of America*, 125(4), 2272-2281.
- Dotevall, H., Lohmander-Agerskov, A., Ejnell, H., & Bake, B. (2002). Perceptual evaluation of speech and velopharyngeal function in children with and without cleft palate and the relationship to nasal airflow patterns. *The Cleft Palate-Craniofacial Journal*, 39(4), 409-424.
- Drechsel, J. S., & Thomson, S. L. (2008). Influence of supraglottal structures on the glottal jet exiting a two-layer synthetic, self-oscillating vocal fold model. *The Journal of the Acoustical Society of America*, 123(6), 4434-4445.

- Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I., Hansen Edwards, J. G., & Zampini, M. L. (2008). Ultrasound imaging applications in second language acquisition. *Phonology and Second Language Acquisition*, 36, 315-328.
- Gildersleeve-Neumann, C. E., & Dalston, R. M. (2001). Nasalance scores in noncleft individuals: Why not zero? *Cleft Palate-Craniofacial Journal*, 38, 106-111.
- Hasegawa, M., & Kern, E. B. (1977). The human nasal cycle. *Mayo Clinic Proceedings*, 52(1), 28.
- Howard, S., & Lohmander, A. (Eds.). (2011). *Cleft palate speech: assessment and intervention*. West Sussex: John Wiley & Sons.
- Khosla, S., Muruguppan, S., Gutmark, E., & Scherer, R. (2007). Vortical flow field during phonation in an excised canine larynx model. *Annals of Otology, Rhinology & Laryngology*, 116(3), 217-228.
- Krakow, R. A., & Huffman, M. K. (1993). Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: An introduction. In M. K. Huffman, & R. A. Krakow (Eds.), *Phonetics and Phonology: Vol. 5: Nasals, Nasalization and the Velum* (pp.3-62). New York: Academic Press.
- Krane, M., & Gary, S. (2004). Aeroacoustics production of fricative speech sounds. *The Journal of the Acoustical Society of America*, 115(5), 2633-2633..
- Kucinski, B. R., Scherer, R. C., DeWitt, K. J., & Ng, T. T. (2006). Flow visualization and acoustic consequences of the air moving through a static model of the human larynx. *Journal of Biomechanical Engineering*, 128(3), 380-390

- Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In M. K. Huffman & R. A. Krakow (Eds.), *Phonetics and Phonology: Vol. 5: Nasals, Nasalization and the Velum* (pp.147-167). New York: Academic Press.
- Martens, J. W., Versnel, H., & Dejonckere, P. H. (2007). The effect of visible speech in the perceptual rating of pathological voices. *Archives of Otolaryngology–Head & Neck Surgery*, 133(2), 178-185.
- Miller, C. J., & Daniloff, R. (1993). Airflow measurements: theory and utility of findings. *Journal of Voice*, 7(1), 38-46.
- Neubauer, J., Zhang, Z., Miraghaie, R., & Berry, D. A. (2007). Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds. *The Journal of the Acoustical Society of America*, 121(2), 1102-1118.
- Reineke, D. M., Baggett, J., & Elfessi, A. (2003). A note on the effect of skewness, kurtosis, and shifting on one-sample t and sign tests. *Journal of Statistics Education*, 11(3), 1-13.
- Sell, D., Mars, M., & Worrell, E. (2006). Process and outcome study of multidisciplinary prosthetic treatment for velopharyngeal dysfunction. *International Journal of Language & Communication Disorders*, 41(5), 495-511.
- Settles, G. S. (2001). *Schlieren and Shadowgraph Techniques: Visualizing Phenomena in Transparent Media*. Berlin: Springer.
- Settles, G. S., Kester, D. A., & Dodson-Dreibelbis, L. J. (2003). The external aerodynamics of canine olfaction. In F. G. Barth, J. A. C. Humphrey & T. W. Secomb (Eds.), *Sensors and Sensing in Biology and Engineering*. (323-335). Vienna: Springer.
- Shawker, T. H., & Sonies, B. C. (1985). Ultrasound biofeedback for speech training: Instrumentation and Preliminary Results. *Investigative Radiology*, 20(1), 90-93.

- Shinwari, D., Scherer, R. C., DeWitt, K. J., & Afjeh, A. A. (2003). Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees. *The Journal of the Acoustical Society of America*, 113(1), 487-497
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137-149.
- Stevens, K. N. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *The Journal of the Acoustical Society of America*, 50(4B), 1180-1192.
- Tang, J. W., Nicolle, A. D., Pantelic, J., Jiang, M., Sekhr, C., Cheong, D. K., & Tham, K. W. (2011). Qualitative real-time schlieren and shadowgraph imaging of human exhaled airflows: an aid to aerosol infection control. *PloS One*, 6(6), e21392.
- Tanner, W. P. Jr., & Swets, J. A. (1954). A decision making theory of visual detection. *Psychological Review*, 61(6), 401-409.
- Titze, I. R. (2006). *The Myoelastic Aerodynamic Theory of Phonation*. Iowa City: National Center for Voice and Speech.
- Van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, 1(3), 227-244.
- Warren, D. W., Dalston, R. M., & Mayo, R. (1993). Hypernasality in the presence of “adequate” velopharyngeal closure. *The Cleft Palate-Craniofacial Journal*, 30(2), 150-154.
- Warren, D. W., Dalston, R. M., & Mayo, R. (1994). Hypernasality and velopharyngeal impairment. *The Cleft Palate-Craniofacial Journal*, 31(4), 257-262.

Appendices

Appendix A Raw Data

A.1 List of Hits, Correct Rejections, Misses and False Alarms

Rater	Hit	CR	Miss	FA	Hit	CR	Miss	FA	Hit	CR	Miss	FA
		V ₁				V ₂				V ₃		
R1	20	13	3	10	18	17	5	6	20	19	3	4
R2	16	18	7	5	20	15	3	8	15	18	8	5
R3	21	15	2	8	19	13	4	10	20	16	3	7
R4	16	18	7	5	20	18	3	5	20	12	3	11
R5	17	22	6	1	19	16	4	7	19	16	4	7
R6	17	18	6	5	20	15	3	8	21	15	2	8
R7	20	13	3	10	22	15	1	8	22	11	1	12
R8	11	21	12	2	16	18	7	5	16	16	7	7
R9	15	20	8	3	19	20	4	3	18	17	5	6
R10	7	19	16	4	11	17	12	6	14	18	9	5
		A ₁				A ₂				A ₃		
R1	*	*	*	*	21	22	2	1	22	20	1	3
R2	23	23	0	0	23	23	0	0	22	23	1	0
R3	19	21	4	2	23	21	0	2	22	21	1	2
R4	23	23	0	0	23	22	0	1	23	23	0	0
R5	23	22	0	1	22	22	1	1	22	23	1	0
R6	23	23	0	0	23	22	0	1	23	22	0	1
R7	21	17	2	6	21	18	2	5	23	16	0	7
R8	23	23	0	0	23	21	0	2	23	22	0	1
R9	23	20	1	3	23	12	0	11	23	20	0	3
R10	23	23	0	0	23	23	0	0	23	23	0	0
		C ₁				C ₂				C ₃		
R1	22	21	1	2	21	21	2	2	21	20	2	3
R2	21	23	2	0	22	23	1	0	23	22	0	1
R3	23	21	0	2	23	20	0	3	23	20	0	3
R4	23	23	0	0	23	23	0	0	23	21	0	2
R5	23	23	0	0	23	22	0	1	23	23	0	0
R6	23	22	0	1	23	23	0	0	23	22	0	1
R7	23	13	0	10	23	19	0	4	23	18	0	5
R8	23	22	0	1	23	20	0	3	23	22	0	1
R9	23	23	0	0	23	21	0	2	23	18	0	5
R10	23	22	0	1	23	23	0	0	23	23	0	0

Note: Absent data is marked by *

A.2 Rates for Hits, Correct Rejections, Misses and False Alarms

Rater	Hit	CR	Miss	FA	Hit	CR	Miss	FA	Hit	CR	Miss	FA
		V_1				V_2				V_3		
R1	0.87	0.57	0.13	0.43	0.78	0.74	0.22	0.26	0.87	0.83	0.13	0.17
R2	0.70	0.78	0.30	0.22	0.87	0.65	0.13	0.35	0.65	0.78	0.35	0.22
R3	0.91	0.65	0.09	0.35	0.83	0.57	0.17	0.43	0.87	0.70	0.13	0.30
R4	0.70	0.78	0.30	0.22	0.87	0.78	0.13	0.22	0.87	0.52	0.13	0.48
R5	0.74	0.96	0.26	0.04	0.83	0.70	0.17	0.30	0.83	0.70	0.17	0.30
R6	0.74	0.78	0.26	0.22	0.87	0.65	0.13	0.35	0.91	0.65	0.09	0.35
R7	0.87	0.57	0.13	0.43	0.96	0.65	0.04	0.35	0.96	0.48	0.04	0.52
R8	0.48	0.91	0.52	0.09	0.70	0.78	0.30	0.22	0.70	0.70	0.30	0.30
R9	0.65	0.87	0.35	0.13	0.83	0.87	0.17	0.13	0.78	0.74	0.22	0.26
R10	0.30	0.83	0.70	0.17	0.48	0.74	0.52	0.26	0.61	0.78	0.39	0.22
		A_2				A_2				A_2		
R1	*	*	*	*	0.91	0.96	0.09	0.04	0.96	0.87	0.04	0.13
R2	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.96	1.00	0.04	0.00
R3	0.83	0.91	0.17	0.09	1.00	0.91	0.00	0.09	0.96	0.91	0.04	0.09
R4	1.00	1.00	0.00	0.00	1.00	0.96	0.00	0.04	1.00	1.00	0.00	0.00
R5	1.00	0.96	0.00	0.04	0.96	0.96	0.04	0.04	0.96	1.00	0.04	0.00
R6	1.00	1.00	0.00	0.00	1.00	0.96	0.00	0.04	1.00	0.96	0.00	0.04
R7	0.91	0.74	0.09	0.26	0.91	0.78	0.09	0.22	1.00	0.70	0.00	0.30
R8	1.00	1.00	0.00	0.00	1.00	0.91	0.00	0.09	1.00	0.96	0.00	0.04
R9	0.96	0.87	0.04	0.13	1.00	0.52	0.00	0.48	1.00	0.87	0.00	0.13
R10	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
		C_3				C_3				C_3		
R1	0.96	0.91	0.04	0.09	0.91	0.91	0.09	0.09	0.91	0.87	0.09	0.13
R2	0.91	1.00	0.09	0.00	0.96	1.00	0.04	0.00	1.00	0.96	0.00	0.04
R3	1.00	0.91	0.00	0.09	1.00	0.87	0.00	0.13	1.00	0.87	0.00	0.13
R4	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.91	0.00	0.09
R5	1.00	1.00	0.00	0.00	1.00	0.96	0.00	0.04	1.00	1.00	0.00	0.00
R6	1.00	0.96	0.00	0.04	1.00	1.00	0.00	0.00	1.00	0.96	0.00	0.04
R7	1.00	0.57	0.00	0.43	1.00	0.83	0.00	0.17	1.00	0.78	0.00	0.22
R8	1.00	0.96	0.00	0.04	1.00	0.87	0.00	0.13	1.00	0.96	0.00	0.04
R9	1.00	1.00	0.00	0.00	1.00	0.91	0.00	0.09	1.00	0.78	0.00	0.22
R10	1.00	0.96	0.00	0.04	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00

Note: Absent data is marked by *

A.3 Values for d-Prime and Bias

<i>d-prime Values for "Video-only" Condition</i>						<i>Bias values for "Video-only" Condition</i>					
Rater	Blocks			Mean	SD	Rater	Blocks			Mean	SD
	1	2	3				1	2	3		
R1	1.29	1.42	2.06	1.59	0.41	R1	-0.48	-0.07	-0.09	-0.21	0.23
R2	1.29	1.52	1.17	1.33	0.17	R2	0.13	-0.37	0.19	-0.01	0.31
R3	1.75	1.10	1.64	1.50	0.35	R3	-0.48	-0.39	-0.31	-0.39	0.09
R4	1.29	1.91	1.18	1.46	0.39	R4	0.13	-0.17	-0.53	-0.19	0.34
R5	2.35	1.45	1.45	1.75	0.52	R5	0.54	-0.21	-0.21	0.04	0.43
R6	1.42	1.52	1.75	1.56	0.17	R6	0.07	-0.37	-0.48	-0.26	0.29
R7	1.29	2.10	1.66	1.68	0.41	R7	-0.48	-0.66	-0.88	-0.67	0.20
R8	1.31	1.29	1.02	1.21	0.16	R8	0.71	0.13	0.00	0.28	0.38
R9	1.52	2.06	1.42	1.67	0.35	R9	0.37	0.09	-0.07	0.13	0.22
R10	0.43	0.59	1.06	0.69	0.33	R10	0.73	0.35	0.25	0.44	0.25
Mean	1.39	1.50	1.44	Tot mean	Tot SD	Mean	0.12	-0.17	-0.21	Tot mean	Tot SD
SD	0.48	0.46	0.34	1.44	0.31	SD	0.48	0.30	0.35	-0.09	0.33

<i>d-prime Values for "Audio-only" Condition</i>						<i>Bias values for "Audio-only" Condition</i>					
Rater	Blocks			Mean	SD	Rater	Blocks			Mean	SD
	1	2	3				1	2	3		
R1	2.00	3.07	2.84	2.64	0.56	R1	0.36	0.18	-0.29	0.08	0.34
R2	4.14	4.14	3.62	3.97	0.30	R2	0.00	0.00	0.26	0.09	0.15
R3	2.30	3.34	3.07	2.90	0.54	R3	0.21	-0.40	-0.18	-0.12	0.31
R4	4.14	3.62	4.14	3.97	0.30	R4	0.00	-0.26	0.00	-0.09	0.15
R5	3.62	3.42	3.62	3.55	0.11	R5	-0.26	0.00	0.26	0.00	0.26
R6	4.14	3.62	3.62	3.79	0.30	R6	0.00	-0.26	-0.26	-0.17	0.15
R7	2.00	2.14	2.56	2.23	0.29	R7	-0.36	-0.29	-0.79	-0.48	0.27
R8	4.14	3.34	3.62	3.70	0.41	R8	0.00	-0.40	-0.26	-0.22	0.20
R9	2.84	2.12	3.13	2.70	0.52	R9	-0.29	-1.01	-0.51	-0.60	0.37
R10	4.14	4.14	4.14	4.14	0.00	R10	0.00	0.00	0.00	0.00	0.00
Mean	3.35	3.29	3.43	Tot mean	Tot SD	Mean	-0.03	-0.25	-0.18	Tot mean	Tot SD
SD	0.95	0.70	0.52	3.36	0.68	SD	0.22	0.33	0.33	-0.15	0.23

<i>d-prime Values for "Combined" Condition</i>						<i>Bias values for "Combined" Condition</i>					
Rater	Blocks			Mean	SD	Rater	Blocks			Mean	SD
	1	2	3				1	2	3		
R1	3.07	2.72	2.48	2.76	0.30	R1	-0.18	0.00	-0.12	-0.10	0.09
R2	3.34	3.62	3.62	3.52	0.16	R2	0.40	0.26	-0.26	0.13	0.35
R3	3.34	3.13	3.13	3.20	0.12	R3	-0.40	-0.51	-0.51	-0.47	0.06
R4	4.14	4.14	3.34	3.87	0.46	R4	0.00	0.00	-0.40	-0.13	0.23
R5	4.14	3.62	4.14	3.97	0.30	R5	0.00	-0.26	0.00	-0.09	0.15
R6	3.62	4.14	3.62	3.79	0.30	R6	-0.26	0.00	-0.26	-0.17	0.15
R7	2.23	2.96	2.81	2.67	0.39	R7	-0.96	-0.59	-0.66	-0.74	0.19
R8	3.62	3.13	3.62	3.45	0.28	R8	-0.26	-0.51	-0.26	-0.34	0.14
R9	4.14	3.34	2.81	3.43	0.67	R9	0.00	-0.40	-0.66	-0.35	0.33
R10	3.62	4.14	4.14	3.97	0.30	R10	-0.26	0.00	0.00	-0.09	0.15
Mean	3.52	3.49	3.37	Tot mean	Tot SD	Mean	-0.19	-0.20	-0.31	Tot mean	Tot SD
SD	0.59	0.52	0.56	3.46	0.47	SD	0.35	0.29	0.24	-0.24	0.25

Appendix B Practice Session Instructions

Nasality Study

Instructions Section

In this study you will judge whether words contain either a:

nasal vowel

or

non-nasal (oral) vowel

Nasal sounds are produced when the velopharyngeal port opens, allowing air and sound to travel and resonate through the nasal cavity

You will see audio and video recordings of words whose vowels are either nasal or not nasal

Words are presented in CV structure

Nasality is associated with nasal resonance:

- when a sound is not nasal (i.e. “oral” sounds), sound resonates primarily in the speaker’s mouth
- when a sound is nasal, sound resonates in both the mouth *and* nose

Let’s practice: the following audio recording is nasal

(click to play)



Nasality can vary by individual speaker. Listen to the resonance of this sound. Try saying it aloud to help feel the nasal resonance.

The following audio recording is not nasal

(click to play)



Repeat the sound. You should feel a difference in how much nasal resonance is produced. In sounds that aren't nasal, less vibration should be felt throughout the nasal cavity.

Listen to them side by side

Nasal

Not Nasal



Nasality is associated with nasal airflow:

- when a sound is not nasal (i.e. “oral” sounds), air flows from the speaker’s mouth
- when a sound is nasal, air flows from both the mouth *and* nose

Note: breathing (especially at the end of a word) looks very similar to nasal airflow associated with nasality. During your assessment, please focus on nasal airflow near the beginning of the recording, not the end.

Let’s practise: this video shows a non-nasal sound (oral)



Notice that at the beginning of the word, airflow only comes from the speaker’s mouth. The speaker breathes out near the end of the word, producing a puff of air from the nose.

Click video to repeat

The following video shows a nasal sound



Notice that at the beginning of the word, airflow comes from both the mouth and nose. The speaker exhales (breath) through the nose at the end of the word.

Click video to repeat

Look at them side by side



Not nasal



Nasal

You will now complete 2 practice sessions; 10 audio recordings, and 10 video recordings.

Rate stimuli as “nasal” or “not nasal” by using your mouse to click the appropriate button. Each recording will play once only.

After practising, you will begin the main experiment:

- You will view stimuli in blocks of audio-only, video-only, or combined audio-video
- Each block of stimuli consists of 46 words. You may take a break at the end of each block.
 - You will rate each block 2-3 times

The experiment will take a little over 1 hour to complete