Brigham Young University

BYU ScholarsArchive

2016-03-01

Speech Adaptation to Kinematic Recording Sensors

Elise Hansen Hunter
Brigham Young University - Provo

Speech Adaptation to Kinematic Recording Sensors

Elise Hansen Hunter

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Christopher Dromey, Chair
Ron Channell
Shawn Nissen

Department of Communication Disorders

Brigham Young University

March 2016

ABSTRACT

Speech Adaptation to Kinematic Recording Sensors

Elise Hansen Hunter
Department of Communication Disorders, BYU
Master of Science

This thesis examined the time course of speech adaptation prior to data collection when using an electromagnetic articulograph to measure speech articulator movements. The stimulus sentence and electromagnetic sensor placement were designed to be sensitive to changes in the fricatives /s/ and /ʃ/. Twenty native English speakers read aloud stimulus sentences before the attachment of six electromagnetic sensors, immediately after attachment, and again at 5, 10, 15 and 20 minutes after attachment. Participants read aloud continuously between recordings to encourage adaptation to the presence of the sensors. Audio recordings were rated by 20 native English listeners who were not part of the production study. After listening to five practice samples, these participants rated 150 stimuli (31 repeat samples) using a visual analog scale (VAS) with the endpoints labeled as *precise* and *imprecise*. An acoustic analysis of the recordings was done by segmenting the fricatives /s/ and /ʃ/ from the longer recording and computing spectral center of gravity and spectral standard deviation in Hertz. Durations of /s/, /ʃ/ and the sentence were also measured. Results of both perceptual and acoustic analysis revealed a change in speech precision over time, with all post attachment recordings receiving lower perceptual scores. Precision ratings beyond the 10 minute recording remained steady. It can be concluded from the results that participants reached a height of adaptation after 10 minutes of talking with kinematic recording sensors attached, and that after the attachment of sensors, speech production precision did not at any point return to pre attachment levels.

ACKNOWLEDGMENTS

I would like to thank Dr. Christopher Dromey for his invaluable feedback and mentoring over the course of my work on this thesis and my education. With the countless meetings, data collection sessions and drafts, he has provided both support and reinforcement to help me complete this project.

I would like to thank my parents for their constant encouragement and support in pursuing my education. They have always inspired me to push myself to the best of my abilities and have provided excellent examples of hard work. I could never have accomplished this achievement without their unwavering support.

Last, but certainly not least, I have to thank my amazing husband Scott. In the pursuit of my master's degree and in the process of writing my thesis, he has been beside me every step of the way. He has encouraged me through hard times and pushed me to become the very best speech language pathologist I can be. I love him so much and know this process would not have been as enjoyable without him.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF THESIS CONTENT

This thesis, *Speech Adaptation to Kinematic Recording Sensors,* is the result of a research project and portions of this thesis may be published as part of articles listing the thesis author as a co-author. The body of this thesis is written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology. An annotated bibliography is presented in Appendix A and an informed consent is presented in Appendix B.

Introduction

Measuring Speech Production

Acoustic recordings allow the researcher to collect data about speech without interfering with the speech mechanism itself. This noninvasive approach can provide valuable quantitative data about phonation, resonance, or articulation. However, there can be some ambiguity in the interpretation of acoustic measures in relation to the movements they represent. This is largely due to motor equivalence, whereby several possible combinations of articulatory movements may result in the same perceptual result. Research has shown that motor equivalence may be involved in the production of many English phonemes. Joseph Perkell (2001) reported that for the vowel /u/ "there is a many-to-one relation between articulations and the vowel acoustics." Thus, the same vowel acoustics could be produced "with more tongue raising and less lip rounding, or vice versa" (p. 3). Motor equivalence allows speakers to accomplish phonetic targets in a variety of ways. While this allows the speaker some flexibility in the complex act of speech production, it also limits the extent to which the researcher may rely on acoustic data to understand articulator movements. Ultimately, the acoustic consequence of speech production is more important to the listener than the movements made by the speaker. However, from a basic science and clinical research perspective, knowing more about the activity of the articulators can be valuable in understanding both normal and disordered speech.

Kinematic measures of speech overcome the limitations inherent in acoustic analysis by directly tracking the gestures of the articulators, thus avoiding the potential ambiguity introduced by motor equivalence. Speech production is a complex neuromuscular task that relies on rapid and precise movements. Since direct access to the activity of the brain is not possible, speech kinematic measures are the closest available reflection of what the brain is doing during speech

production (van der Merwe, 1997). Measuring movement in disordered speech allows clinical researchers a greater understanding of what may be happening in articulatory control to contribute to the communication problem. This information may ultimately help clinicians develop more functional and efficient treatments by targeting the disordered movements on the basis of scientific evidence.

Measuring articulator movement is, by nature, more invasive than obtaining acoustic measurements. It involves attaching transducers to the articulatory structures. Invasive tools such as these, while allowing the direct measurement of speech, have the potential to change the natural speech movements in two different ways. They may alter the way that speech sounds and also the way speech feels to the speaker.

The Role of Feedback

Feedback is an important part of everyday life. It contributes to the process of controlling the way the body interacts with the environment in order to complete daily tasks such as walking, speaking, or driving. There are two main forms or systems of feedback. The first is illustrated by driving along a winding road. In performing this task, the brain is receiving continuous visual and tactile feedback to help the driver navigate each turn. This is referred to as a closed-loop feedback system because there is continuous feedback which directly influences ongoing performance (van der Merwe, 1997). The second feedback system is referred to as an open-loop system. This can be illustrated by the act of shooting a basketball. In contrast to the previous analogy of driving a car, in the act of shooting a basketball, all preparations and adjustments are carried out before the ball leaves the hand. The athlete cannot make any adjustments to the shot after this point. In an open-loop system, afferent input is "relatively

unimportant" as it does not directly influence performance while it is happening (van der Merwe, 1997).

Typical adult speech motor control is considered to rely on open-loop processes. The production of phonemes occurs too rapidly to allow for the continuous feedback that occurs in a closed-loop system. On a phoneme-by-phoneme basis, closed-loop control cannot account for the speed of speech production. Phonemes are produced in a matter of milliseconds (Nooteboom & Doodeman, 1980). This allows insufficient time for acoustic output to be processed by the brain and the articulatory movements altered (Zheng, Munhall, & Johnsrude, 2010). Although feedback is too slow for concurrent adjustments to articulation, it still remains an important factor for long-term quality control.

Auditory and Tactile Feedback

In infancy, speech acquisition occurs as the child experiments with the speech mechanism using feedback during babbling. The infant learns through trial and error by moving the articulators and evaluating the resulting acoustic output. Both auditory and tactile feedback play an important role in the development of speech and language (Rees, 1972).

Speech is a complex sensorimotor skill, which integrates both auditory and tactile feedback for quick and precise sound production. Although their roles are different, auditory and tactile feedback co-occur during speech production. Researchers agree that sensory input or feedback is an "integral part of movement control and coordination" in speech (van der Merwe, 1997, p. 3). Tactile-kinesthetic input is direct feedback and many argue that it plays a larger role and is faster than auditory feedback in the control of speech production (van der Merwe, 1997). Other researchers have shown the dominance of tactile-kinesthetic feedback by experimentally

altering both auditory and tactile-kinesthetic feedback during speech production and measuring the compensation for each (Lametti, Nasir, & Ostry, 2012).

While it is clear that tactile-kinesthetic input is integral to speech output, auditory input also makes a vital contribution to the quality and accuracy of speech production. The importance of auditory feedback is revealed by the disruptive effects of altered auditory feedback (AAF). In studies where auditory feedback is either altered or delayed, the quality and precision of speech production is compromised (Yates, 1963). As noted earlier, auditory feedback plays a larger role during development than in mature speech. This is illustrated in Waldstein's study (1990) on the effects of postlingual deafness on speech production. Adults with previously normal speech who lose their hearing, and therefore auditory feedback, are still able to maintain intelligible speech. Yet over time, as Waldstein reports, the lack of auditory feedback affects speech precision and quality (1990). In this way, it can be concluded that auditory feedback after speech development, while still important, primarily acts as an overall quality control mechanism for speech production.

Perturbation

Perturbation of the speech system inherently alters speech production and consequently both auditory and tactile-kinesthetic feedback. While perturbation of the speech mechanism can occur in everyday situations (for example, when a person has a dental retainer), it also occurs when kinematic transducers are attached to the articulators, bite blocks are placed between the molars, or pseudopalates are used for articulation therapy. These instruments introduce a foreign object into the vocal tract that temporarily alters the morphology of the oral cavity. This would be considered a static perturbation, because the size and shape of the transducer or device does not change and it is present for all spoken sounds. In contrast, dynamic perturbations are not seen

in everyday situations. They are artificial and only occur in a small number of speech experiments in specialized labs. Dynamic perturbations are caused when a device (such as a robotic arm) introduces an atypical disturbance to the movements involved in speech production. An example would be a push on the jaw during production of a bilabial consonant; this methodology was used in a research study by Lametti et al. (2012). This unexpected disturbance to the movement altered the typical path of the jaw for the bilabial gesture, and thus represented a dynamic experimental perturbation.

Research into both dynamic and static perturbation has revealed valuable information about speech motor control and the integration of sensory information into motor output. These insights have been gained by studying the compensations made by speakers in response to altered sensory feedback. After the initial perturbation, all of the compensations of an individual speaker are collectively referred to as adaptation (McFarland, Baum, & Chabot, 1996). McFarland et al. measured the compensations made by multiple speakers after structural modification of the oral cavity by a pseudopalate. From perceptual data collected initially after placement of the pseudopalate and again after fifteen minutes, McFarland et al. concluded that although speech is perturbed by the structural modification, adaptation takes place over time.

McFarland et al. (1995) also discovered in a separate study that while compensations are made over time for altered sensory feedback, adaptation is incomplete. This is due to the fact that rather than returning to preperturbed natural speech, which is what adaptation would suggest, the speech of the individual has changed. Thus, while speech may begin to sound more like the individual's natural speech, its production has been adapted to compensate for the perturbation. In this way, the use of kinematic instruments to measure speech movements may not actually allow measurement of truly natural speech, but rather adapted speech. Research into the

adaptations that are made by the speech system upon the introduction of a perturbation has the potential to reveal important insights into speech motor control (Aasland, Baum, & McFarland, 2006).

It is understood that adaptation takes place (McFarland et al., 1996), yet there is little research concerning the average amount of time it takes for an individual to reach the height of adaptation (or when compensations have essentially reached a plateau). The amount of time allocated by experimenters for compensation and adaptation to the presence of speech transduction devices varies widely. In one study on speech adaptation to the presence of a pseudopalate, the researchers allowed forty-five minutes (McAuliffe, Robb, & Murdoch, 2007), while another similar study of speech adaptation to a dental retainer relied on data gathered after two weeks of adaptation time (Hamlet, 1985). Other studies of adaptation have used varying durations between fifteen minutes and an hour (Aasland et al., 2006; McFarland et al., 1996). With this amount of variation in the literature, research that could reveal the average amount of time needed for an individual to reach the height of adaptation to static perturbations would be beneficial to speech kinematic researchers for future studies.

Research involving the measurement of articulatory movement relies on a variety of technologies. Electromagnetic articulography is one approach that is currently used in laboratories around the world. These systems use sensors adhered to articulators and magnetic field technology to measure the three-dimensional position of the sensors over time. The Northern Digital Instruments (NDI) Wave system (Waterloo, Ontario, Canada) is one type of electromagnetic articulograph, and was the focus of the present study. It was anticipated that the presence of the sensor coils on the articulators and the small wire from each senor exiting the mouth would perturb speech production. Therefore, the purpose of this study was to learn how

much time would be needed after the attachment of the sensors for the speaker to adapt and speak naturally. This information may be valuable in future studies using this technology by providing researchers with a guide concerning the amount of time needed for adaptation before the collection of data. The collection and analysis of both acoustic and perceptual data in the present study allowed the measurement of adaptation effects in the speech of research participants.

Methods

The present study was part of a larger project that addressed several different research questions using data from the same speakers. There were two components of this study: the recording of audio data and the perceptual evaluation by listeners coupled with an acoustic analysis of selected sounds. Both components are discussed below.

Speech Recordings

Participants. Ten women ages 20 to 34 ($M = 25.1$, $SD = 4.01$) and ten men ages 20 to 32 ($M = 25.3$, $SD = 3.5$), all native speakers of English with no history of speech or hearing disorders, participated in the study. They were seated in a sound booth for the audio recordings.

Instrumentation. The NDI Wave system sensors were present for this experiment, but the system was not activated to collect kinematic data. Six sensors were used for each speaker. Five sensors were glued to the participants' articulators using Glustitch PeriAcryl90 (Glustich Inc., Delta, British Columbia, Canada). Two sensors were glued along the midline of the tongue. One was 1 cm posterior to the tip and the other was at the tongue front/back midpoint. Another sensor was glued to a 5mm by 5mm piece of Stomahesive wafer (ConvaTec, Greensboro, NC), which was then attached to the lower incisors at midline. The last two sensors were glued to the lips at midline: one was at the lower lip vermillion border and the other at the upper lip

vermillion border. The reference sensor was attached to an eyeglass frame without lenses, worn

by the participant. The reference sensor served as the origin of the coordinate system for all other

sensors to correct for head movement during recordings. A condenser microphone (AKG model

C2000B) connected via a Focusrite Scarlett 2i2 analog to digital audio converter recorded the

audio signal as the participant spoke. The microphone was 30 centimeters from the mouth.

Audacity 2.0 software was used to record the audio signal at 44.1 kHz, and subsequently to

segment the microphone data for analysis.

Stimuli and procedure. After reading and signing an IRB approved consent form,

participants sat in a sound booth for the collection of data. Stimuli were presented visually in a

48-point black font on white paper on the wall of the sound booth in front of each individual.

The participants read aloud a list of six sentences (see Table 1), five times through. Sentences

were read aloud before the attachment of sensors, immediately after attachment, and again at

five, ten, fifteen, and twenty minutes after attachment. Sentence 1 was the focus of the present

study and was the only sentence segmented and used in the perceptual analysis. Between

recordings of stimuli, participants continuously read aloud from a book or newspaper to adjust to

the presence of the sensors on their articulators.

Table 1

Sentences Read Aloud by Participants

Sentence Order	Sentence
1	It's time to shop for two new suits.
2	A good AC should keep your car cool.
3	It's never too soon to choose the right.
4	One warm morning a boy was mowing the lawn.
5	We do agree the loud noise is annoying.
6	There's no good reason they would go down there.

Acoustic and Perceptual Evaluation

Participants. Ten women ages 22 to 32 (M = 25.5, SD =2.95) and ten men ages 21 to 31 (M = 26.2, SD=3.04), all native speakers of English with no history of speech or hearing disorders, participated as raters in the perceptual evaluation.

Instrumentation. The raters sat in a sound booth at a computer workstation to complete the perceptual evaluation task. They listened to stimuli through loudspeakers that they adjusted to a comfortable loudness. The participants selected, listened to, and rated stimuli using a custom Matlab (2014b) application.

Stimuli and procedure. Sentence 1 from Table 1 was used in the perceptual evaluation. The first correctly produced sentence recording of the five repetitions was segmented and amplitude-normalized so that the listening level was consistent across samples. After reading and signing an IRB approved consent form, participants listened to five practice samples with examples of precise and imprecise speech production included to acquaint them with the range of

stimuli they would be hearing. Participants then listened to 119 samples (with 31 randomly repeated sentences) for a total of 150 stimuli, each sample lasting approximately three seconds. The sequence of stimulus presentation was randomized for the study, but not separately for each listener. After listening to each sample, participants rated the sample using a visual analog scale (VAS) on a computer screen with the endpoints labeled *precise* and *imprecise.* Participants were permitted to listen to the stimuli as many times as necessary.

Acoustic analysis. The phonemes /s/ and /ʃ/ were identified using the combined audio waveform and spectographic display to segment the fricatives from the longer recording. Using the Praat (version 5.4) spectral slice function, the center of gravity and standard deviation in Hertz were computed. These measures were extracted from the first three correct productions at each time recording and were averaged in Microsoft Excel. The duration of /ʃ/ in "shop," /s/ in "suits" and the duration of the sentence for the first three correct productions was also measured using the Praat combined audio waveform and spectrographic display and averaged in Microsoft Excel.

Statistical analysis. The means for three repetitions for each measure were prepared in Excel for statistical analysis with SPSS (version 21). The acoustic and perceptual variables were tested for significant change over time by performing a repeated measures ANOVA with concurrent contrasts comparing each recording against the preattachment recording. Acoustic variables included spectral center of gravity, spectral spread, and duration for /ʃ/ and /s/ as well as utterance duration. In any instance where the assumption of sphericity was violated, the Huynh-Feldt corrected degrees of freedom were used. Perceptual variables included listener ratings of speech precision from zero to 100, with zero representing imprecise and 100 precise speech production.

Results

Perceptual Ratings

   Rater reliability. During the perceptual evaluation, 31 samples were repeated to evaluate

intra-rater reliability. Only the data from raters with a correlation between original and repeated

ratings above 0.74 were included in the perceptual analysis. This resulted in the use of rating

from 12 raters. The remaining eight raters' correlations ranged from 0.601 to 0.736, and their

data were deemed insufficiently reliable to use in the main analysis. For the 12 reliable raters,

inter-rater consistency was evaluated with an intraclass correlation coefficient (ICC). The single

measures ICC was .656 and the average measures ICC was .958, $F(118,1298)=23.842$, $p<.001$

   Adaptation effects. Repeated measures ANOVA revealed a significant change in

articulatory precision over time. All postattachment ratings were lower than preattachment

ratings (see Table 2 for descriptive statistics). Furthermore, the concurrent contrast tests revealed

that all recordings after the preattachment recording were rated as significantly poorer in

precision. Speech precision ratings changed little after the 10-minute recording (see Figure 1 and

Table 2); all post-attachment perceptual ratings remained lower than the ratings of the

preattachment recording. Data collection for one male participant at one recording time was

missed; therefore, the analysis was run for 19 subjects.

Table 2

Descriptive Statistics for the Perceptual Ratings Over Time (0 represents imprecise and 100 represents precise speech production)

Recording Time	Mean	Standard Deviation
Pre-sensor Attachment	89.8	6.1
Post-sensor Attachment	43.5	22.5
5 minutes post	51.5	22.7
10 minutes post	60.5	23.7
15 minutes post	60.1	20.9
20 minutes post	61.3	21.0
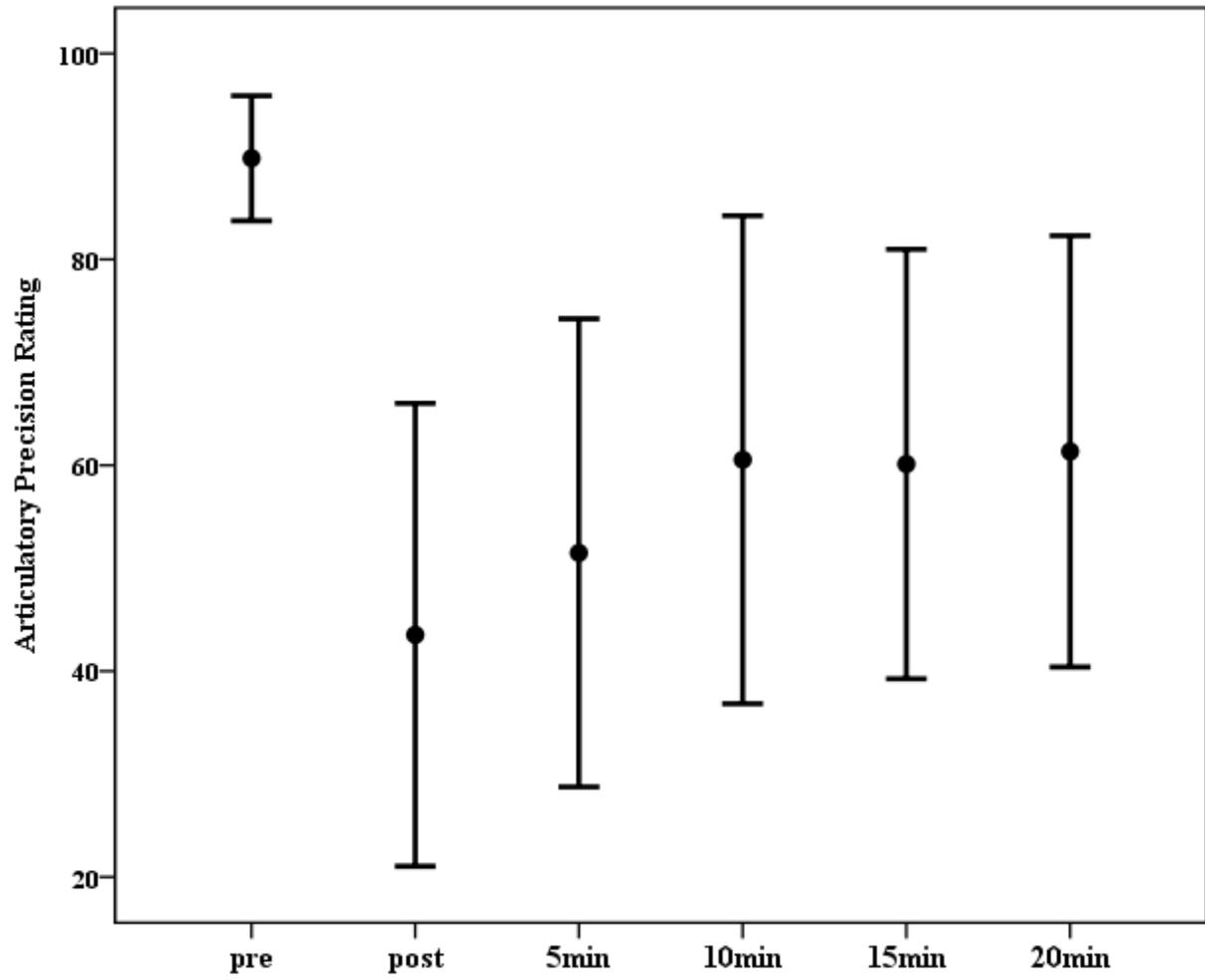
Figure 1. Mean and standard deviation of perceptual ratings of articulatory precision over time.

Table 3

Repeated Measures ANOVA Within-Subjects Contrasts for Each Perceptual Rating Compared to the Preattachment Condition

Recording Time	*F*	*Sig.*	*Partial Eta Squared*
Immediately post	72.945	$p < .001$	.802
5 minutes	46.353	$p < .001$	.720
10 minutes	25.300	$p < .001$	.584
15 minutes	30.772	$p < .001$	.631
20 minutes	28.264	$p < .001$	.611

Acoustic Measures

The descriptive statistics for acoustic measures are reported in Tables 4-10. Results of the Repeated measures ANOVA tests are reported in Table 11 and the contrast results in Table 12.

Duration. The duration of the sentence changed significantly, and contrasts revealed that recordings beyond 10 minutes were shorter than the preattachment recording. The duration of /s/ changed significantly, and contrasts revealed that all recordings after sensor attachment were shorter than the preattachment recording. The duration of /ʃ/ changed significantly, and contrasts revealed that the duration in recordings at 5, 10, and 20 minutes postattachment were shorter than in the preattachment recording.

Spectral measures. The spectral center of gravity for /s/ changed significantly, and contrasts revealed that for recordings immediately post and at 10 and 20 minutes, this measure was lower than in the preattachment recording. The spectral spread for /s/ changed significantly and contrasts revealed that all postattachment recordings had a lower value than the preattachment recording. The spectral center of gravity for /ʃ/ changed significantly, and

contrasts revealed that all post attachment recordings had a higher value than the preattachment recording. There was a significant male/female difference, with women having a higher spectral center of gravity than men $F(1, 17) = 9.353$, $p = .007$, E.S.= .355. The spectral spread for /ʃ/ changed significantly and contrasts revealed that it was higher for all postattachment recordings than the preattachment recording.

Table 4

Mean and Standard Deviation of Sentence 1 Duration in Seconds for Each Recording Time

Recording Time of Sentence 1	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	2.412	2.225	0.3352	0.3980
Postattachment	2.491	2.324	0.3199	0.2719
5 minutes post	2.250	2.193	0.2782	0.3264
10 minutes post	2.173	2.101	0.2538	0.2051
15 minutes post	2.227	2.065	0.2358	0.2072
20 minutes post	2.164	2.016	0.2163	0.4694
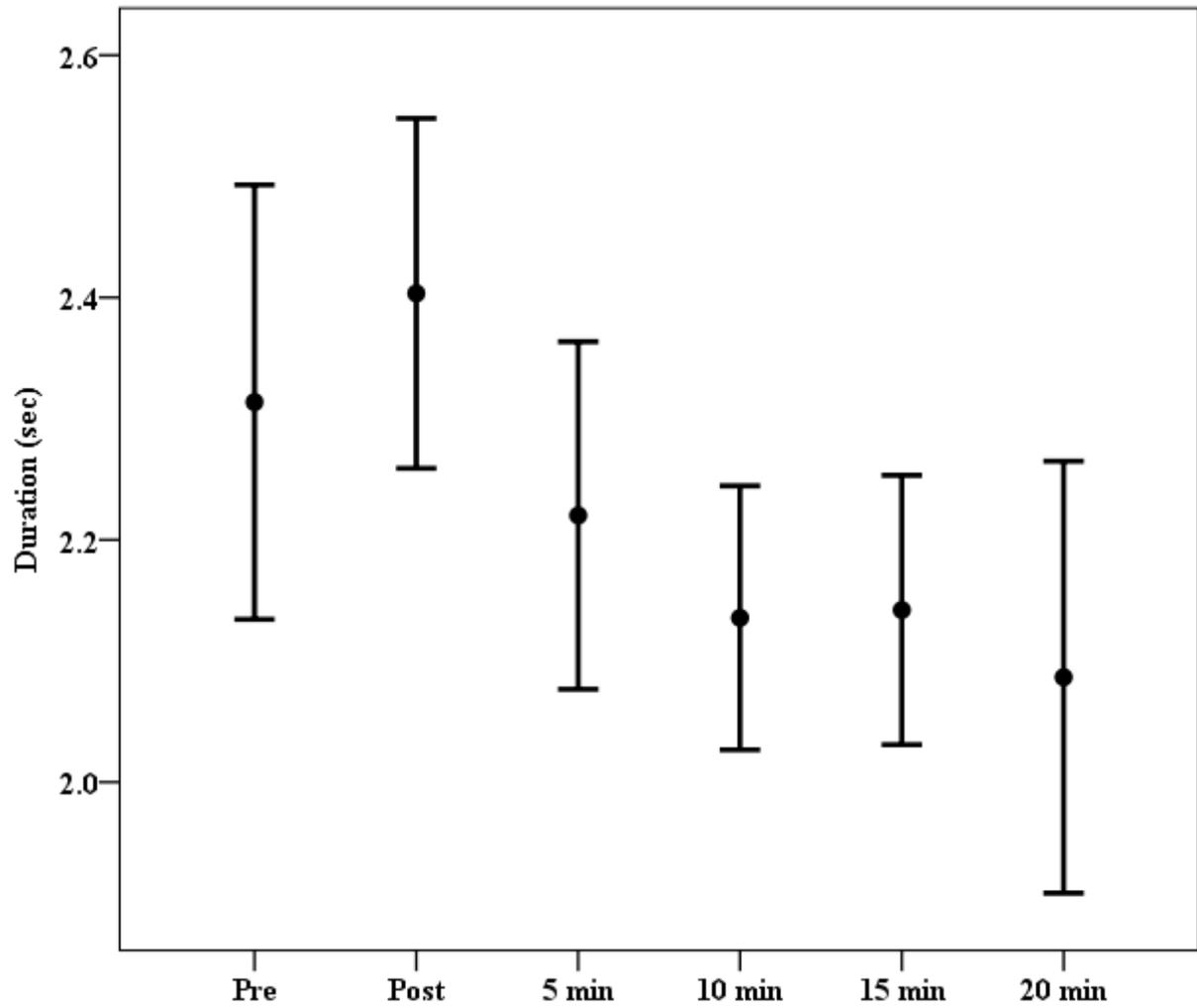
Figure 2. Mean duration of sentence 1 in seconds over time.

Table 5

Mean and Standard Deviation of /s/ Duration in Seconds for Each Recording Time

Recording	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	0.151	0.134	0.0296	0.0195
Postattachment	0.142	0.132	0.0339	0.0167
5 minutes post	0.138	0.122	0.0267	0.0107
10 minutes post	0.137	0.124	0.0215	0.0096
15 minutes post	0.109	0.101	0.0192	0.0077
20 minutes post	0.119	0.121	0.0221	0.0127
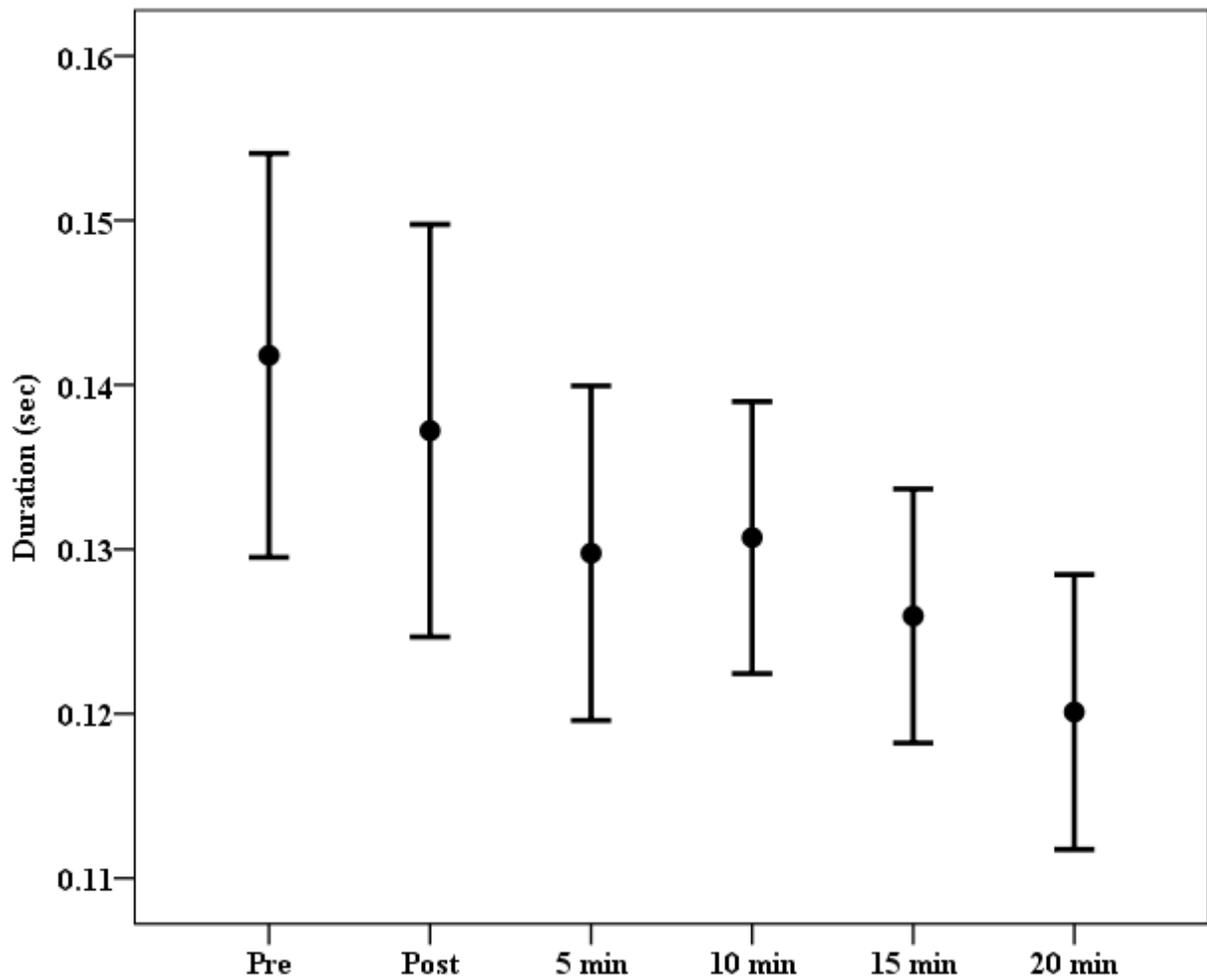
Figure 3. Duration of /s/ in seconds over time.

Table 6

Mean and Standard Deviation of /ʃ/ Duration in Seconds for Each Recording Time

Recording Time of /ʃ/	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	0.114	0.108	0.0252	0.0168
Postattachment	0.110	0.114	0.0251	0.0099
5 minutes post	0.101	0.099	0.0184	0.0076
10 minutes post	0.098	0.099	0.0146	0.0073
15 minutes post	0.109	0.101	0.0192	0.0077
20 minutes post	0.098	0.097	0.0158	0.0072
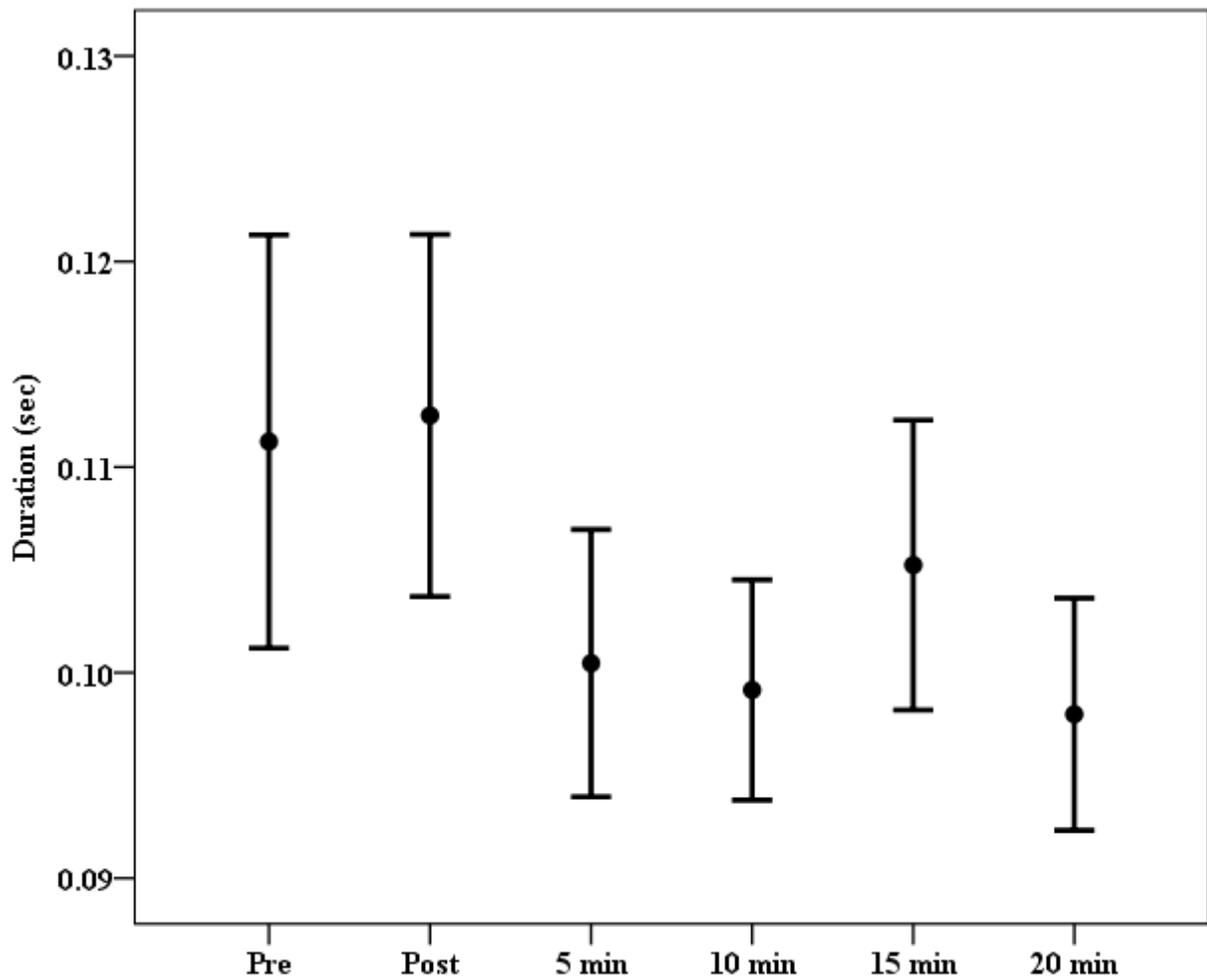
Figure 4. Duration of /ʃ/ in seconds over time.

Table 7

Mean and Standard Deviation of /s/ Spectral Center of Gravity in Hertz for Each Recording

Time

Recording time	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	5262.85	5848.04	455.04	754.16
Postattachment	4890.52	5524.40	373.87	1138.35
5 minutes post	4774.42	5477.76	937.04	1513.16
10 minutes post	4767.23	5530.29	504.93	907.23
15 minutes post	4852.25	5617.51	707.27	1271.71
20 minutes post	4900.48	5459.15	480.98	1164.91
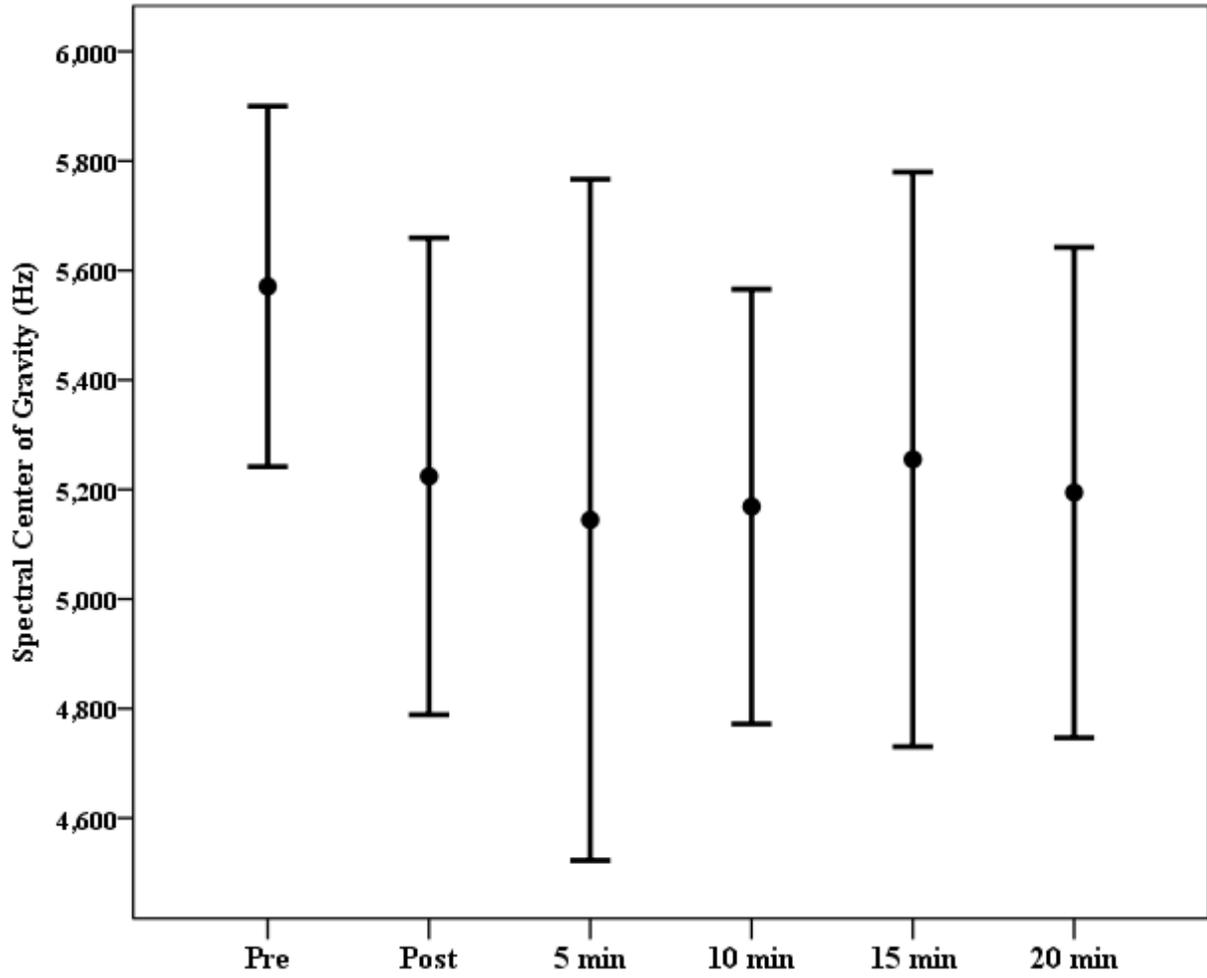
Figure 5. Spectral center of gravity of /s/ in Hertz over time.

Table 8

Mean and Standard Deviation of /s/ Spectral Spread in Hertz for Each Recording Time

Recording time	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	1810.83	1792.73	223.31	254.98
Postattachment	2120.77	2204.44	480.07	429.00
5 minutes post	2166.17	2003.35	552.06	369.24
10 minutes post	1969.20	2123.80	338.11	298.86
15 minutes post	1939.18	1963.18	503.17	276.65
20 minutes post	2104.77	2029.63	480.05	303.26
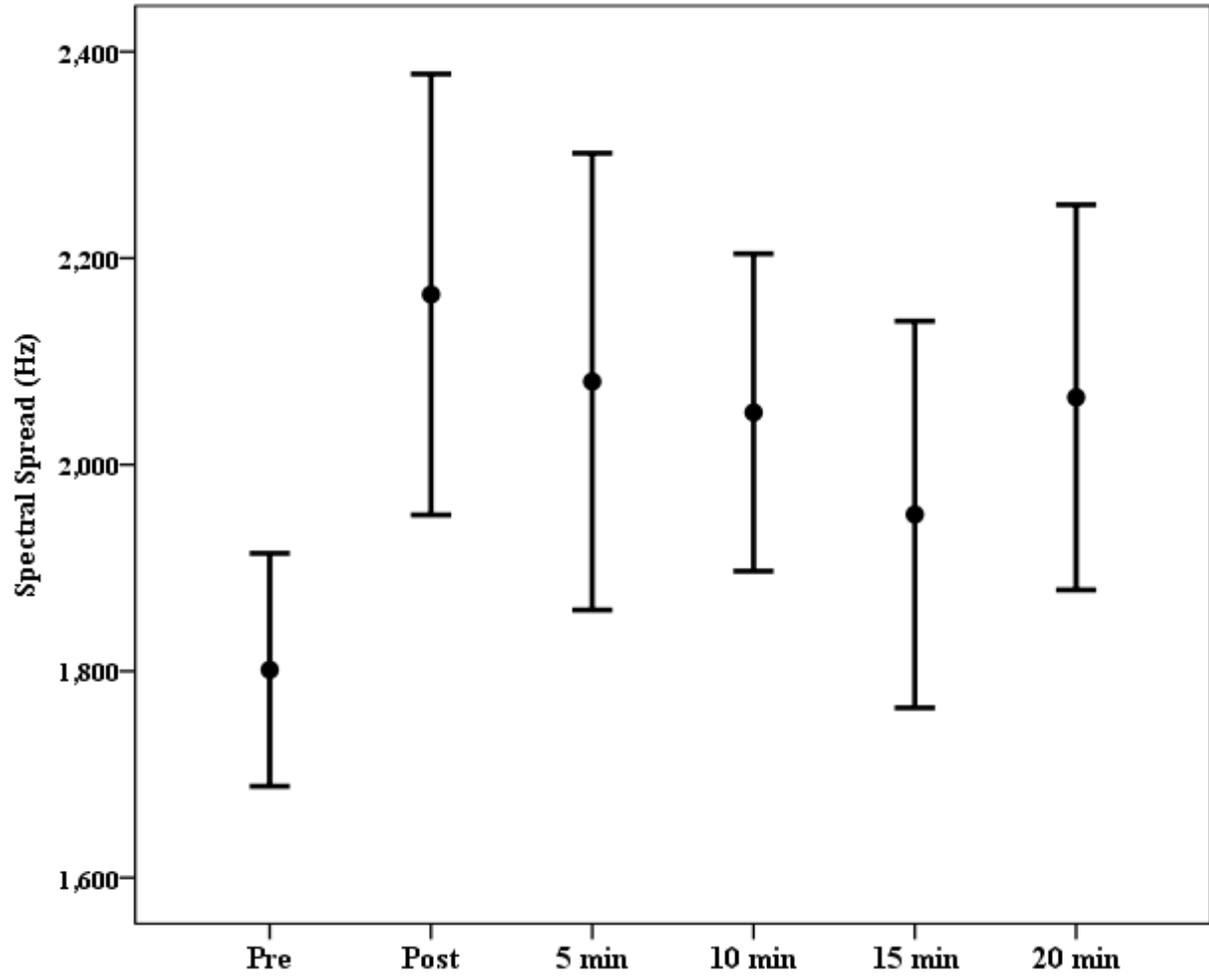
Figure 6. Spectral spread of /s/ in Hertz over time.

Table 9

Mean and Standard Deviation of /ʃ/ Spectral Center of Gravity in Hertz for Each Recording

Time

Recording time	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	6760.56	8482.30	837.54	906.20
Postattachment	5780.49	6188.22	1729.66	960.06
5 minutes post	5937.45	6757.46	1555.24	785.41
10 minutes post	6332.54	6536.41	795.31	638.28
15 minutes post	5697.53	6876.90	1343.33	615.51
20 minutes post	5941.11	6959.75	258.05	492.03
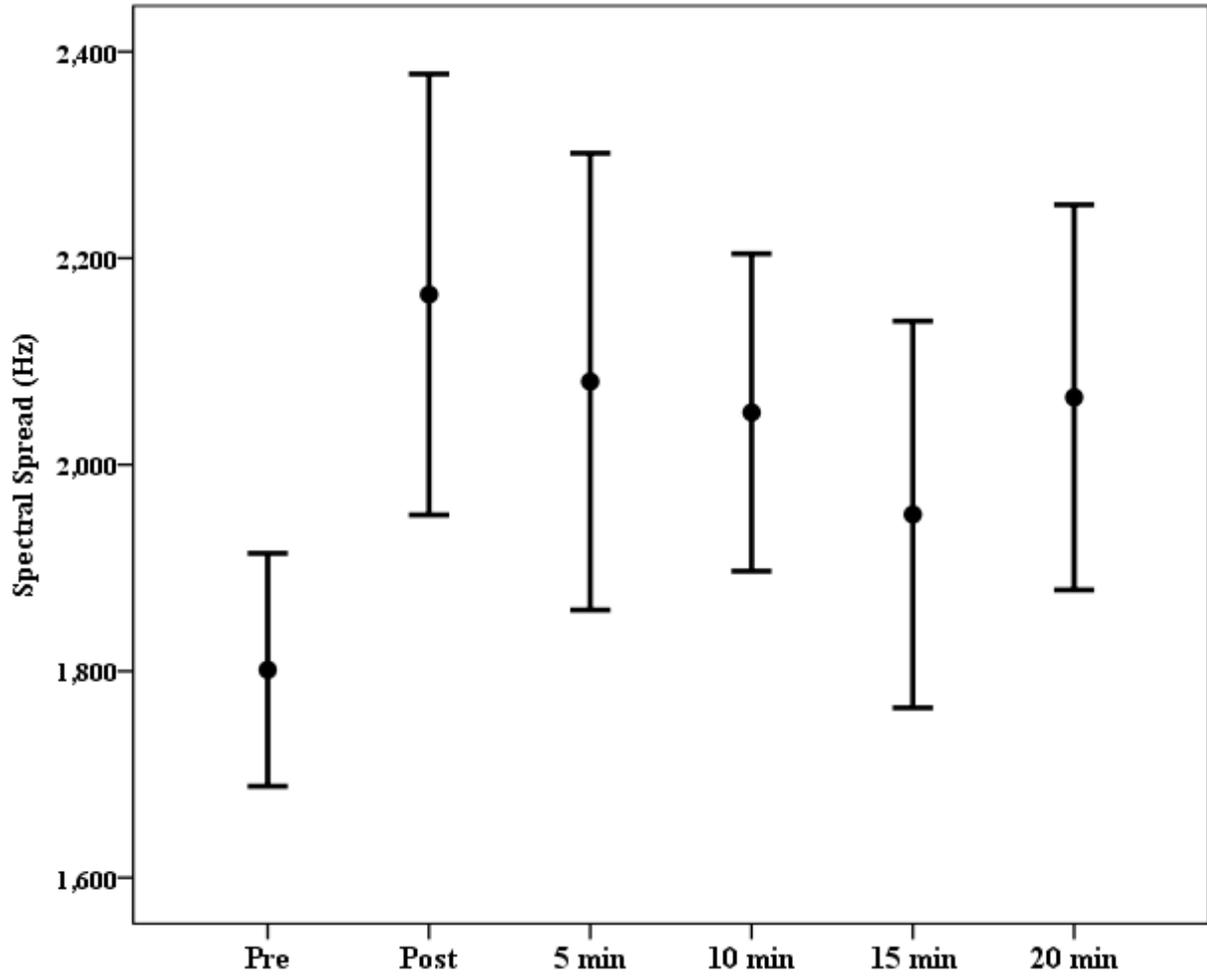
Figure 7. Spectral center of gravity of /ʃ/ in Hertz over time.

Table 10

Mean and Standard Deviation of /ʃ/ Spectral Spread in Hertz for Each Recording Time

Recording time	Mean		Standard Deviation (*SD*)	
	Male	Female	Male	Female
Preattachment	1861.93	1991.51	381.64	739.80
Postattachment	2172.13	2570.50	832.69	731.80
5 minutes post	2064.84	2464.16	525.74	686.66
10 minutes post	1770.58	2322.77	655.84	535.24
15 minutes post	2107.57	2132.44	737.67	423.94
20 minutes post	2029.71	2369.67	628.98	403.85
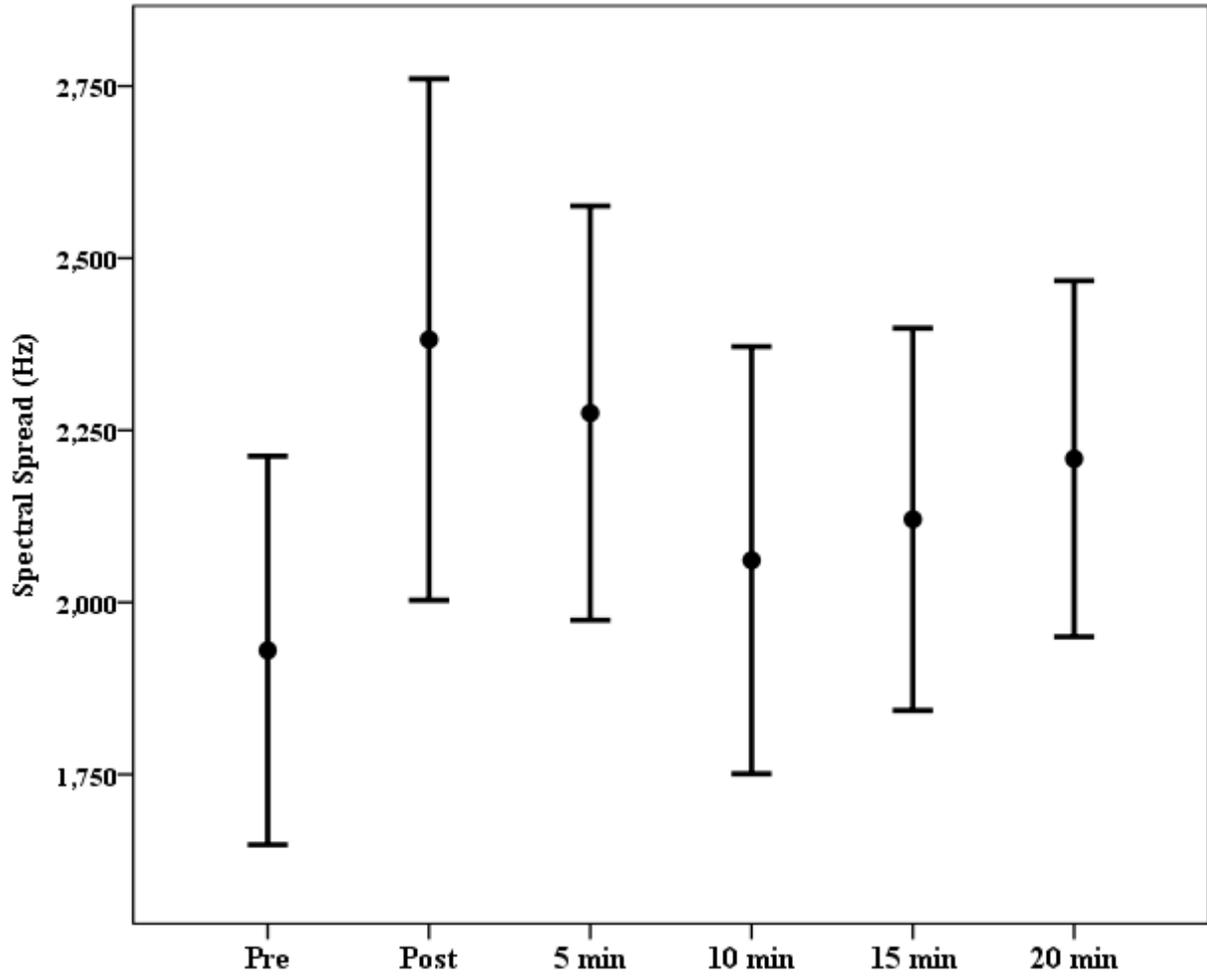
Figure 8. Spectral spread of /ʃ/ in Hertz over time.

Table 11

Results of Repeated Measures ANOVA for Duration, Spectral Center of Gravity and Spectral

Spread for /s/ and /ʃ/

Measure	d.f.	*F*-ratio	*p*-value	E.S.
Duration of sentence	4.708, 80.030	7.010	< .001	.292
Duration of /s/	3.943, 67.030	15.495	< .001	.477
Duration of /ʃ/	4.258, 72.387	5.173	< .001	.233
Spectral center of gravity of /s/	3.187, 54.172	1.248	.302	.068
Spectral spread of /s/	5.000, 85.000	4.520	.001	.210
Spectral center of gravity of /ʃ/	4.161, 70.744	6.589	< .001	.279
Spectral spread of /ʃ/	5.000, 85.000	2.073	.077	.109

Table 12

Results of Concurrent Contrasts. Recording Times for Each Measure are Compared with the Preattachment Recording

Measure	Post		5 min		10 min		15 min		20 min	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Duration of sentence 1	1.895	.187	1.689	.211	4.558	.048	6.532	.020	7.115	.016
Duration of /s/	1.305	.269	10.687	.005	4.790	.043	45.651	< .001	11.947	.003
Duration of /ʃ/	0.057	.815	5.681	.029	6.149	.024	1.226	.284	8.889	.008
Spectral Center of Gravity of /s/	4.984	.039	2.063	.169	6.479	.021	2.028	.173	3.059	.098
Spectral Spread of /s/	14.824	.001	9.912	.006	11.404	.004	3.932	.064	8.169	.011
Spectral Center of Gravity of /ʃ/	14.558	.001	9.345	.007	17.822	.001	17.924	.001	26.909	< .001
Spectral Spread of /ʃ/	4.903	.041	3.177	.093	0.501	.488	1.162	.296	2.442	.137

Discussion

The purpose of the present study was to examine the amount of time needed after the attachment of kinematic recording sensors for speakers to adapt and speak naturally. Results of both the perceptual and acoustic analysis revealed a change in speech precision over time, with all postattachment recordings receiving lower perceptual scores. Results from both analyses also revealed that precision beyond the 10-minute recording remained similar. Therefore, it can be concluded that participants reach a height of adaptation after 10 minutes of talking with kinematic recording sensors attached, and that after the attachment of sensors, speech production precision did not return to preattachment levels.

Perceptual Ratings

All postattachment perceptual ratings were lower than preattachment ratings. This is consistent with the findings from Aasland et al. (2006) study of the effects of static perturbation in the form of a pseudopalate. These authors also found that quality ratings following perturbation were considerably lower than preperturbed speech. An earlier research study by McFarland et al. (1996) also supported these results. Although bite blocks, rather than recording sensors, were used as a source of static perturbation, there were significant quality changes over time, indicating that speech adaptation had occurred. The perceptual results of the present study also revealed that speech precision changed little after the 10-minute recording. This is consistent with the findings of Aasland et al. (2006), who reported that recordings were given similar perceptual ratings 15 minutes after the introduction of the static perturbation.

Acoustic Measures

      Duration. The present study found that the durations of the entire sentence, of /s/ and of /ʃ/ changed significantly over time. Productions beyond the 10-minute recording were shorter than in the preattachment recording (with the duration of /ʃ/ at 20 minutes being the exception). These results are consistent with results from McFarland et al. study (1996) on adaptation to changes in the structure of the oral cavity using an artificial palate. Their study revealed significant changes over time, with shorter productions after insertion of the pseudopalate. However, it should be noted that McFarland et al. study did not specifically consider /s/ or /ʃ/, but rather many different vowels, consonants, and utterances. Shorter durations of the whole sentence, /s/, and /ʃ/ in the present study could be a result of practice and familiarity effects from speaking the stimuli repeatedly.

      Spectral measures. Spectral measure results revealed significant changes over time for the spectral center of gravity and spectral spread of /s/ and /ʃ/. Contrasts revealed that the spectral center of gravity and spectral spread for /s/ were lower after sensor attachment than in the preattachment recording. These results are consistent with previous studies which found lower center of gravity and spectral spread for the fricative /s/ after the introduction of a static perturbation (McFarland et al., 1996). Contrasts revealed that the spectral center of gravity and spectral spread for /ʃ/ were higher postattachment than in the preattachment recording. These results are also consistent with those of McFarland et al. (1996), who found higher center of gravity of /ʃ/ after the presentation of a static perturbation. However, a previous study (McAuliffe et al., 2007) reported different results for the spectral center of gravity for /ʃ/ (slightly lower after the presentation of a static perturbation). This however is most likely due to the use of a pseudopalate rather than sensor coils, which has a different structurally perturbing influence.

The present study also found a significant male/female difference, with women having a higher spectral center of gravity than men. This finding can be attributed to the slightly smaller oral cavity of women and thus smaller resonating space, which causes fricatives to resonate at a higher frequency.

Adaptation

The purpose of the current study was to discover the amount of time needed for a participant to reach a plateau of adaptation to the presence of recording sensors before the collection of kinematic data. The perceptual and acoustic measures both revealed that speech performance changed little after the 10-minute recording. For this reason, when using adhered kinematic recording sensors, we find no practical benefit for an adaptation time longer than 10 minutes prior to the collection of data. Previous studies regarding speech adaptation to static perturbation have found similar results. Aasland, et al. (2006) found that perceptual ratings and acoustic measures changed little after their 15-minute postperturbation recording time. However, there are other studies that collected data over much longer adaptation periods. Hamlet (1985) studied adaptation to a dental prosthesis over a period of two weeks. McAuliffe et al. (2007) examined speech adaptation to the presence of an EPG pseudopalate over a period of three hours. Aasland et al. also examined speech adaptation to an EPG palate, but over a period of one hour. However, none of these previous studies measured speech adaptation to the presence of articulograph recording sensors. Kinematic recording sensors do not typically remain in place as long as dental appliances and artificial palates. The dental adhesive used to adhere sensors to anatomy in the oral cavity is easily loosened by the movement of the structures in connected speech. This increases the likelihood of sensors falling off, thus making it valuable to know whether a shorter adaptation period would be sufficient. The findings of the current study

suggest that a 10-minute adaptation period prior to data collection may be sufficient, because the adaptation effects appear to plateau beyond this point.

It is also important to note that in respect to adaptation results, most speakers followed the same trend but one or two did not adapt well and one adapted remarkable quickly. It is understood that within a group, individuals will react differently. Other studies reported similar findings of individual outliers among participants (McFarland et al., 1996).

Adapted Speech versus Natural Speech

The current study revealed substantial speech changes following attachment of the sensors. Overall, perceptual ratings and acoustic measures did not return to their preattachment levels. This finding is consistent with McFarland and Baum's (1995) study on articulatory perturbation in which adaptation was not complete, meaning adaptation did not return speech to its preperturbed state. These results suggest that as a speaker becomes accustomed to a perturbation, what is produced is no longer natural speech, but rather adapted speech. The goal of using kinematic recording devices is to measure the movements of natural speech. However, as discussed previously, these devices necessarily perturb speech. As a consequence, the individual is no longer producing natural speech, but rather adapted speech. Even if results of the current study had revealed that speech returned to preattachment levels of perceptual and acoustic performance, the speech would still not have been natural, given the adaptation that would have allowed it to sound unperturbed. Although, the current technology is measuring at least partially adapted speech rather than natural speech, the information gained is still valuable to understanding the kinematic details of speech production.

Limitations of the Present Study and Directions for Future Research

The present study used one of the six sentences from a larger study for the analysis of speech adaptation over time. This sentence was designed to be sensitive to the presence of sensors at the tip and middle of the tongue, which is where the fricatives /s/ and /ʃ/ are formed. Thus, by using a stimulus sentence with a high proportion of the fricatives /s/ and /ʃ/, participants were more vulnerable to perturbation caused by the senor coils. One potential limitation of the present study is the possibility of the over-representation of the fricatives /s/ and /ʃ/ in the stimuli. Using a more phonetically balanced sentence could show less of an impact of perturbation on typical connected speech.

While the present study found significant adaptation to kinematic recording sensors, there is still a need for future research in this area. Studies using more than two individual kinematic recording sensors on the surface of the tongue could provide more information. The surface of the tongue is larger in comparison to other oral structures and there is room for the addition of more sensors. The tongue, of all oral speech structures, often contributes the most to speech production and is most easily perturbed by the presence of electromagnetic sensors. Therefore, additional studies with the presence of more sensors on the surface of the tongue and differing placements would provide valuable information about the adaptation of speakers to static perturbation. Additional sensors on the tongue could be placed on the midline, as was done in the present study, or could be placed closer to the edges. Each placement would provide different adaptation information. This study has provided valuable insights into speech adaptation to kinematic recording sensors and future research with additional sensors and different placements could expand current knowledge on the subject.

The present study examined the amount of time needed after the attachment of kinematic recording sensors for speakers to adapt and speak more naturally. Perceptual and acoustic analysis results revealed changes in speech precision over time, and also revealed that perceived precision beyond the 10-minute recording did not change. Therefore, it was concluded that participants reach a height of adaptation after 10 minutes of talking with kinematic recording sensors attached. These results offer insight into the complex patterns of speech adaptation. Furthermore, by providing researchers with a guide for the amount of time needed for adaptation before the collection of data, these results are of practical use for this technology in future research.

References

Aasland, W. A., Baum, S. R., & McFarland, D. H. (2006). Electropalatographic, acoustic, and

perceptual data on adaptation to a palatal perturbation. *Journal of the Acoustical Society*

of America, 119, 2372-2381. doi: 10.1121/1.2173520

Hamlet, S. L. (1985). Speech adaptation: An aerodynamic study of adults with a childhood

history of articulation defects. *Journal of Prosthetic Dentistry, 53*, 553-557. doi:

10.1016/0022-3913(85)90647-X

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production

revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of*

Neuroscience, 32, 9351-9358. doi: 10.1523/JNEUROSCI.0404-12.2012

McAuliffe, M. J., Robb, M. P., & Murdoch, B. E. (2007). Acoustic and perceptual analysis of

speech adaptation to an artificial palate. *Journal of Clinical Linguistics and Phonetics,*

21, 885-894. doi: 10.1080/02699200701576827

McFarland, D. H., & Baum, S. R. (1995). Incomplete compensation to articulatory perturbation.

Journal of the Acoustical Society of America, 97, 1865-1873. doi: 10.1121/1.412060

McFarland, D. H., Baum, S. R., & Chabot, C. (1996). Speech compensation to structural

modifications of the oral cavity. *Journal of the Acoustical Society of America, 100*, 1093-

1104. doi: 10.1121/1.416286

Nooteboom, S. G., & Doodeman, G. J. (1980). Production and perception of vowel length in

spoken sentences. *Journal of the Acoustical Society of America, 67*, 276-287. doi:

10.1121/1.383737

Perkell, Joseph. (2001). *The Sensorimotor Control of Speech Production: Model and Data*.

Cambridge, MIT.

Rees, N. S. (1972). The role of babbling in the child's acquisition of language. *British Journal of Disorders of Communication, 7*, 17-23. doi: 10.3109/13682827209011549

van der Merwe, A. (1997). A theroretical framework for the characterization of pathological speech sensorimotor control. In Malcolm R. McNeil (Ed.), *Clinical Management of Sensorimotor Speech Disorders* (pp. 1-25). New York, NY: Thieme.

Waldstein, R. S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America, 88*, 2099-2114. doi: 10.1121/1.400107

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin, 60*, 213-232. doi: 10.1037/h0044155

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience, 22*, 1770-1781. doi: 10.1162/jocn.2009.21324

APPENDIX A: Annotated Bibliography

Aasland, W. A., Baum, S. R., & McFarland, D. H. (2006). Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation. *Journal of the Acoustical Society of America, 119*, 2372-2381. doi: 10.1121/1.2173520

Objective: The purpose of this study was to observe how the speech production system compensates for a palatal perturbation caused by an electropalographic (EPG) appliance. *Methods:* The researchers collected electropalatographic, acoustic and perceptual data from nine young adult participants. They were native speakers of English with no history of speech or hearing disorders. Each participant was fitted with two custom EPG palates. One was thin (1mm thick) to mimic normal palatal structure, while the other was thick with a 6mm acrylic buildup on the alveolar ridge to perturb fricative production. The stimulus, /asa/, was used to obtain graphic, acoustic and perceptual data of medial /s/ production at all recording times. Ten repetitions of the stimuli were recorded without a pseudopalate in place to obtain a baseline for the acoustic and perceptual data. After the initial recording, ten repetitions of the stimuli were recorded with both thin and thick palates every 15 minutes for a period of one hour. At one hour additional recordings of two sets of ten repetitions of the stimuli were collected without any palate to measure any residual effects. The authors recorded all productions of /asa/ directly onto a computer using Articulate Assistant software. Acoustic analysis consisted of measuring medial /s/ duration, the mean of the spectral distribution and spectral standard deviation. EPG analysis was completed to determine whether compensatory tongue positioning occurred during perturbation of /s/ production. Three EPG measures were computed and compared for different palates and recording times. Whole contact was computed based on the percentage of EPG sensors contacted over time. Alveolar-palatal contact was computed as the percentage of sensors contacted in the five first anterior rows of electrodes. Variability index was used to reflect the "stability of articulatory gestures" by computing the percent frequency of activation over time. The separate perceptual experiment involved 10 native English participants who did not participate in the previous collection of data. Participants in the perceptual experiment rated recorded /s/ productions from five randomly selected participants from the previous experiment on a scale of 1 to 5 (1 being very poor quality and 5 being very good quality). Raters were familiarized with the scale by an example of a typical production which was taken from a pre-palate production of /asa/. All rating samples were presented randomly. *Results:* Acoustic analysis ANOVA results showed that after the insertion of either palate, the spectral mean was significantly lower than for the pre-palate recording. However, 15 minutes after palatal insertion this pattern was reversed and remained the same through the rest of the recording times. Post hoc analysis of the interaction of no palate, thin palate and thick palate showed significant differences in each condition over time. EPG analysis revealed relatively few changes in tongue-palate contact over time for all palate conditions. Results from the perceptual experiment showed that mean perceptual quality ratings for productions with the thick palate were considerably lower than those with the thin palate. Supporting acoustic results, these ratings were reversed after 15 minutes. After the hour of practice time, ratings for productions with both thin and thick palates were similar and higher than the initial recording after insertion. However, all ratings of productions with either palate

were lower than ratings for productions with no palate. *Conclusion:* Results showed the flexibility of the speech production system in its ability to adapt to oral-articulatory perturbation. There were few changes in tongue-palate contact over time, while acoustic and perceptual results indicated an increase in precision and quality over time. This suggests that compensations for /s/ were not made by tongue- palate contact changes but rather by changes in tongue shape. *Relevance to the current work*: This study provides evidence of speech adaptation over time to static perturbation based on both acoustic and perceptual data.

Hamlet, S. L. (1985). Speech adaptation: An aerodynamic study of adults with a childhood history of articulation defects. *Journal of Prosthetic Dentistry, 53*, 553-557. doi: 10.1016/0022-3913(85)90647-X

Objective: The purpose of this study was to determine if adults with a history of fluency and/or articulation deficits in childhood experience more difficulty with adapting to perturbed speech. *Method:* Three groups of subjects participated in the study: 9 subjects with no history of speech or language deficits, 5 former lispers, and 4 subjects with a history of /l/ or /r/ distortion. Each subject was fitted with an acrylic dental retainer with a 4mm thickness over the alveolar ridge. The dental prothesis was worn constantly by all subjects for a 2 week period. Three data recording sessions were held within that time period: before insertion of the retainer, after 1 day and after 2 weeks. Data were recorded using an audio signal and oral air flow was captured by a face mask with pneumotach and pressure transducer. During data collection, subjects read a randomized list of phrases that contained /s/ in vowel-consonant-vowel contexts. Five repetions of each phrase were recorded and analysed. *Results:* All subjects self reported the most difficulty with /s/ regardless of a previous disorder and none reported that the perturbation recreated previous speech deficits. Aerodynamic results of subjects with no history of deficits show air flow values of /s/ were lower at one day post insertion than at 2 weeks. Former lispers and subjects with previous articulation deficits did not show a significant difference in mean air flow between 1 day and two weeks post insertion. For all participants, mean air flow was greater before inserstion than at the two other recording times. *Conclusion*: Both subject groups with a hisotry of previous deficits showed evidence of slower adaptation and compensation for /s/. *Relevance to the current work:* The study examined speech adaptation and compenstation to static perturbation over an extended period of time.

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience, 32*, 9351-9358. doi: 10.1523/JNEUROSCI.0404-12.2012

Objective: The purpose of this study was to determine whether auditory or somatosensory feedback plays a larger role in the quality of speech production. *Methods:* 75 subjects between the ages of 18 and 40 participated in the study. All participants were native English speakers with no history of speech, language or hearing disorders. Custom acrylic and metal dental appliances were made for each participant for both the upper and lower teeth. The lower dental appliance was attatched to a robotic arm that tracked jaw movement and also applied forces to the jaw. The upper dental appliance connected the

upper jaw to two articulated arms which held the participants head motionless. Participants were seated in front of a computer which displayed either the word "head" or "had" for the subject to speak aloud. Participants were asked to speak the word continously at a comfortable pace and loudness, closing the mouth completely between each production until the word was removed from the screen. Researchers altered both auditory and somatosensory feedback, either simultaneously or alone, to find which influenced speech production more. Somatosensory feedback was altered using a robotic arm that changed the motion path of the jaw, while auditory feedback was altered by changing the frequency of the the first format of the vowel sound before playing back the audio through headphones. 14 subjects experienced somatosensory perturbation alone and 14 different subjects experience auditory perturbation alone. *Results:* Results of the study showed that the effects of the different perturbations were independento of each other. Somatosensory perturbation did not alter "the sound of the voice" and auditory perturbation did not alter the movment path of the jaw. Analysis also showed that applying the perturbations in combonation to subjects did not affect the amount of compensation. However, those subjects who compensated for somatosensory perturbation demonstrated less or no compensation to the auditory perturbation. Of the 75 subjects in the study, 53% adapted only to auditory perturbation, 26% adapted to both somatorsensory and auditory perturbations and 21% adapted to only somatosensory perturbation. *Conclusion:*These results indicate that while compensations were made by subjects for both auditory and somatosensory perturbations, each subject showed "sensory preference" to one or the other during speech motor learning. *Relevance to the current work:*This study provides evidence of the importance of feedback to adaptation and compensations during perturbed speech production.

McAuliffe, M. J., Robb, M. P., & Murdoch, B. E. (2007). Acoustic and perceptual analysis of speech adaptation to an artificial palate. *Journal of Clinical Linguistics and Phonetics, 21*, 885-894. doi: 10.1080/02699200701576827

Objective: The purpose of this study was to examine to process of speech adaptation to a standard electropalatographic practice palate. *Methods:* Eight female adult speakers with no history of speech or hearing deficits and who also were undergraduate speech pathology majors participated in the study. A custom EPG palate (containing no electrodes or lead wires) was made for each subject. Twelve CVC words embedded in a carrier phrase were used as stimuli and were repeated five times under four speaking conditions, resulting in each subject producing 60 CVC words. The four speaking conditions for data collection were normal speech (no palate), immediately post-insertion, 45 minutes post-insertion and 3 hours post-insertion Word-initial consonants used were /t/, /k/, /s/ and /ʃ/. Post-vocalic consonants used were /t/ or /p/. Vowels used were /i/, /a/ and /u/. CVC words were presented to subjects in random order across all four conditions. Subjects read a consonant-vowel-consonant phrase prior to, immediately following, 45 minutes post, and 3 hours after inserting the palate. Data were recorded using a head-mounted microphone at a sampling rate of 44kHz with a Sony Digital Audio Tape recorder in a sound treated room. Perceptual and acoustic analyses were performed on the initial consonant-vowel which closely examined consonants /t/, /k/ and /s/ and vowels /i/, /a/ and /u/. Seven seperate undergraduate speech pathology students served as

listeners for the perceptual evaluation. The third repetition of the five CVC productions of each subject was used for the perceptual evaluation. Listeners rated the CVC productions, given in randomized order, using a visual analog sliding scale of 0 (normal precision) through 10 (severe imprecision). Acoustic analysis was done using PRAAT software to compare sement durations, first and second vowel formant frequencies and consonant spectra. Of the perceptual samples, 25% were re-rated to test intra-rater reliability. *Results:* Repeated measures ANOVA revealed significant increase in the imprecision of consonants between pre-insertion and immediately post-insertion as well as 45 minutes post. The imprcision reduced significantly from 45 minutes post to 3 hours post. Perceptual ratings revealed that the best consonant precision occured during the pre-insertion condition. Spectral analysis results showed significant change across all conditions. *Conclusions:* Speech adaptation to an artificial palate does take place over time. Consonants, vowel formant frequencies, and vowel duration are relativiely unaffected by the insertion of an EPG palate. The greatest percieved imprecision was gound in word-initial consonant articulation after the insertion of the palate, but the impresision returned to normal following 45 minutes to 3 hours of adaptation time. *Relevence to the current work:* The study uses both acoustic and perceptual analysis to evaluate the process of speech adaptation over time. Perceptual analysis used a visual analog scale using the term "precision" to rate participant productions.

McFarland, D. H., & Baum, S. R. (1995). Incomplete compensation to articulatory perturbation. *Journal of the Acoustical Society of America, 97*, 1865-1873. doi: 10.1121/1.412060

Objective: The purpose of this study was to examine the role that auditory and somatosensory feedback play in compensation to perturbation during speech production. *Method:* Fifteen native female speakers of French with no history of communication disorders and who also passed an audiometric screening participated in the study. Stimuli for this study included three vowels produced in isolation [i a u], voiceless stop consonants [p t k] produced preceeding the vowels, and two voiceless friciatives [s s̠] under the same vowel condition as the voiceless stops. Each stimulus was produced 10 times by participants in random order under perturbed and unperturbed conditions. Perturbed conditions included the presence of a large bite block and the precence of a small bite block. Recordings of all stimuli under all conditions were analyzed with both temporal and spectral measures. *Conclusions:* Overall, results showed that compensations to a large or small bite block were not complete and never returned to their pre-perturbed production. Analysis also illustrated that compensations for vowels may have developed over time, while the effects of perturbation may last longer for consonants due to the articulatory precision needed for correct production. *Relevance to the current work:* This study examined the adaptation of speech over time using a bite block for static perturbation. The study also used spectral analysis to examine the extent and timeline of consonant adaptation.

McFarland, D. H., Baum, S. R., & Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *Journal of the Acoustical Society of America, 100*, 1093-1104. doi: 10.1121/1.416286

Objective: The purpose of this study was to examine speech production adaptations and compensations to modifications made to the physical anatomy of the oral cavity. *Method:* Fifteen native female speakers of French with no history of communication disorders and who also passed an audiometric screening participated in the study. Two artificial palates were made for each of the 15 participants by a dentist. The two palates differed in alveolar-palatal thickness. The first palate was 6 mm thick at the midline and the second was approximately 3 mm thick at the midline. Stimuli for the study included three vowels produced in isolation [i a u], voiceless stop consonants [p t k] produced preceeding the vowels and two voiceless fricitives [s ʃ] under the same vowel condition as the voiceless stops. Five repetitions were elicited from the participants in random order under perturbed and unperturbed conditions. These three experimental conditions were (1) no artificial palate, (2) thin artificial palate and (3) thick artificial palate. Data were collected using a digital audio tape recorder and directional microphone. Vowel and consonant durations, spectral mean and formant frequencies were determined by segmenting waveforms. Perceptual analysis was also conducted using 10 native French speakers with no history of speech or hearing disorders. Isolated vowel and consonant segments from data collection in all three conditions were used to create 6 perceptual tests (one for each phoneme and condition). Raters listened to consonants isolated from their vowel context to avoid contaminating effects. Stimuli were presented in random order to the listeners via headphones and they were then asked to rate the stimuli on a five-point scale (with the anchor words 'unintelligible' and 'perfect'). *Results:* There were significant changes in spectral mean for fricatives under perturbed contidions, with few changes in stop consonants and vowels. Perceptual data supported these patterns and gave evidence of improvements of speech production over time. Results were also compared to a similar previous study using bite blocks. *Conclusions:* The study found virtually no significant differences in acoustic and perceptual parameters of vowels under normal or palate conditions. Fricatives, on the other hand, were "highly susceptible" to the perturbing effects of an artificial palate. Perceptual ratings supported these results, with significantly lower ratings for fricatives under palatal conditions. Speech compensation for a subset of speech sounds appears to improve over time. The compensation to bite blocks and artificial palates differs in its effect on either vowels, consonants or fricatives due to differing effects on the jaw or alveolar ridge. *Relevance to the current work:* The study used both acoustic and perceptual analysis to examine speech adaptation to static perturbation in the form of an artificial palate.

Nooteboom, S. G., & Doodeman, G. J. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America, 67*, 276-287. doi: 10.1121/1.383737

Objective: The purpose of this study was to determine how vowel segment durations in spoken utterances are represented in auditory sensory storage. A secondary purpose was to study the degree to which syntactic and auditory context affect the identification of vowel length in the Dutch langauge. *Method:* Four male Dutch speakers in their early twenties with no history of speech or hearing disorders participated in the study. Two target words were used, one with a long Dutch vowel /a/, one with a short Dutch vowel /ɑ/ and each with the same initial and final consonants (/t/ and /k/). The two target words

were included in five different sentence types. Each participant read ten sentences, each with an embedded target word, first to themselves, and then aloud. Sentences were recorded on a tape recorder and converted to oscillograms using a visicorder. Ten male native Dutch speakers with no history of speech or hearing deficits participated in the perceptual portion of the study. Listeners were presented with stimuli through earphones in a quiet room. Listeners had two push buttons in front of them: one labeled with TAAK and the other TAK. The listeners were instructed to push the button of the vowel they heard, and when in doubt to guess. *Results:* A two-way analysis of variance revealed signifiant effects of sentence type on vowel duration of both /a/ and /ɑ/. Systematic differences in durations of both vowels in different syntactic posistions were found. Although the results did follow a pattern for vowel durations for syntactic structure, individual vowel durations were not predictable from syntactic sturcture alone. *Conclusions:* The authors drew two main conclusions from the results of the study. The first conclusion was that the accuracy of vowel durations in spoken sentences are "at least as good as would be predicted from most data on duration discrimination of "isolated nonspeech sounds with comparable durations". The second was that the internal criterion listeners use to distinguish between long vowels and short vowels based on vowel duration can be adjused to the auditory and phonetic structure of the surrounding speech in a sentence. *Relevence to the current work:* This study used perceptual listeners or raters to analyze the duration of vowel lengths.

Perkell, Joseph. (2001). The Sensorimotor Control of Speech Production : Model and Data. *Cambridge, MIT.*

> *Objective:* The purpose of this article was to outline a model of the sensorimotor control of speech production. The model has two main systems of speech production and speech perception. A complex visual mapping of both systems was presented to support the text. The model was designed to be supported by speaker data regarding brain function, physiology, anatomy, speech motor control, biomechanics and acoustics. The model explains that the goals for some articulatory movements for certain phonemes are contained in the auditroy-temporal domain. The model also outlines that speech movement planning is dependent on numerous mappings that are first obtained and then maintained using auditory feedback.

Rees, N. S. (1972). The role of babbling in the child's acquisition of language. *British Journal of Disorders of Communication, 7*, 17-23. doi: 10.3109/13682827209011549

> *Objective:* The purpose of this article was to discuss two major theories concerning the development of language in children. The theroies discussed were the learning theory and maturational approach. The article outlined the learning theory as an approach to language acquisition in which babbling is involved in the development of phonological abilities. In contrast, the maturational approach involves no important association between babbling and development of language. The article discussed commonly held views concerning the relationship between babbling and language acquisition. The first of these is the essential and important role of babbling in the development of "articulation skills". The child uses babbling as a practice for future speech production both by using

sensory and acoustic feedback. *Relevance to the current work:* This article outlines how the development of language incorporates both sensory and acoustic feedback through practicing.

van der Merwe, A. (1997). A therotical framework for the characterization of pathological speech sensorimotor control. In Malcolm R. McNeil (Ed.), *Clinical Management of Sensorimotor Speech Disorders* (pp. 1-25). New York: vThieme.

Objective: The purpose of this chapter was to discuss the role of sensorimotor control in speech production. Van der Merwe suggests that a knowledge of the complex intricaces of speech production is essential to the treatment of neurogenic speech disorders. Much of the framework is based on animal and human kinematic speech studies. The theoretical framework van der Merwe suggests is a transformation "of the speech code from one form to another as seen from a brain behavior perspective". This view of the framework places emphasis on the sensorimotor interface. In contrast to the past three-stage speech model of linguistic encoding, programming and execution, this framework presents and adds the motoric aspects of speech production. Van der Merwe discusses the importance of sensorimotor integration in speech motor control and the essential role it plays in speech coordination. Two different types of motor control, open and closed loop systems, were discussed. In an open-loop control system afferent input is relatively unimportant, while in a closed-loop control system afferent input is of utmost importance. Speech production is, according to most researchers, considered a open-loop motor control system. *Relevence to the current work:* This article outlines the importance of sensorimotor integration in speech motor control. It also defines open-loop versus closed-loop motor control systems.

Waldstein, R. S. (1990). Effects of postlingual deafness on speech production: Implications for the role of auditory feedback. *Journal of the Acoustical Society of America, 88*, 2099-2114. doi: 10.1121/1.400107

Objective: This study investigated the effects of postlingual deafness on the production of speech with the purpose of understanding the role of auditory feedback is the speech production system. To accomplish this objective, the study systematically characterized the effects of postlingual deafness on speech usuing spectral and temporal properties of consonants, vowels and suprasegmentals over time. *Method:* Seven postlingually deaf native speakers of English participated in the study. All participants had a hearing loss of 120 dB or greater and lost their hearing after the age of 5. All of the seven particpants also used speech regulary to communicate in natural life settings. Seven separate participants with normal hearing and matching ages were used as the control group. All participants recorded word and sentence productions focusing on consonants, vowels or suprasegmentals. Voice-onset time (VOT) was determined for consonant productions and the first two formant frequencies were calculated for vowel productions. Suprasegmentals were analyzed by obtaining intonation contours using a correlation algorithm. Mean fundamental frequency values and standard deviations of each sentence were also calculated. *Results:* Acoustic analysis of the participant productions resulted in significant results showing that postlingual deafness affects the production of all speech

sounds (consonants, vowels and suprasegmentals). Postlingually deaf participants exhibited shortened VOT in voiceless stop productions, a reduction in the formant frequency ranges used in vowel production, and longer overall sentence duration. *Conclusions:* These results suggest that auditory feedback affects "the monitoring and maintaining of phonetic precision in all classes of speech sounds". Although phonological distinctions remained generally intact for postlingually deaf participants, production was less precise. *Relevance to the current study:* Continual auditory feedback affects the overall precision of speech production over time.

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin, 60*, 213-232. doi: 10.1037/h0044155

> *Objective:* The purpose of this article is to discuss the effects of delayed auditory feedback on speech production. Delayed auditory feedback has been shown to affect speech production in multiple ways. Many subjects who speak with delayed auditroy feedback prolong vowels, repeat consonants, increase utterance intensity, and implement a range of other articulatory changes. Although many changes in speech production with delayed auditory feedback are similar, there are many differences from individual to individual. This is believed to be due to physiological differences. *Relevance to the current study:* This article discusses the role and influence of auditory feedback on speech production.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience, 22*, 1770-1781. doi: 10.1162/jocn.2009.21324

> *Objective:* The purpose of this study was to investigate the neural connection between motor commands and sensory feedback during speech production. *Method:* Twenty one men and women participated in the study. An fMRI was used to identify regions of the brain with activity during speech production. The collected fMRI information was further used to observe whether auditory feedback was located in the predicted area or not. The study also compared the real-time brain activity obtained during passive listening to results obtained during the active production of speech sounds. *Results:* There was an increase in activity in the superior temporal gyrus region during speech production when auditory feedback did not match the predicted auditory output. The superior temporal gyrus was also activated durig listening when speech was detected. *Conclusions:* These results suggest that the speech system relies on self-monitoring and feedback. It can then be assumed that the system is involved in controling articulatory planning. It also supports the idea that speech perception overlaps with self-monitoring in the superior temporal gyrus area. *Relevance to the current study:* This study revealed the importance of feedback during speech production in relation to neural connections and motor commands.

APPENDIX B: Informed Consent

Consent to be a Research Subject

....

Introduction
This research study on speech adaptation is being conducted by Professor Christopher Dromey and graduate student Elise Hunter at Brigham Young University to learn about speakers' adaptation to sensors attached in and around the mouth. You were invited to participate because you are a native English speaker with no history of speech or hearing disorders.

Procedures
If you agree to participate in this research study, the following will occur:

- you will sit in a sound booth at a computer station
- you will listen over loudspeakers and will be able to adjust the loudness to a comfortable level for you
- you will listen to a series of sentences and rate each one on a sliding scale to rate speech clarity
- examples of imprecise and precise sentences will be given to familiarize you with the range of sentences that will be heard
- you will be able to change ratings of previous samples if you want to
- you will listen and rate sentences for approximately fifteen minutes, but no more than 30 minutes

Risks/Discomforts
Possible mild fatigue.

The researchers will minimize risks of fatigue by allowing you to work at your own pace and take breaks if needed.

Benefits
There will be no direct benefits to you. It is hoped, however, that through your participation researchers may learn about the process of speech adaptation when recording sensors in and around the mouth are used to measure speech.

Confidentiality
The research data will be kept on a password protected computer and only the researcher will have access to the data. At the conclusion of the study, all identifying information will be removed and the data will be kept in the researcher's locked office.

Compensation
You will receive $5 for your participation; compensation will not be prorated.

Participation

Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely.

Questions about the Research

If you have questions regarding this study, you may contact Elise Hunter at elisehhunter@gmail.com or Christopher Dromey at 801-422-6461 or dromey@byu.edu for further information.

Questions about Your Rights as Research Participants

If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Name (Printed):_____Signature _____Date: _____