Critical Review and Experimental Study:
Exploring the Temporal Characteristics of the Vowels Associated with Intelligible Alaryngeal Speech*

Lauren Perduk
M.Cl.Sc. (SLP) Candidate
University of Western Ontario: School of Communication Sciences and Disorders

This report presents findings from a critical review of literature and the results of a study examining the temporal characteristics of the vowels produced by intelligible alaryngeal speakers. The critical review includes four between groups studies, two of which compared esophageal speech (ES) to laryngeal (normal) for vowel duration and vocal onset time respectively, one of which compared tracheoesophageal (TE), ES, and laryngeal speech for both vowel duration and vocal onset time (VOT), and one which examined differences in vowel durations in ES speech and how they relate to acceptability and intelligibility ratings. The study examined altering vowel durations in TE speech to determine if altered length would result in perceptual changes. This involved the transcription of TE speech samples by listeners (n=12). Findings from the critical review suggest differences in timing measures as related to different speech modes. Overall, more intelligible ES speech was associated with longer vowel durations than laryngeal speech and the shortest VOTs. The TE speakers consistently had the longest vowel durations, but also displayed longer, closer to normal VOTs. It was determined that the most skilled alaryngeal speakers are able to manipulate timing to become closer to that of laryngeal speech. The study found that shortening the vowels of individual speakers to more closely approximate lengths of laryngeal speakers resulted in increased intelligibility of voiceless stops /t/ and /k/.

Introduction

Individuals who undergo total laryngectomy require the use of alternative modes of voice production. The most commonly used methods are esophageal speech (ES), electrolarynx, and tracheoesophageal speech (TE).

One factor affecting intelligibility in the alaryngeal population is that the voiced- voiceless distinction is difficult for these speakers and voiceless consonants in alaryngeal speech are often perceived as voiced (Doyle, Danhauer, and Reed, 1988). This distinction is often made by differentiating VOT for the following vowel, thereby also altering vowel duration.

It is known that ES speakers are able to effect systematic and linguistically appropriate variation in the timing of vocal onset, showing that vowel duration, along with VOT, may a be manipulable factors for alaryngeal speakers (Christensen, Weinberg, and Alfonso, 1978). These vowels, when immediately following stop consonants may affect consonant perception, because characteristics of stops are often modified by co-articulation with the following vowel (Baken, 1987). Studies have examined both differences in vowel durations and VOT in these speech modes as compared to laryngeal speakers as well as intelligibility of these modes. (Doyle, Danhauser, & Reed, 1988; Christensen & Alfonso, 1976; Christensen, Weinburg, & Alfonso, 1978; Robbins, Christensen, & Kemper, 1986; Swisher, 1980).

Currently, little is known regarding the direct effect of manipulating these temporal factors on the resulting intelligibility of the consonants surrounding the vowel, particularly in TE speech. Previous studies have shown common characteristics surrounding the vowels in alaryngeal speech modes when only intelligible samples are considered (Christensen & Alfonso, 1976; Christensen, Weinburg, & Alfonso, 1978; Robbins, Christensen, & Kemper, 1986; Swisher, 1980) and are discussed further within this review.

Clinically, this information is important because of the lack of research that exists in this population directly related to methods to improve intelligibility in their speech. These results may help to provide guidelines to create targets for temporal features of vowels during voice therapy following laryngectomy. Additionally, increased intelligibility in alaryngeal speech is directly related to increased acceptability ratings (Swisher,

**This paper was created as a required assignment for the CSD9639 Evidence Based Practice for Clinicians course at Western. While it has been evaluated by course instructors for elements of accuracy and style, it has not undergone formal peer-review.*

1980). This research will provide information regarding factors that may improve voice-related quality of life by affecting acceptability of speech.

Objectives

The purpose of this paper sought to critically review the existing background literature related to temporal measurements of the vowels in both TE and ES speech modes in the alaryngeal population. Specifically, it examines VOT and vowel durations and how they present in intelligible speakers. With this research as a foundation, the next objective was to perform an experimental study where naïve listeners transcribed TE speech samples that included manipulations of vowel durations.

Ultimately, the goal was to determine which temporal characteristics were associated with the most intelligible alaryngeal speech and how these measures may be clinically relevant in the practice of Speech-Language Pathology with those who have undergone laryngectomy.

Study 1: Critical Review

Methods

Search Strategy

Computerized databases including PubMed, PsychINFO, and ASHA publications were searched using the following search strategy: [((alaryngeal) OR (tracheoesophageal) OR (esophageal)) AND ((vowel) OR (vocal onset))]. Reference lists accompanying several of the selected articles were also used to obtain additional literature.

Selection Criteria

Studies included in this review were required to include measurements of at least one temporal aspect (VOT or vowel duration) compared across a number of TE or ES speakers. Articles also needed to include a measure of intelligibility (either rating or exclusion criteria) as judged by transcriptions of speech samples.

Data Collection

The literature search yielded four peer-reviewed articles. These articles were all between groups studies including; Christensen & Weinburg (1976); Christensen, Weinburg, & Alfonso (1978); Robbins, Christensen, & Kemper (1986) and Swisher (1980).

The between groups study design was used to allow direct comparison of the alaryngeal modes to each other as well as to laryngeal speakers on various temporal measures. All studies contained a laryngeal group as a control and samples were randomized within the groups. This provides a higher level of evidence and still allows the comparison of alaryngeal speakers as compared to laryngeal.

Results

Christensen & Weinburg (1976) compared the vowel durations of a group to adult male ES speakers (n=10) to a group of adult male laryngeal speakers (n=9). All ES speakers were rated as having above-average to excellent ES speech (Weinburg & Bennett, 1972). All speakers were recorded saying 32 symmetrical CVC syllables in a sentence frame. Consonants /p/, /t/, /k/, /b/, /d/, /g/, /s/, and /z/ and vowels /i/, /ɪ/, /ɑ/, and /u/ were transcribed by listeners with experience transcribing. An ANOVA revealed increased vowel durations in intelligible samples for ES speakers, which was due exclusively to longer vowels in voiceless consonant environments compared to laryngeal. Comparisons made within the ES group found that the ES speakers' vowel durations were significantly longer in voiced consonant environments as compared to voiceless environments.

Although this study had a small number of participants, the stimuli were recorded repeatedly, which allowed the extraction of a larger number of speech samples (ES =1660; laryngeal =1440). This study also explored a wide range of articulatory positions for vowels. Limitations included that there was no detail providing regarding demographics including age or other voice related health concerns of the control group.

In order to rate the ES speakers as being above-average to excellent, their speech was compared to highly rated speakers from a previous study done by the same authors. This rating type may have resulted in listener bias as they were judged only by the authors. However, broad transcription from 10 experienced listeners of these samples provided additional confirmation of speech intelligibility. The criteria used for acceptance of speech samples (80% intelligible vowel along with 80% intelligible consonants) was highly restrictive and was confirmed with an additional agreement task, yielding a high, 95% inter-rater reliability. Criteria to measure vowel durations was well defined for both voiceless and voiced consonant environments and were taken from a previous study to ensure validity (Peterson & Lehiste, 1960). These measurements were confirmed by the investigator who re-measured the vowels and found no significant differences between measurements, showing high test-retest reliability of these measurements.

Comparisons were made of the vowel durations of the two groups as well as for each individual vowel using appropriate statistics (ANOVA). Overall, this study provides a compelling evidence that vowel durations of intelligible ES speech are significantly longer than those of laryngeal speech, but these differences are only observed in voiceless consonant environments.

Christensen, Weinberg, & Alfonso (1978) conducted further analysis of speech samples from an earlier study (Christensen & Weinburg, 1976) to obtain measurements of VOT in order to determine whether ES speakers could systematically vary the VOT of different speech sounds with the same manner of production. Only speech samples previously deemed to have acceptable intelligibility were analyzed to determine VOT. The VOT measurements were based on previously validated guidelines (Lisker & Abramson, 1967). The researchers performed an ANOVA to determine that the VOT values associated with voiceless stops had much shorter lag intervals than that of laryngeal speech. No differences were observed for voiced stops. There was also a significant increase in VOT as placement moved from labial to velar, which is consistent with the pattern of laryngeal speech.

For the same reasons previously identified, the methods, participants, and designs showed strong validity and reliability with the exception of some information lacking regarding the control group. Guidelines to measure VOT were well defined to ensure consistency. Appropriately, ANOVA was used to compare these measurements.

Overall, the results of this study provide compelling evidence. The findings are consistent with those of the previous research when considering that the detection of the earlier onset of voicing in the voiceless consonants may serve in part to account for the lengthening of the vowels in these same samples.

Robbins, Christensen, & Kemper (1986) examined both vowel duration and VOT in the speech of TE, ES, and laryngeal speakers using three aged-matched group (n=15 per group). Three CVC words were selected as stimuli and contained different vowels, as well as voiceless stops in initial and final position (/kup/, /kɑp/, and /pik/). Each participant read each of the three stimuli within a carrier phrase three times. Listeners used a forced choice to make judgments regarding the intelligibility of each of the samples. Only samples deemed intelligible were analyzed for vowel duration or VOT. The study found main effects for both group and vowel with TE speech having the longest durations on all vowels, (followed by ES and laryngeal). Laryngeal speakers had the longest VOTs, (followed by TE and ES).

Participant groups were aged-matched with well-defined characteristics for inclusion. All TE speakers had the same prosthesis and had received no previous voice therapy. All ES participants had undergone speech therapy following laryngectomy and all laryngeal speakers had no previous history of voice disorder. These eligibility considerations assured no obvious advantages to particular speakers within groups.

The use of three different vowels allowed for examination of a range of articulatory positions, and although comparisons could not be made between voiced and voiceless samples, the consonant environments were varied to allow articulatory movement from front to back and back to front. However, because this direction was only examined in one direction for each vowel, the effects of the vowel on the results of this directional change were not controlled.

The use of practicing SLPs as listeners ensured familiarity with speech judgments, however the use of the forced choice for transcription restricted the errors that the listener may have heard. This may not have been as reliable of an intelligibility measure as a broad transcription task. The number of repetitions of stimuli allowed for the production of a large total number of stimuli (n=135) from each group. To ensure intelligibility of the samples, only words with 80% of the vowels identified correctly were analyzed for vowel durations and likewise for consonant identification and VOT. Although this resulted in some differences in the samples analyzed for each measure, the intelligibility measures displayed face validity in targeting only the part of each word that was to be measured. Inter-rater reliability of these judgments was high and reported to be over 85%.

Appropriate statistics (A split-plot factorial ANOVA) were used to allow comparison of both between groups (speaker type) and within groups (vowels) variables. Overall, this study provides compelling evidence that alaryngeal speakers retain the laryngeal patterns for VOT as they pertain to placement. The use of previously validated measures for temporal characteristics and high inter-rater reliability ensured the data was comparable across other studies, although the transcription task format may have overestimated intelligibility measures.

Swisher (1980) compared the phonation time of the vowels in ES speech to laryngeal speakers and also to

ratings of the speech samples. This study also included a measure of oral pressures that will not be discussed in detail. ES speakers (*n*=10) and age matched laryngeal controls (*n*=10) were used as participants. Four stop consonants (/p/, /b/, /t/, /d/) and two fricatives (/s/, /f/) were combined with vowels /ɪ/ and /ɑ/ and final /p/ to create CVC stimuli which were recorded in a carrier phrase. Each speaker recorded each phrase 4 times, along with the first 3 sentences of the Rainbow Passage. Twenty-five listeners rated the Rainbow Passage sentence on a 5 point scale of acceptability and transcribed the CVC stimuli. These ratings and transcriptions showed a positive correlation. Higher intelligibility also strongly correlated to shorter vowel durations for both vowels in the ES group.

The participant groups were aged-matched and had well-defined characteristics for inclusion, stating that all ES speakers had completed speech therapy and all laryngeal speakers had no previous voice problems. All listeners had completing training in aspects of speech and voice therapy and were experienced with speech transcription. These raters displayed high inter-rater reliabilities of over 0.80 on measures of acceptability and intelligibility. Although acceptability measured on a 5 point scale is somewhat subjective, the reported correlation to intelligibility scores (0.35) gave a moderate level of validity to this measure.

Averages of the second and third repetitions for each phrase were analyzed for durations. This provided a measure which could better account for individual variation in duration. The measures of vowel duration were also well defined. Durational measurements had high test-retest reliability over a 27 day span (0.89).

Appropriate parametric and non-parametric statistics (Spearman Rank Order and Pearson Product Moment correlations) were used to identify correlations between acceptability and intelligibility as well as between vowel duration and intelligibility (-0.95 for /ɑ/ and -0.79 for / ɪ/) in ES.

Overall this study provides compelling evidence for correlations between shorter vowel durations and more intelligible speech. The consistent final /p/ in the CVC structure may have affected durations in some environments, but otherwise the study displayed moderately high levels of validity and reliability.

Discussion

The measured temporal characteristics of alaryngeal speech displayed consistent patterns as they relate to speaker type, consonant environment, and vowel position. All measurements for vowel duration and VOT had been previously validated and the guidelines provided consistency of analysis between all four reviewed studies (Peterson & Lahiste, 1960; Lisker & Abramson, 1967)

Taken together, findings of the reviewed studies suggest that ES speech displays longer overall vowel durations than laryngeal, but this difference only exists due to increased vowel durations in voiceless consonant environments. By increasing the distinction between the durations in voiced and voiceless environments, ES speakers were also able to increase intelligibility. Consistently, ES speakers also displayed the shortest VOT lag intervals and earlier onset of voicing than laryngeal speakers.

Upon integration of TE speech, it was found that both ES and TE speakers followed the normal pattern of increasing VOT as position changed from front to back, however laryngeal speakers displayed consistently longer VOTs than alaryngeal speakers. TE speakers were found to have the longest vowel durations of all three groups, but also had closer to normal VOTs than the ES group. TE speakers also produced more intelligible samples overall then the ES group when restrictions were put on inclusion based on intelligibility of samples.

Overall, findings suggest that as temporal measures of vowels more closely approximate those of laryngeal speech, intelligibility increases. Manipulation of temporal aspect of the vowels in TE speech and how they relate to intelligibility are further explored in the study to follow.

Study 2: Pilot Study

TE speech is pulmonary driven, making it most similar to normal laryngeal voicing in this regard (Maruthy, Mallet, and Bellur, 2014). Throughout previous research there have been various comparisons of characteristics of ES, TE and laryngeal speakers. This research has shown evidence of laryngeal speakers being rated as most intelligible with TE speakers being rated as significantly more intelligible than ES; including when speakers switch from ES to TE speech mode (Doyle, Danhauer, and Reed, 1998; Law, Ma, and Yiu, 2009; Bridges, 1991). This same intelligibility difference was also significant when isolating only stop consonants. (Doyle, Danhauer, and Reed, 1988). Despite higher intelligibility in TE speech, the voiced, voiceless distinction is difficult, and voiceless consonants in TE speech are often perceived as voiced (Doyle, Danhauer, and Reed,

1988). It is known that vowels immediately following stop consonants may affect consonant perception (Baken, 1987).

The following study aims to examine whether manipulations of vowel duration have an effect on the intelligibility ratings of TE speakers. More specifically, vowels will be presented in stop consonant environments to examine differences in the perception of the initial stop consonant as a result of vowel duration. This study will allow further understanding of a) the factors which affect intelligibility in the TE speech mode, b) how vowel duration plays a role in initial stop consonant perception, and c) if the intelligibility of certain stop consonants is more susceptible to differences in vowel durations (including voiced-voiceless distinction) in TE speech.

Methods

Speaker Samples

High quality audio recordings of CVC words containing all initial stop consonants and the same vowel (æ) from 10 male TE speakers were selected from a library of TE speech samples.

Listeners

Ten undergraduate and graduate students with experience in International Phonetic Alphabet (IPA) transcription were recruited as listeners in the study. All of the listeners were considered naïve to voice disorders and alaryngeal speech as they did not have any formal exposure or education in this area. Listeners were native English speakers and had no history of hearing concerns.

Speech Stimuli

Speakers provided recordings of 66 CVC words (Weiss & Basili, 1985). Of these, 6 stimuli were extracted to create a list including one speech sample for each of the initial stop consonants from each of the 10 speakers. The vowel /æ/ was held consistent. Each syllable also had a final stop consonant, with the exception of /kætʃ/.

These isolated samples were transferred to a personal computer and saved as WAV files using the acoustic software *Audacity*. The duration of the vowel in each sample was measured by a research assistant with training in acoustic measures of speech. These measurements were confirmed by the investigator. Each sample was then subjected to two manipulations of vowel duration; one which increased vowel duration by 20% and one which decreased vowel duration by 20%.

Samples were randomized within each group and 3 randomized lists of 60 stimuli were created for each of the manipulations for a total of 9 lists. Six stimuli were repeated at the end of each list as an internal validity measure.

Procedures

Each participant completed the task in one single session lasting approximately 45 minutes. When listeners arrived for individual listening sessions they were informed that they would be transcribing recordings of abnormal voice samples. Listeners were told they would navigate three separate 66 item lists of CVC syllables for a total of 198 speech samples and were presented with one list from each of the manipulations. They were instructed to transcribe each of the samples using the IPA. Listeners were allowed to listen to the samples as many times as they needed to be confident in their transcription of the samples, but not to go back to a sample once a judgment had been made. Participants transcribed these samples onto three separate sheets, each of which was numbered (to correlate to the numbers of the samples). Participants were not given any further information about the vowels or consonants included. The manipulations of the stimuli were not revealed.

The presentations of the stimuli were randomized both within and between manipulation type lists (the order of manipulation type was varied as well as the order of the stimuli within each manipulation type).

Reliability

The final 6 items (10%) of each list were duplicated from the first 60 items. Transcriptions of these items were compared to the transcription of the same items within the list in order to measure internal validity. Correlations between the initial consonants of these items were 0.92.

Results

Independent samples t-tests found a significant difference in the scores for short vowels /t/ (M=4.67, SD=1.67) and long vowel /t/ (M=3.08, SD=1.88) conditions; $t(22)=0.829$, $p = 0.040$. A significant difference was also found for short vowel /k/ (M=7.17, SD=2.12) and long vowel /k/ (M=5.25, SD=1.81) conditions; $t(22)=2.37$, $p = 0.027$. These results indicated greater listener ability to distinguish initial /t/ and /k/ when these sounds preceded vowels of shorter durations.

Results also showed a general trend toward greater

intelligibility of voiceless stops as position moved from front to back (Figure 1).

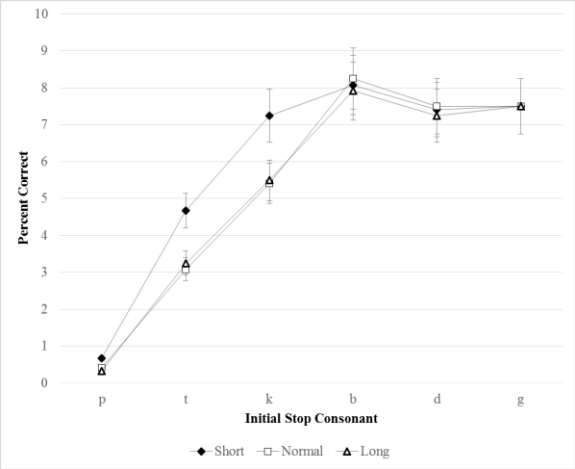

Figure 1: Average correct responses for each initial stop consonant by vowel duration.

Discussion

The results of the independent study serve to further reinforce the findings that as temporal measures come closer to approximating those of laryngeal speech, intelligibility increases. In voiceless consonant environments (for /t/ and /k/) shortening (and subsequently approximating laryngeal speech values) the duration had a positive effect on the listeners' correct perception of these sounds.

Additionally, this study found an overall increase in intelligibility of the voiceless consonant sounds as placement moved from front to back. This finding may relate to results from the reviewed studies which found a pattern of increased VOT, and closer approximations to laryngeal, as the vowels moved from front to back

Clinical Implications

Literature Review
Overall, it was consistently found that speech samples which included vowel durations and VOTs which more closely approximated to laryngeal speakers were rated as more intelligible.

Clinically, it is hypothesized that as ES becomes more skillful, the manipulation of the articulators can be used to maximize air supply and allow production of consonants and vowels that are closer to normal. Particularly the distinction between voiced and voiceless phonemes is of importance in alaryngeal speech (Doyle, Danhauer, and Reed, 1988). This information may provide increased detail pertaining to a method of training ES speakers in creating the voiced-voiceless distinction to increase overall intelligibility.

Independent Study
Similar patterns were observed in the TE population, indicating that a similar voiced-voiceless distinction can be made through manipulation of vowel duration.

Alaryngeal speakers may have increased intelligibility if they are instructed to initiate and terminate phonation at the proper time and therefore decrease the vowel phonation time in voiceless stop environments. In addition, higher ratings of intelligibility may correlate to higher acceptability. This increase has potential implications related to the voice related quality of life for alaryngeal speakers.

Acknowledgements

The experimental study was supported by the Voice Production & Perception Laboratory, Western University. Special thanks to Philip C. Doyle, PhD, CCC-SLP for the provision of speaker data

References

Baken, R. (1987). *Clinical Measurement of Speech and Voice*. Virginia: College-Hill Press.

Bridges, A. (1991). Acceptability ratings and intelligibility scores of alaryngeal speakers by three listener groups. *International Journal of Language & Communication Disorders, 26*(3), 325-335.

Christensen, J.M., & Alfonso, P.J. (1976). Vowel duration characteristics of esophageal speech. *Journal of Speech and Hearing Research, 19,* 678-689.

Christensen, J.M., Weinberg, B., & Alfonso, P.J. (1978). Productive voice onset time characteristics of esophageal speech. *Journal of Speech and Hearing Research, 21,*56-62.

Doyle, P.C., Danhauser J.L., & Reed C.G. (1988). Listeners' perceptions of consonants produced by esophageal and trachesophageal talkers. *Journal of Speech and Hearing Disorders, 53,* 400-407.

Law, I. K. Y., Ma, E. P. M., & Yiu, E. M. L. (2009). Speech intelligibility, acceptability, and communication-related quality of life in

Chinese alaryngeal speakers. *Archives of Otolaryngology–Head & Neck Surgery*, *135*(7), 704-711.

Lisker, L., & Abramson, A. S., (1967). Some effects of context on voice onset time in English stops. *Language Speech, 10,* 1-28.

Maruthy, S., Mallet, M.K., & Bellur, R. (2014). Comparison of esophageal and tracheoesophageal speech modes in dual mode alaryngeal speakers. *Journal of Laryngology and Voice, 4,* 6-11.

Peterson, G.E., & Lehiste, J. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32,* 693-703.

Robbins, J., Christensen, J., & Kemper, G. (1986) Characteristics of speech production after tracheoesophageal puncture: Voice onset time and vowel duration. *Journal of Speech and Hearing Research, 29,* 499-504.

Swisher, W.E. (1980). Oral pressures, vowel durations, and acceptability ratings of esophageal speakers. *Journal of Communication Disorders, 13,* 171-181.

Weiss, M.S., & Basili, A.G., (1985). Electrolaryngeal speech produced by laryngectomized subjects: Perceptual characteristics. *Journal of Speech and Hearing Research, 28,* 294 300.