Critical Review: Does Fundamental Frequency (F_0) Modulation Positively Impact Auditory-Perceptual Features of Electrolaryngeal (EL) Speech?

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

Abstract

This study presents a critical review of research examining the effect of F0 modulation on auditory-perceptual features of EL speech. Included is a critical evaluation of five articles pertaining to various modulations of F0. Overall, the results of this paper suggest that F0 modulation can positively impact auditory-perceptual features of EL speech.

Introduction

In 2013, there were an estimated 89,081 people living with laryngeal cancer in the United States alone ("Cancer Stat Facts", 2014). For some individuals with laryngeal cancer, surgical treatment is recommended. This results in the removal of part or all of the larynx, including the vocal cords. A number of therapies and devices may be used to provide alternative means of spoken output for affected individuals. Such methods of alaryngeal speech include: Tracheoesophageal speech, esophageal speech, and the use of an electrolarynx (EL). The EL provides an external vibratory source in order to achieve alaryngeal voicing utilizing the vocal tract. Research has suggested that following total laryngectomy, many people rely on the EL as a source of alaryngeal voicing (Hillman, Walsh, Wolf, Fisher, & Hong, 1998). Despite the fact that many people rely on this mode of voicing, the EL has been judged as the poorest alaryngeal method for producing intelligible speech in relation to other techniques (Williams & Watson, 1985).

In order to fully understand the quality of EL speech, a number of auditory-perceptual outcome measures such as acceptability and naturalness have been used in research to highlight deviations in EL speech from that of laryngeal voicing (Bennett & Weinberg, 1973). Understanding how F0 modulation impacts auditory-perceptual measures is vital for improving EL speakers' verbal communication. It is therefore the goal of this critical analysis to determine whether such modulations could positively impact auditory-perceptual ratings of EL speech.

Objectives

This critical review analyzes studies that have included F0 modulations and their influence on auditory-perceptual measures of EL speech as rated by naïve listeners.

Methods

Search Strategy

Online databases (PubMed, Western Libraries, and Google Scholar) were searched using the following terms:

[(fundamental frequency) OR F0 OR (pitch) AND (acceptability) OR (intelligibility) OR (naturalness) OR (VRQOL) OR (auditory-perceptual features) OR (auditory-perceptual measures) AND (electrolaryngeal speech) OR (EL speech) OR (artificial larynx)].

Selection Criteria

Studies included examined the effects of F0 modulation on EL speech, and were required to include one of the following auditory-perceptual measures: naturalness, intelligibility, or acceptability. No restrictions were chosen for age or languages spoken by the participants recruited for the studies.

Data Collection

Papers analyzed in this review include two mixed factors designs (level 2a evidence), one within-groups repeated measures study (level 2a evidence), one single group post-test only study (level 3 evidence), and one case study (level 4 evidence) (Archibald, 2009).

Results

Nagel et al. (2012) completed two studies with mixed designs to examine the effects of F0 on EL speech intelligibility. Thirty-four and 22 non-laryngectomized speakers (50% male) participated for each study, respectively. For study one, each participant was recorded producing EL speech at three different F0 values while reading six sentences from a well-recognized assessment of dysarthric speech. Naïve listeners transcribed recordings, and percent words understood was calculated. For study two, naïve

listeners judged whether an EL speech recording (at 75Hz, 130Hz, or 175Hz for each participant) sounded feminine or masculine, and provided acceptability ratings on a 100mm visual analog scale. During a subsequent session, the listeners made acceptability judgments with the speaker sex revealed.

Appropriate statistical analyses revealed no significant differences in intelligibility based on speaker sex, but significantly better intelligibility for the lowest F0 value regardless of speaker sex in study one. In study two, the results revealed that when making speaker judgements, the listeners used different strategies depending on whether the speaker was male or female. Speech acceptability was influenced by intelligibility, such that the more intelligible the speech was, the more acceptable it was perceived to be. However, listeners were also influenced by F0 and the gender of the speaker, when making judgements about female speakers' acceptability.

Overall, this study provides suggestive evidence that utilizing a F0 value of 75Hz for monotone EL devices is significantly better than F0 values set at 130Hz or 175Hz for speech intelligibility. However, in terms of acceptability, there appears to be a trade-off between female speakers being optimally intelligible and sounding acceptably feminine. This is due to the fact that female EL speakers were judged as significantly more female sounding as the F0 of the EL increased.

Meltzner & Hillman (2005) devised a mixed study design to investigate the impact of adjusting F0 in EL speech to match that of typical speech produced by non-EL users. Speech samples from one male and one female non-laryngectomized speakers were recorded. The speakers produced a sentence with all voiced phonemes, and one with a mixture of both voiced and voiceless sounds. The speech samples were produced with either a frequency modulated or monotone EL, or using natural speech (among other manipulations not of interest to the present review). Naïve listeners chose which sentence sounded most natural from sentence pairs, and then judged typicality on a 10mm VAS scale. This was completed for all conditions including resynthesized monotone speech based on the natural sample. All procedural details were well described by the authors.

Appropriate statistical analyses revealed that modulating the F0 of EL speech made it significantly more natural sounding than EL speech produced with a constant F0 value. Nevertheless, EL speech was rated as significantly less acceptable than that of normal or laryngeal monotone speech.

Overall, this study provides suggestive evidence that modulating F0, therefore providing intonation, results in significantly more natural sounding speech than that produced with a constant F0 value.

Watson & Schlauch (2009) completed a mixed design case study, which investigated the effect of variable frequency information on intelligibility with one EL speaker. The participant read sentences aloud while the F0 information was adjusted, and conversely, while the F0 information was held constant. All sentences were phonetically balanced, declarative, and had relatively low sematic probability. Twenty naïve listeners were instructed to transcribe the sentences read by the EL speaker. Prior to listener transcriptions, noise was added to the speech files in order to mimic realistic speech-in-noise conditions.

Appropriate inferential statistics were completed, indicating that EL speech intelligibility was significantly better with variable F0 information, compared to the monotone condition. Speech intelligibility was also determined to be significantly better with varied F0 information, compared to a variable F0 speech sample resynthesized to have a flat F0 of 65Hz.

In summary, this paper provides suggestive evidence that modulating F0 information such that variable intonation patterns are incorporated into EL speech can help improve overall speech intelligibility.

Nagle & Heaton (2016) completed a within-groups repeated measures study examining speech naturalness as evaluated by ten listeners unfamiliar with EL speech. Eight male alaryngeal speakers who used a TruTone EL as either their primary (n=6) or backup mode of communication (n=2), read sentences using both an electromyographic-electrolarynx (EMG-EL with submental control) or a handheld TruTone. EMG-EL practice was provided for one hour or less prior to recording. The naïve listeners were asked to rank the naturalness of each recording on a scale of 0-100.

Appropriate inferential statistics were conducted, indicating that speech produced using the EMG-EL device was equally or more natural sounding compared to speech produced with the TruTone EL for three of the five speakers. Overall, this paper provides evidence that submental F0 modulation through the EMG-EL, provides a significantly more natural speech signal than that of the manual F0 modulation of the TruTone EL device.

Yan et al. (2014) completed a single-group post-test only study investigating how different driving signals of

EL prostheses are associated with listeners' ratings of intelligibility and acceptability. Specific to this critical analysis, the authors sought to determine the best five F0 values (ranging in increments of 3 Hz from 90-140 Hz) that corresponded to the best auditory-perceptual scores. In order to conduct this study, a minishaker system (similar to a conventional EL) was designed to be used by participants. Two laryngeal and two alaryngeal male speakers produced recordings for the perceptual ratings experiment. All participants were Cantonese speakers, and the sentences produced were from a Chinese passage. Listeners rated intelligibility and acceptability based on the following six factors: 1) voice quality 2) articulation proficiency 3) quietness of speech 4) pitch variation 5) accuracy in tone production 6) overall speech intelligibility. These speech attributes were rated on an equal-interval seven-point scale.

Appropriate statistical analyses revealed that results indicated a F0 of 108Hz was considered the most acceptable for speech, when providing a monotone signal. Overall, this study provides evidence demonstrating that adjusting the F0 of EL speech to a specific F0 (108Hz) can have positive impacts on auditory-perceptual ratings of EL speech.

Discussion

The studies reviewed in this paper provide evidence that F0 modulations such as introducing intonation and adjusting F0 values can have positive impacts on auditory-perceptual measures of EL speech. For example, for monotone EL devices, a relatively low F0 value of 75Hz increases intelligibility ratings for male speakers (Nagel et al., 2012). Interestingly, Yan et al. (2014) found the F0 value of 108Hz to yield the best acceptability ratings provided by naïve listeners. Therefore, it appears the modulation of F0 in monotone EL speech to either 75Hz or 108Hz (depending on the language) can positively impact auditory-perceptual measures of EL speech. Since intelligibility influences acceptability (Nagel et al., 2012), it is surprising that these two studies differed in which F0 value related to best auditory-perceptual ratings. However, this discrepancy could be explained by differences between the languages spoken by participants in each study. Since Yan et al. (2014) recruited Cantonese speakers, and the paper by Nagel et al. (2012) only had English speakers, the importance of tonal qualities of each language may have affected the listener judgements in each study. In a paper by Keating & Kuo (2012) which compared Mandarin (a tonal language) with English, it was found that the average F0 of Mandarin was higher than that of English. This may explain why the acceptability rating was the highest for 108Hz and not a lower F0 value for the Cantonese speakers in the study by Yan et al. (2014).

Based on the findings of this review, the auditory-perceptual measure of naturalness was also improved by utilizing submental EMG-EL pitch control devices (Nagle & Heaton, 2016), and by varying F0 information through intonation (Meltzner & Hillman, 2005). Furthermore, intelligibility ratings were also much higher for EL speech containing pitch inflection compared to monotone productions (Watson & Schlauch, 2009). Overall each of the five studies reviewed in this critical analysis provide evidence that F0 modulation can positively impact various auditory-perceptual measures of EL speech.

Of the papers included in the present review, only Nagle et al. (2012) included female participants as speakers, and found that for these females, acceptability ratings were poorer as the F0 value was lowered. However, intelligibility ratings were highest across genders, for the lowest F0 value included, which was 75Hz. Since this value corresponds to a typical F0 for a male alaryngeal speaker, it is not surprising that female speakers could not be identified as female by the naïve listeners. Since no other article in this review included female speakers, it is difficult to know whether this finding would be replicated by other researchers. The lack of female speakers within this critical review is not the only limitation of the presented articles. Majority of the papers in this review have a small numbers of participants. All papers except the one put forth by Nagle et al. (2012) had eight or less speakers participate. This limitation means that the statistical power is smaller than would be expected from a larger participant group, and can be generalized to only people matching the demographics of the participants in each study. It should also be noted that the studies conducted by Nagle et al. (2012) and Nagle & Heaton (2016) only recruited participants who were laryngeal speakers. Laryngectomees (who are the population utilizing EL devices) differ from laryngeal speakers because their vocal tract length tends to be shorter following laryngectomy surgery. This results in higher formants from that produced by laryngeal speakers (Nagle & Heaton, 2016). Nonetheless, the information gleaned from each of these studies is important for understanding how F0 modulation can positively impact auditory-perceptual measures of EL speech

Conclusion

The studies reviewed suggest that F0 modulations can positively impact auditory-perceptual measures of EL speech. Specifically, varying F0 throughout the speech signal, setting F0 to a low value (75-108Hz), and

utilizing submental F0 control improved auditory-perceptual measures of EL speech. Future research is needed in order to determine how EL speech is rated across different languages in terms of F0 modulations. This future direction is important considering the findings revealed by both the Yan et al. (2014) and Nagle et al. (2012) studies. Lastly, future research should aim to investigate more thoroughly the impact of F0 modulations on auditory-perceptual ratings of EL speech in female speakers.

Clinical Implications

Overall, the papers presented in this critical review provide suggestive evidence that F0 modulation has positive impacts on auditory-perceptual measures of EL speech. Despite the limitations of these studies, clinicians such as speech-language pathologists can utilize the information they provide in order to better care for clients. Specifically, these findings should be considered during pre- and post-operative counselling for patients undergoing laryngectomy surgery. For the female laryngectomee, counselling could include information about the apparent trade-off between acceptability and intelligibility ratings noted in the literature. It is important for female patients to have this information when considering the use of an EL, so they can be as informed as possible about the impact of F0 modulation. Furthermore, it is important for speech-language pathologists to provide knowledge to patients about the cost and benefits of each EL device. Some F0 features have the potential to improve auditory-perceptual ratings, but may come at a higher price. Overall, the evidence from this critical review can be utilized by speech-language pathologists and other professionals, to help laryngectomee patients make the best choice for themselves regarding alaryngeal speech modes including EL devices.

References

Archibald, L.M.D. (2009). Experimental design decision tree. Personal collection of L.M.D. Archibald, Western University, London, ON.

Bennett, S., & Weinberg, B. (1973). Acceptability ratings of normal, esophageal, and artificial larynx speech. *Journal of Speech, Language, and Hearing Research, 16*, 608-615.

Dollaghan, C. (2007). The handbook of evidence based practice in communication disorders. Baltimore: Paul H. Brookes Publishing Co.

Hillman, R.,Walsh, M.,Wolf, G., Fisher, S., & Hong,W. (1998). Functional outcomes following treatment for advanced laryngeal cancer. Part I: Voice preservation in advanced laryngeal cancer. Part 2: Laryngectomy rehabilitation: The state of the art in the VA System. *Annals of Otology, Rhinology and Laryngology, 172*, 1–27.

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *Journal of Acoustical Society of America, 132*(2).

Meltzner, G.S., & Hillman, R.E. (2005). Impact of aberrant acoustic properties on the perception of sound quality in electrolarynx speech. *Journal of Speech, Language, and Hearing Research, 48*(4), 766-779.

Mendenhall, W., Morris, C., Stringer, S., Amdur, R., Hineman, R.,Villaret, D., & Robbins, K. T. (2002). Voice rehabilitation after total laryngectomy and postoperative radiation therapy. *Journal of Clinical Oncology, 20*, 2500–2505.

Nagle, K.F., Eadie, T.L., Wright, D.R., & Sumida, Y.A. (2012). Effect of fundamental frequency on judgments of electrolaryngeal speech. *American Journal of Speech-Language Pathology, 21,* 154-166.

Nagle, K.F., & Heaton, J.T. (2016). Perceived naturalness of electrolaryngeal speech produced using sEMG-controlled vs. manualpitch modulation. *Interspeech 2016,* 238-242.

National Cancer Institute. *Cancer Stat Facts: Larynx Cancer*. (2014). Retrieved from: https://seer.cancer.gov/statfacts/html/laryn.html

Watson, P.J. & Schlauch, R.S. (2009). Fundamental frequency variation with an electrolarynx improves speech understanding: A case study. *American Journal of Speech-Language Pathology, 18,* 162-167.

Williams, S., & Watson, J. (1985). Differences in speaking proficiencies in three laryngectomee groups. *Archives of Otolaryngology, 111,* 216 219.

Yan, N., Ng, M.L., & Lee, T. (2014). Improving the sound quality of an electronic voice box. In Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific (pp. 1-4). IEEE.