Occurrence Frequencies of Acoustic Patterns of Vocal Fry in American English Speakers

Nassima B. Abdelli-Beruh Ph.D.

Department of Communication Sciences & Disorders,

Post Campus @ Long Island University

720 Northern Boulevard, Brookville, NY 11548

Thomas Drugman

TCTS Lab @ University of Mons

Faculté Polytechnique de Mons

Boulevard Dolez, 31, B-7000 MONS, Belgium

R. H. Red Owl

Department of Educational Leadership & Administration,

& Doctoral Program in Interdisciplinary Studies

Post @ Long Island University

720 Northern Boulevard, Brookville, NY 11548

Address Correspondence to Nassima Abdelli-Beruh, Ph. D.

Department of Communication Sciences & Disorders,

Post Campus @ Long Island University

720 Northern Boulevard, Brookville, NY 11548

Phone: 516-299-2436, Fax: (516)-299-3151

Email: Nassima.Abdelli-Beruh@liu.edu

Abstract

Summary: Objective.
The goal of this study was to analyze the occurrence frequencies of three individual acoustic patterns (A, B, C), and of vocal fry overall (A + B + C) as a function of gender, word position in the sentence (Not Last Word vs. Last Word), and sentence length (number of words in a sentence).

Study design.
This is an experimental design.

Methods.
Twenty-five male and 29 female AE speakers read the Grandfather passage. The recordings were processed by a Matlab toolbox designed for the analysis and detection of creaky segments, automatically identified using the Drugman-Kane (KD) algorithm. The experiment produced subsamples of outcomes, three that reflect a single, discrete acoustic pattern (A, B, or C) and the fourth that reflects the occurrence frequency counts of Focal Fry Overall without regard to any specific pattern. Zero-truncated Poisson regression analyses were conducted with Gender and Word Position as predictors, and Sentence Length as a covariate.

Results.
The results of present study showed that the occurrence frequencies of the three acoustic patterns and vocal fry overall (A + B + C) are greatest in sentence-final position but are unaffected by sentence length. The findings also reveal that AE female speakers exhibit Pattern C significantly more frequently than Pattern B and the converse holds for AE male speakers.

Conclusions.
Future studies are needed to confirm such outcomes, assess the perceptual salience of these acoustic patterns and determine the physiological correlates of these acoustic patterns. The findings have implications for the design of new excitation models of vocal fry.

INTRODUCTION

The present study is an extension of the work by Abdelli-Beruh, Wolk and Slavin[1], Wolk, Abdelli-Beruh, and Slavin [2] and Drugman, Kane and Gobl [3]. Abdelli-Beruh et al. [1] and Wolk al. [2] reported that vocal fry is more frequently perceived at the end of sentences than elsewhere in sentences, which intimates, in accordance with many studies, that vocal fry serves as a syntactic marker. [1-2, 4-17] They also found that the prevalence of vocal fry is greater in the speech of female than male American English (AE) speakers engaged in a reading task, suggesting, in agreement with previous findings that vocal fry might serve as well as a gender marker. [1-2, 5, 7, 12-13, 17, 19-20] Using an automated detection algorithm based on the KD (Kane-Drugman) features, Drugman, et al. [3] identified three different acoustic patterns (A, B, C) of vocal fry in the speech of 11 speakers of four different languages engaged in various speaking tasks.

The present study expands on the three acoustic patterns (A, B, C) associated with vocal fry, which were documented by Drugman et al. [3] They applied the KD perceptually guided automated detection algorithm to speech samples from a small number of linguistically diverse speakers engaged in various speaking conditions. In the present study, the KD algorithm was applied to speech productions obtained from a large sample (25 males and 29 females) of AE speakers engaged in a reading task. The present research sought also to continue the work of Abdelli-Beruh et al. [1], and Wolk et al. [2] by testing whether the occurrence frequencies of each of the three acoustic events and those of vocal fry overall (A +B + C) vary across word position (within vs. end of sentence). This study further addresses whether the frequencies of occurrences observed are consistent with the evidence of previous work based on perceptual data[1, 2], which have

indicated that vocal fry tends to be predominantly perceived at the end of sentences. Moreover, it investigated whether the gender difference observed in the perceptual data[1, 2] is reflected in the occurrence frequencies of A, B, and C acoustic patterns. Because the effects of the predictors (i.e., gender and word position) considered in this study might be influenced by sentence length (operationalized as the number of words in each sentence), the analysis also incorporated that measure as a covariate. The findings of the present study have implications for automated speech recognition programs and speech synthesis as it might help improve the naturalness of the synthesized tokens. [55]

Various terms co-exist in the literature to describe this voice pattern. In the field of speech language pathology, the prevalent terms are vocal fry, pulse register, creaky voice, creak, whereas in the field of psycholinguistics the terms "irregular phonation", "pulse phonation" "glottalization" or "laryngealization" are most frequently used. [1-2, 7, 16-17, 19, 22, 31-36] It is not known whether there are substantial vibratory, acoustical and auditory differences associated with these different labels but for Hollien[32] pulse register is "undoubtedly" synonymous with vocal fry, glottal fry, creak and strohbass (p. 2); Monsen and Engebreston [37] were of the opinion that vocal fry and creaky voice can be used interchangeably; Redi and Shattuck-Hufnagel [17] used glottalization and creak alternatively. Gorden and Ladefoged [38] used creaky phonation or vocal fry to describe the same vocal phenomenon. Imaizumi and Gauffin [39] considered creak to be a "special case" of vocal fry. Laver [28], however, distinguished creak from creaky voice. The interchangeability in the terminology may be confusing but it attests to the fact that these distinctions may not be linguistically relevant. In this paper, the auditory criterion used in the KD approach is that of "rough quality with the additional sensation of repeating

impulses". [40] The roughness voice quality and its concomitant multiple glottic pulses have been documented previously. [17, 32, 34, 41-42]

Early acoustic analyses of voice intentionally produced with vocal fry show that the perception of vocal fry is associated with a specific range of fundamental frequency (F_0) that lies below that of the modal register for both males and females. [32, 34, 41, 43-48] In addition to its unique frequency range, Hollien [32] states that vocal fry also differs from modal and falsetto registers in its amplitude ranges (i.e., lower than for the other registers) and in its frequency make up as voice spectra show single or double pulses.[45] These distinct acoustic features, which are associated with unique vocal fold length, vocal fold thickness and vibratory patterns are associated with the distinctive perception of vocal fry. [32, 34, 49]

More recently, Redi and Shattuck-Hufnagel [17] using a combination of perceptual and acoustic criteria describe four types of acoustic patterns associated with glottalization: 1) aperiodicity (i.e., glottal pulses irregular in duration from period to period); 2) creak (i.e., glottal pulses of low fundamental frequency accompanied by damping); (3) diplophonia (i.e., glottal periods with systematic repetition in shape, duration, or amplitude); 4) glottal squeak (i.e., or a sudden shift to relatively high sustained F_o of very low amplitude" (p. 414). They reported that the latter type of glottalization occurred infrequently, while the former ones have been previously documented by Huber. [52]

Drugman et al. [3] analyzed excerpts of conversational speech samples produced by four American English speakers, two Japanese, two Swedish and two Finish speakers. Using an automated detection program for creak based on features developed by Kane et

al., [40], Drugman et al., [53] and Ishi, Sakakibara, Ishiguro and Hagita, [54], they reported three patterns of glottal events present in the speech of the majority of the speakers regardless of their native tongue. Pattern A is characterized by a very irregular temporal structure, in which the period between glottal peaks does not follow any predictable pattern (see Figure 1). In contrast to pattern A, pattern B is characterized by a more regular periodicity and the presence of two prominent excitatory peaks (see Figure 2). The first peak is the Glottal Opening Instant (GOI) and it is likely secondary to a sudden opening of the glottis. The second peak corresponds to the Glottal Closure Instant (GCI). The glottal opening period, defined as the timespan between the GOI and the subsequent GCI, has been shown to be rather constant across the creaky segments of a given speaker. The GCI is generally followed by a long glottal closed phase.[6] Similarly to pattern A, however, pattern B presents a marked discontinuity at the glottal closure instants. Pattern C, like pattern B, is characterized by regular periods such as visible in modal voice but it exhibits F_o below 50 Hz. Unlike pattern B, pattern C does not present any secondary peak (see Figure 3). Pattern C likely corresponds to what Ishi et al. [54] called the "single-pulse patterns" (p. 49) and the "creak" category described by Redi and Shattuck-Hufnagel. [17]

Drugman et al. [3] examined the frequencies of acoustic patterns A, B and C and found that speakers use jointly more than one acoustic pattern, a fact that has been documented previously.[17, 54- 56] They also reported that pattern A is the most frequently used by nine speakers out of 11, except for two male speakers (BDL -USA and MV-Finland) in the speech of whom pattern B was most frequent. Interestingly, pattern B had been reported to be most frequent in Drugman et al. [57] and Raitio et al. [58] Such

discrepancy in the results might suggest that previous excitation models may have been inadequate to detect the temporal irregularities characteristics of A or it might suggest that the prevalence of a given pattern varies with characteristics of the speakers or the task involved. The speakers, who predominantly used patterns B, were both males and were both involved in reading tasks. The remaining speakers, males and females, for whom there was a higher prevalence of pattern A than of any other pattern, were involved in conversational speech. It is important to further investigate the factors that might contribute to such discrepancies as vocal fry plays an important role in pragmatics, para-linguistics, meta-linguistics (emotion), and socio-linguistics. [1, 2, 4-5, 7-19, 21-25, 27-30]

It may be the case that the gender difference reported in the literature, where AE females produce more fry than AE males might be linked to the prevalence of one acoustic pattern over another. Better understanding of the factors that determine the use of one pattern over another by a gender group or within the phrasal structure is necessary for adequately modeling the feature characteristics of fry so that such factors are incorporated in a speech synthesis model for increased adequacy and naturalness.

Method

Subjects

Twenty-five male and 29 female native speakers of American English were recorded. They were between the ages of 18 and 25 years. At the time of the recording, they did not acknowledge having any known history of anatomical or physiological oral defects, vocal pathology and no known hearing disorders.

Apparatus

The tokens were recorded on one channel of a digital recorder (Marantz, model 300) at a sampling rate of 44,100 samples per second. A high definition microphone (Shure, Beta 58A) was positioned on a stand, 30 cm from the speaker's mouth. All files were then imported from the left channel of the recordings onto an IMac. Audacity software was used for the acoustic measurements.

Automatic Detection of Vocal Fry/Creak using KD features

The recordings were processed by a Matlab toolbox especially designed for the analysis and detection of creaky voice.[6, 40] For each utterance, the creaky segments were automatically identified using the algorithm proposed in Drugman et al.[6] This algorithm works as follows. Every 10 ms, ten acoustic features are extracted. These features are tailored for the characterization of creaky voice and some were proposed in previous studies.[40, 54] The first and second derivatives of these features are appended in order to model the speech dynamics. The resulting 30-dimensional vectors are fed into an Artificial Neural Network (ANN) discriminatively trained to draw a binary decision about the presence of creaky voice. The training material consisted of about 200 minutes of manually annotated data from 11 speakers in 4 languages (US English, Finnish, Swedish and Japanese) involved in different communication contexts. This algorithm clearly outperformed other state-of-the-art approaches and obtained a relatively high performance with F1[1] scores varying between 0.6 and 0.8. In this way, each utterance was associated with a list of the creaky voice segments (characterized by their starting and ending times) identified by the aforedescribed algorithm. Each segment was further

[1] *The F1 score is a measure used in statistical analysis of binary classification which combines true positives, false positives and false negatives into one single metric*

manually annotated by the second author into one of the 3 creaky patterns discussed in Drugman, et al.[6] For this purpose, both the speech and the linear prediction residual waveforms were visualized and the pattern annotation strictly followed the same criteria as in Drugman, et al. [6]

Design and material

The Grandfather passage was used in this study. It is a standard reading material commonly used by speech pathologists in clinical settings to test an individual's ability to produce connected speech (See Appendix A).

Study design

The study design comprises 100 individual word tokens (see Appendix A), some of which occur multiple times for a total of 130 pattern opportunities (i.e., experimental conditions under which a specific pattern might be observed). Replicating those pattern opportunities for females and males yielded an overall study design of 260 distinct opportunities for each pattern to be observed at any frequency level greater than zero. Multiple subjects provided data for each pattern opportunity.

Data analysis

The experiment produced subsamples of outcomes that were measured at two levels of specificity. The first three subsamples include individual outcome measures that reflect a single, discrete acoustic pattern (A, B, or C) detected by the KD algorithm (Drugman, et al. [3]) as described above. The fourth subsample includes a combined outcome measure that reflects the occurrence frequency counts of Focal Fry Overall without regard to any specific acoustic pattern of vocal fry. Each of the three vocal fry acoustic patterns and the overall vocal fry outcome measure was analyzed under varying

conditions related to two hypothesized predictors (i.e., gender and whether the word occurs at the ends of a sentence) and a covariate (i.e., sentence length as operationalized by number of words). The occurrence frequencies of the individual acoustic patterns and vocal fry overall were analyzed using a series of four zero-truncated Poisson regression models appropriate for predicting and explaining discrete count outcomes. Although traditional nonparametric statistical approaches such as Pearson chi-square may be used in analyzing count data, nonparametric routines generally have lower statistical power, are not feasible for the analysis of extended ranges of frequency counts, and are limited in their ability to account for the effects of multiple predictors and/or covariates. Zero-truncated Poisson regression enabled testing of the effects of the two hypothesized predictor variables and the covariate. The assumptions associated with the Poisson distribution were confirmed by estimating a zero-truncated negative binomial regression model for each pattern and testing the statistical significance of the respective overdispersion factors alpha from each using a likelihood ratio chi-square. The goodness of fit of each of the four zero-truncated Poisson models was confirmed with Wald's chi-square and McFadden's pseudo R^2. The discovered effects were presented visually in margins plots of the predicted conditional ($Pattern_{A,\ B\ or\ C} > 0$ or $Pattern_{A+B+C} > 0$) probabilities for the frequencies based on the predictors found to be statistically significant ($\alpha = .05$) in each model. Analyses were conducted using Stata/IC version 14.

Procedure

The speakers were leaning their back against the wall of the sound proof room. The speakers were asked to move minimally during the session. To keep the distance between the mouth of the speaker and the microphone somewhat constant, a 30-cm folder

was affixed to the stand to which the microphone was attached. The speakers were asked to keep this distance constant throughout the recording.

Participants were asked to use their natural speaking intensity and pitch level. They were asked to read the passage once and they were asked whether they felt they were using their natural speaking intensity and pitch level. When they responded positively, recordings commenced.

Results

From the 260 total opportunities (i.e., number of unique and non-unique words spoken by each subject) for vocal fry to occur in any pattern, separate subsamples were extracted for those opportunities where vocal fry was observed in any pattern and for the opportunities where it was observed in specific patterns. The four subsamples consist of: 129 opportunities in which it was observed as Pattern A; 99 opportunities in which it was observed as Pattern B; 77 in which it was observed as Pattern C; and 160 opportunities in which vocal fry was observed in any pattern. Because multiple subjects provided data for each pattern opportunity, the fourth sample includes cases in which more than one pattern was observed and, therefore, contains fewer cases than the sum of the pattern subsamples.

Zero-truncated Poisson regression analyses were conducted on those subsamples focusing on vocal fry pattern opportunities as the unit of analysis in order to model and explain the frequency counts of the respective individual patterns and of vocal fry overall. The dispersion assumptions of zero-truncated Poisson regression for each analysis were found to be sufficiently satisfied as evidenced by non-statistically significant overdispersion factors alpha in the associated zero-truncated negative binomial regressions (all $P(\alpha) > 0.05$). The results of the zero-truncated Poisson regression

analyses for the three acoustic patterns and for vocal fry overall are presented in Tables 1-4. The conditional probabilities of the frequencies of occurrence of each pattern and of vocal overall are shown in the margins plots in Figures 1-3, reflecting the non-linear nature of the probabilities at varying frequency levels. Only statistically significant predictors are modeled in the margins plots.

[Insert Tables 1-4 about here.]

[Insert Figures 1-3 about here.]

The Poisson regression analyses produced highly statistically significant models (all $p < 0.001$) with excellent goodness of fit (Wald $\chi^2_{(4)}$ for A = 239.64, Wald $\chi^2_{(4)}$ for B = 137.08, Wald $\chi^2_{(4)}$ for C = 194.94, Wald $\chi^2_{(4)}$ for Vocal Fry Overall = 946.28) and adequate predictive values (Pseudo R^2 for A = 0.33, Pseudo R^2 for B = 0.36, Pseudo R^2 for C = 0.42, Pseudo R^2 for Vocal Fry Overall = 0.60). The results of these analyses are used below to evaluate the effects of speaker gender (Male vs. Female) and word position in the sentence (Not Last Word vs. Last Word), after adjusting for the potential covariance with sentence length (number of words in each sentence).

Gender effects

As shown in Figure 1, the analyses revealed that gender affects the frequencies of the occurrence of Patterns B and C, when considered in the context of word position and sentence length. The effect of gender is more pronounced in the case of Pattern C, where females tend to manifest the pattern with a greater frequency than males and females are more likely to exhibit Pattern C than B at any level of frequency greater than 2. Males exhibit Pattern B more frequently than females and are unlikely to exhibit Pattern C at any level of frequency above one. No statistically significant gender effect was found for

the conditional frequencies of Pattern A individually or Vocal Fry Overall (see Tables 1 and 4).

Word position effects

At higher frequencies (15 or greater) of occurrences, the cumulative conditional probabilities for all three individual patterns are greatest when the word is the last in the sentence, as shown Figure 2. When the frequency of occurrence is between 10 and 15, the cumulative conditional probability for Pattern B is greater when the occurrence opportunity is at the last word of the sentence than when it is within the sentence and also for any word position for instances of Patterns A or C. At lower occurrence frequencies between 2 and 7, the results are opposite, with the conditional probabilities of all acoustic patterns being greater in the non-final word position. When considering Vocal Fry Overall without regard to specific patterns, the influence of final word position is clearly dominant as depicted in Figure 3.

Sentence length effects

The study found no systematic, statistically significant effect of the covariate sentence length after gender, word position, and the potential interaction of those factors were taken into account (see Tables 1-4).

Effects on Vocal Fry Overall

Having identified the effects of the two hypothesized predictors and the covariate within each of the three acoustic patterns, the analysis then focused on the effects of those factors and the existence of the specific acoustic pattern in explaining the frequency of Vocal Fry Overall. As evident in Figure 3, at all levels of frequency of the occurrence of Vocal Fry Overall greater than 2, Pattern A is systematically more probable than either of

the other two patterns, which have some degree of periodicity. Pattern C, which has the most periodicity has the lowest cumulative probability at any frequency level. Although not readily apparent in the figure, the cumulative conditional probability of Pattern C (0.23) is lower than that of Pattern B (0.25) and Pattern A (0.28) even at a frequency level of 1.

Discussion

The present study sought to test whether the occurrence frequencies of the three acoustic patterns associated with vocal fry, and of vocal fry overall without regard to any specific pattern are affected by gender, word position and sentence length.

Data analyses show that gender does not significantly affect the occurrence frequencies of vocal fry overall. This finding is in apparent disagreement with previous findings based on acoustic measures.[5, 7, 17-20] It is also in disagreement with studies based on the perceptual evaluation of speech samples produced by AE male and female speakers, which reveal that listeners perceive vocal fry to be more frequent in the speech of female than male speakers.[1, 2] The present analyses, however, reveal a subtle gender difference in the prevalence of one acoustic pattern over another. While speakers of the two gender groups use Pattern A at about the same occurrence frequency, female speakers produce more often Pattern C than Pattern B, whereas males do the converse. It is not clear for what might account for such preference. Both Patterns B and C have fairly regular temporal features but they differ in the presence versus absence of secondary excitatory patterns, which is the peak with the most energy in between two GCI (see figure 2). It is not clear what account for this gender difference but additional studies are needed to confirm such outcomes, assess the perceptual salience of these acoustic

patterns and determine whether the gender difference previously reported in the literature, where AE females are perceived to produce more vocal fry than AE males might be linked to the prevalence of one acoustic pattern over another.

Furthermore, the data show that the occurrence frequencies of all three acoustic patterns and of vocal fry overall are much higher on the last word alone than on any other word within any sentence. This acoustic finding is in concordance with previous findings that documented a higher prevalence of fry at phrase boundaries, which has been taken to suggest that vocal fry is a syntactic marker in many languages [1, 5, 8, 11, 15-17, 44, 34] as well as in speakers of many other languages.[2, 4, 7, 9-10, 12-14]

The present findings also show that pattern A, which unlike patterns B and C does not have any periodic component, is the most frequently occurring acoustic pattern in the creaky speech of both male and female AE speakers. The prevalent use of acoustic pattern A is in agreement with previously documented findings in American English and in other languages. [3, 17] Drugman et al. [3] reported that nine out of 11 speakers of different languages (Swedish, Finnish, AE, Japanese) used pattern A most frequently. Out of the five AE speakers in their study, four speakers predominantly used pattern A over the other two patterns. Redi and Shattuck-Hufnagel [17] also reported that five of the six professional and seven of the eight non-professional voice users exhibited significantly more often aperiodic patterns than creak (classification 2) patterns. The category of glottalization described by Redi and Shattuck-Hufnagel [17], which they called "aperiodicity" (classification 1) resembles pattern A. Pattern A is also very similar in its description to an acoustic pattern identified by Ishi, Ishiguro and Hagita [55 pg. 2058] as having

"very low fundamental frequencies with discrete glottal pulses, and eventual irregularity in periodicity." They frequently detected it in 11 out of 15 passages.

The finding that pattern A is the most frequently occurring acoustic pattern is, however, in disagreement with other findings that show that pattern B is the most frequently occurring pattern in creaky speech segments. [57-58] This apparent discrepancy might be ascribed to the fact that the modeling system developed by Drugman et al. [57] and Raitio et al. [58] used two speakers (BDL and MV), who were primarily B-pattern users, therefore biasing the design of the excitation model. The findings of the present study show the importance of a large sample of speakers in the configuration of voice models.

The results also showed that the occurrence frequencies of the three acoustic patterns and those of vocal fry overall are not significantly affected by the number of word in a sentence. There is no other study to our knowledge that has assessed the contribution of sentence length on the occurrence frequencies of vocal fry. Additional studies are therefore needed to corroborate such a finding.

In sum, the results of present study showed that the occurrence frequencies of three acoustic patterns documented by Drugman et al. [3] and the occurrence frequencies of vocal fry overall (A + B + C) are greatest in sentence-final position and are unaffected by sentence length. The findings also reveal that AE female speakers exhibit Pattern C significantly more frequently than Pattern B and the converse holds for AE male speakers.

Using an automated recognition program, this work is a first attempt to document the occurrence frequencies of three distinct acoustic patterns in the speech of a large sample of AE speakers engaged in a reading task. It is not clear whether naïve listeners

can perceptually distinguish between the three acoustic patterns and hence whether they use it at a linguistic, para-linguistic, meta-linguistic, and/or pragmatics level. Future studies need to assess the perceptual salience and the weighted contribution of these three acoustic patterns and determine whether there is a need to increase the complexity of a new speech synthesizer by incorporating these three patterns. There also a need to investigate the laryngeal physiology associated with either of these three acoustic patterns associated with vocal fry.

Appendix A

You wish to know all about my grandfather. Well, he is nearly 93 years old, yet he still thinks as swiftly as ever. He dresses himself in an ancient, black frock coat, usually minus several buttons. A long, flowing beard clings to his chin, giving those who observe him a pronounced feeling of the utmost respect. When he speaks his voice is just a bit cracked and quivers a trifle. Twice each day he plays skillfully and with zest upon a small organ. Except in the winter when the snow or ice prevents, he slowly takes a short walk in the open air each day. We have often urged him to walk more and smoke less but he always answers, "Banana oil!" Grandfather likes to be modern in his language.

References

1. Abdelli-Beruh, N., Wolk, L., & Slavin (2014). Prevalence of Vocal Fry in Young Adult Male American English Speakers, *Journal of Voice*, 26(20), 185-190.

2. Wolk, L., Abdelli-Beruh, N-B, and Slavin, D. (2012). Habitual Use of Vocal Fry in Young Adult Standard American English Speakers, *Journal of Voice*, 26(3), e111-e116.

3. Drugman, T., Kane, J., & Gobl (2013). Data-driven Detection and Analysis of the Patterns of Creaky Voice, *Computer Speech & Language*, vol. 28, issue 5, pp. 1233-1253.

4. Belotel-Grenié, A., & Grenié, M. (2004). "The Creaky Voice Phonation and the Organization of Chinese Discourse" [on-line], *International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages*, 28-30.

5. Dilley, L., Shattuck Hufnagel, S., & Ostendorf, M. (1996). Glottalization of Word-Initial Vowels as Function of Prosodic Structure, *Journal of Phonetics*, 24(4), 423–444.

6. Drugman, T., Alku, P., Alwan, A., & Yegnanarayana, B. (2014). Glottal Source Processing: from Analysis to Applications. *Computer Speech & Language*. Vol. 28, issue 5, pp. 1117-1138.

7. Henton, C. G., & Bladon, R.A.W. (1988). Creak as a Socio-Phonetic Marker. In L. Hyman and C.N. Li. (Eds.). *Language, Speech and Mind: Studies in Honor of Victoria A. Fromkin*. Beckenham, Croon Helm, 3-29.

8. Kreiman, J. (1982). Perception of Sentence and Paragraph Boundaries in Natural Conversation. Journal of Phonetics, 10, 163–175.

9. Kane, J., Pápay, K., Hunyadi, L., & Gobl, C. (2011). On the Use of Creak in Hungarian Spontaneous Speech. *Proceedings of 13th International Congress of Phonetic Sciences*. 1014-1017.

10. Lehiste, I. (1965). Juncture. *Proceedings of the 5th International Congress of Phonetic Sciences*, 172-200.

11. Lehiste, I., 1979. Sentence Boundaries and Paragraph Boundaries: Perceptual Evidence. In *The Elements: a Para-Session on Linguistic Units and Levels*, Chicago Linguistics Society, 15, 99-109.

12. Local, J. K., Wells, W.H.G., & Sebba, M. (1985). Phonology of Conversation. *Journal of Pragmatics*. 9, 309-330

13. Local, J. K., Kelly, J., & Wells, B. (1986). Towards a Phonology of Conversation: Turn-Taking in Tyneside English. *Journal of Linguistics*. 22, 411-437.

14. Ogden, R. (2001). Turn-holding, Turn-yielding, and Laryngeal Activity in Finnish Talk-in-Interaction. *Journal of the International Phonetics Association*, 31, 139-152.

15. Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and Glottal Stop. In G. Docherty, & D.R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 90–119). Cambridge: Cambridge University Press.

16. Surana, K. & Slifka, J. 2006. Is Irregular Phonation a Reliable cue Towards the Segmentation of continuous speech in American English? In: *ICSA International Conference on Speech Prosody*. Dresden, Germany.

17. Redi, L. & Shattuck-Hufnagel, S. (2001). Variations in the Realization of Glottalization in Normal Speakers. *Journal of Phonetics*, 29, 407-429.

18. Byrd, D. (1994). Relation of Sex and Dialect to Reduction. *Speech Communication*, 15, 39-54.

19. Dilley, L., & Shattuck Hufnagel, S. (1995). Variability in Glottalization of Word Onset

Vowels in American English, *Proceedings of the 5th International Congress of Phonetic Sciences*, 95, 4, 586-589.

20. Dilley, L., Shattuck Hufnagel, S., & Ostendorf, M. (1994). Prosodic Constraints on Glottalization of Vowel Initial Syllables in American English. *Journal of the Acoustical Society of America*, 94, 2978-2979.

21. Allen, J. (1970). The Glottal Stop as a Junctural Correlate in English. *Journal of the Acoustical Society of America*, 40(1), 57-58.

22. Böhm, T. & Shattuck-Hufnagel, S. (2007). "Utterance Final Glottalization as Cue for Familiar Speaking Recognition", In *INTERSPEECH-2007*, 2657-2660

23. Carlson, R., Gustafson, K., & Strangert, E. (2006). Cues for hesitation in speech synthesis. *Proceedings of Interspeech*, Pittsburgh, USA, 1300–1303.

24. Espy-Wilson, C., Manocha, S., Vishnubhotla, S., (2006). A new set of Features for Text-Independent Speaker Identification. *Proceedings of Interspeech (IC- SLP)*, Pittsburgh, Pennsylvania, USA, 1475–1478.

25. Elliot, J. R., (2002). The Application of a Bayesian Approach to auditory Analysis in Forensic Speaker Identification. *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, 315–320.

26. Bladon, R.A., Henton, C.G. & Pickering, J.B. (1984). Towards an Auditory Theory of Speaker Normalization. *Language Communication,* 4,59-69.

27. Gobl, C., & Ní Chasaide, A. (2003). The Role of Voice Quality in Communicating Emotion, Mood and Attitude. *Speech Communication*, 40, 189-2002.

28. Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

29. Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2005). Voice Quality and f0 cues for Affect Expression. *Proceedings of Interspeech*, Lisbon, Portugal, 1849–1852.

30. Yuasa, I. P. (2010). Creaky Voice: a new Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women. *American Speech*, 85, 4, 315-337.

31. Ewender, T., Hoffmann, S., & Pfister, B. (2009). Nearly Perfect Detection of Continuous F0 Contour and Frame Classification for TTS Synthesis, *In Proceedings of Interspeech*, 100-103.

32. Hollien, H. (1972). On Vocal Registers. *Communication Sciences Laboratory Quarterly Report*, 10, (1), 1-33.

33. Hollien, H. (1974). On Vocal Register. *Journal of Phonetics*, 2, 125-144.

34. Hollien, H. & Wendahl, R.W. (1968). Perceptual Study of Vocal Fry. *Journal of the Acoustical Society of America*, 43(3), 506-509.

35. Moore, P. & Von Leden, H. (1958). Dynamic Variations in the Vibratory Pattern in the Normal Larynx. *Folia Phoniatrica*, 10(4), 205-238.

36. Slifka, J. (2007). Irregular Phonation and its Preferred Role as Cue to Silence in Phonological Systems. *Proceedings of the 16th International Congress of Phonetic Sciences*. 228-232.

37. Monsen and Engebreston (1977)

38. Gordon, M., & Ladefoged, P. (2001). Phonation Types: a Cross-Linguistic Overview. *Journal of Phonetics*, 29 (4), 383–406.

39. Imaizumi S,. & Gauffin J. (1991). Acoustic and Perceptual Characteristics of Pathological Voices: Rough, Creak, fry and Diplophonia. *Annual Bulletin Res. Institut Logopedia Phoniatria*, 25, 109-19.

40. Kane, J., Drugman, T., & Gobl, C., (2013). Improved Automatic Detection of Creak. *Computer Speech and Language*, 27(4), 1028–1047.

41. Blomgren, M., Chen, Y., Ng, M., & Gilbert, H. (1998). Acoustic, Aerodynamic, and Perceptual Characteristics of Modal and Vocal Fry Registers. *Journal of the Acoustical Society of America*, 103, 2649-2658.

42. Timcke, R., von Leden, H., & Moore, P. (1959). Laryngeal Vibrations: Measurements of the Glottic Wave. *Archives of Otolaryngology*, 68, 1–19.

43. Allen, E., & Hollien, H. (**1973**). A Laminagraphic Study of Pulse (Vocal fry) Phonation, *Folia Phoniatrica*. **25**, 241–250.

44. Hollien, H., Moore, P., Wendahl, R.W., & Michel, J. F. (1966). On the Nature of Vocal Fry. *Journal of Speech and Hearing Research*, 9, 245-247.

45. Hollien, H. & Michel, J.F. (1968). Vocal Fry as a Phonational Register. *Journal of Speech and Hearing Research*, 11, 600-604.

46. McGlone, R.E. & Shipp, T. (1971). Some Physiologic Correlates of Vocal-Fry Phonation. *Journal of Speech and Hearing Research*, 14, 769-775.

47. Murry, T. (1971). Subglottal Pressure and Airflow Measures During Vocal Fry Phonation. *Journal of Speech and Hearing Research*, 14, 544-551.

48. McGlone, R. (1967). Air Flow During Glottal Fry. *Journal of Speech and Hearing Research*. 10, 299-304.

49. Hollien, H., Damsté, H., & Murry, T. (1969). Vocal Fold Length During Vocal Fry Phonation. *Folia Phoniatrica*, 21, 179-198.

50. Whitehead, R. W., Metz, D.E., and Whitehead, B.H. (1984). Vibratory Patterns of the Vocal Folds During Pulse Register Phonation," *Journal of the Acoustical Society of America,* 75(4), 1293–1297.

51. Wendhal, R. W., Moore, P., & Hollien, H. (1963). Comments on Vocal Fry. *Folia Phoniatrica*. 15, 251-5.

52. Huber, D. (1988). *Aspects of the Communicative Function of Voice in Text Intonation*. Ph.D. thesis, University of Goteborg/Lund.

53. Drugman, T., Kane, J., & Gobl, C. (2012b). Resonator-based Creaky Voice Detection. *Proceedings of Interspeech*, Portland, Oregon, USA.

54. Ishi, C., Sakakibara, K., Ishiguro, H., & Hagita, N. (2008a). A Method for Automatic Detection of Vocal fry. *IEEE Transactions on Audio, Speech, & Language Processing,* 16 (1), 47–56.

55. Ishi, C. T., Ishiguro, H., Hagita, N. (2007). Acoustic and EGG Analysis of Pressed Phonation. *Proceedings of International Conferences of Phonetic Sciences*, Saarbrucken, Germany, 2057– 2060.

55.Yoon, T. J., Zhuang, X., Cole, J., & Hasegawa-Johnson, M. (2006, May). Voice quality dependent speech recognition. In International Symposium on Linguistic Patterns in Spontaneous Speech.

56. Ishi, C. T., Ishiguro, H., & Hagita, N. (2010). Acoustic, Electroglottographic and Paralinguistic Analysis of "Rikimi" in Expressive Speech. *Proceedings of Speech Prosody*, Chicago, USA, 1–4.

57. Drugman, T., Kane, J., & Gobl, C. (2012a). Modeling the Creaky Excitation for Parametric Speech Synthesis. *Proceedings of Interspeech*, Portland, Oregon, USA. . pp. 1424-1427.

58. Raitio, T., Kane, J., & Drugman, T. (2013). HMM-based Synthesis of Creaky Voice. *Proceedings of Interspeech*. Lyon, France. pp. 2316-2320.

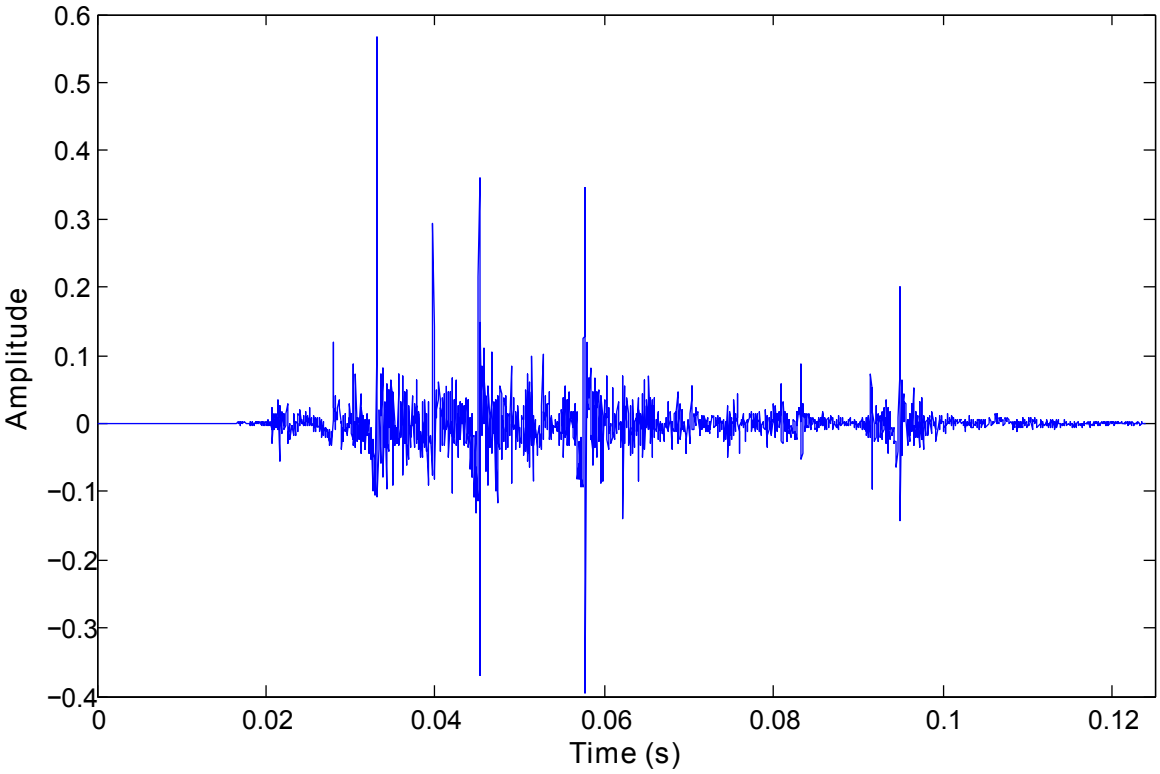

Figure 1 shows a sample of an aperiodic pattern (Pattern A): Linear prediction residual for a typical Pattern A creaky voice segment. The waveform exhibits clear discontinuities with a highly irregular temporal structure. These peaks appear sporadically, and the inter-peak duration does not seem to follow a clear deterministic rule as it is the case for regular patterns.

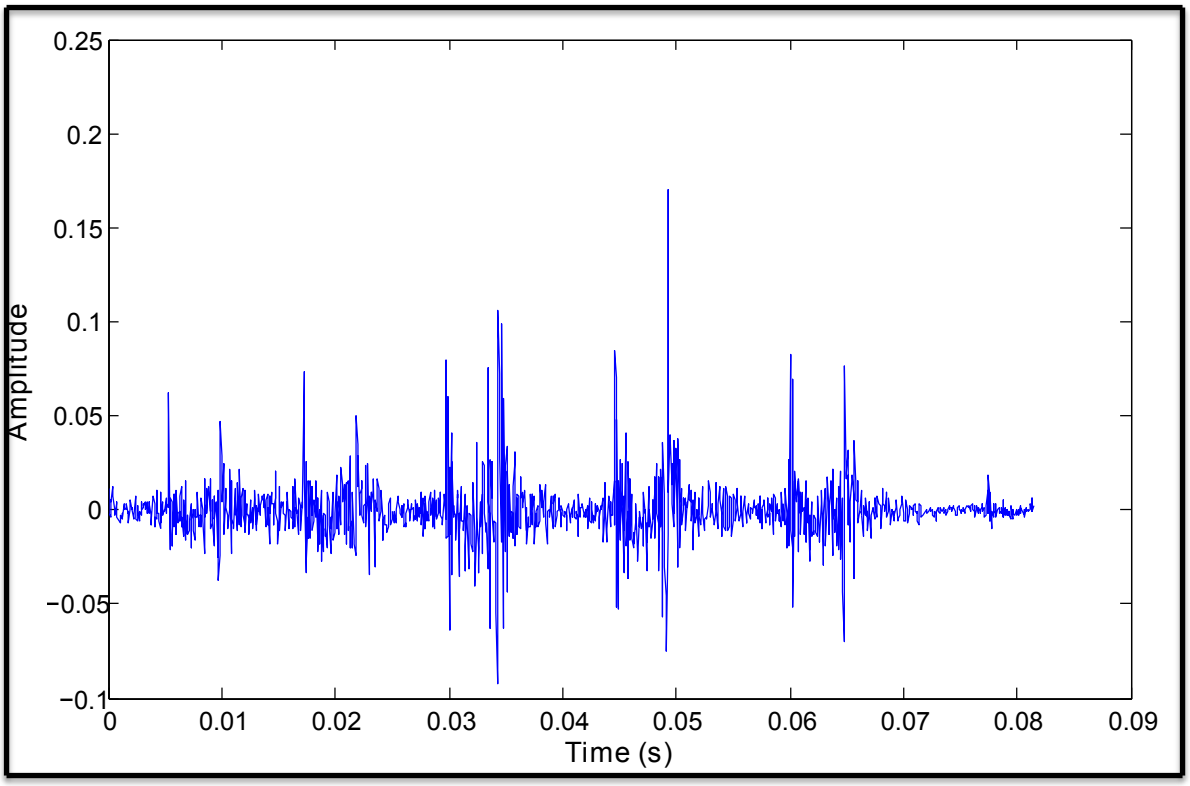

Figure 2 shows a sample of Periodic Double pattern (Pattern B): Linear prediction residual for a typical Pattern B creaky voice segment. The waveform exhibits fairly regular temporal characteristics, with a stable pattern comprising two clear discontinuities. In this example, the main excitation peak called Glottal Closure Instant (GCI) is visible at 0.05 seconds, and it is preceded by a so-called secondary peak associated with a sudden opening of the glottis at the so-called Glottal Opening Instant (GOI), as seen at 0.045 seconds.

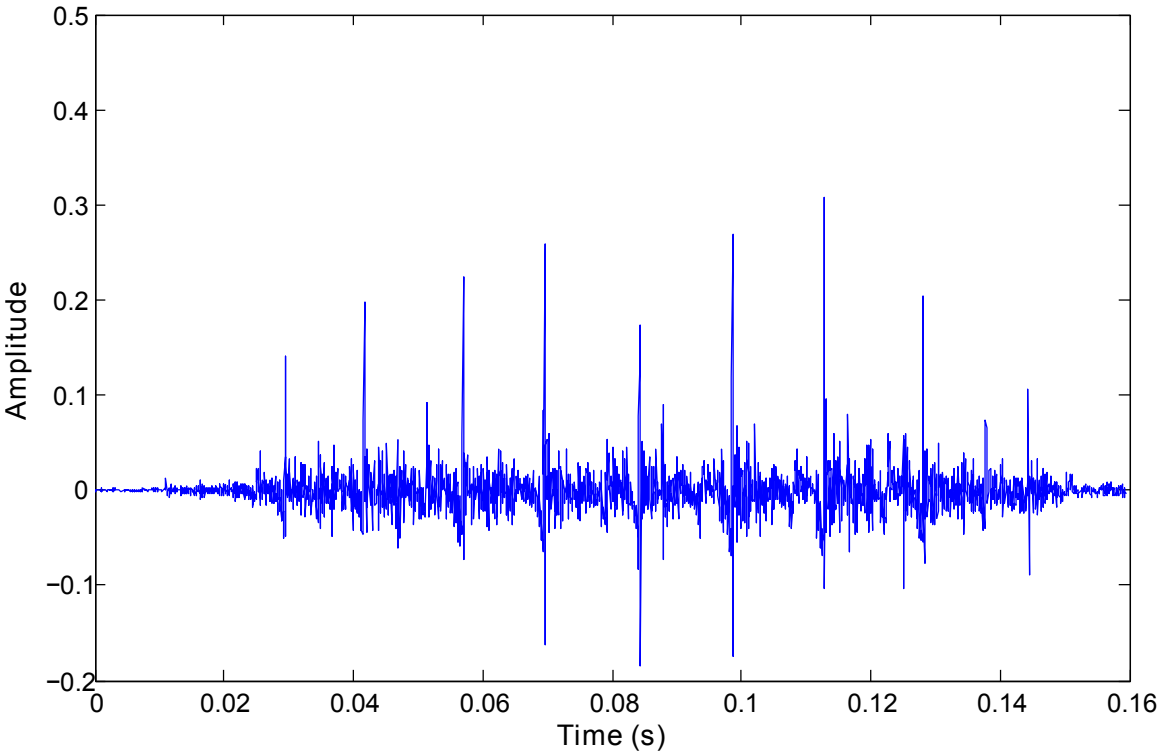

Figure 3 shows a Periodic Simple pattern (Pattern C): Linear prediction residual for a typical Pattern C creaky voice segment. The waveform exhibits fairly regular temporal characteristics with a single discontinuity per cycle. This excitation is similar to that of modal voice, but the fundamental frequency is much lower (in the present figure between 65 and 70Hz). Unlike Pattern B, pattern C does not display strong secondary excitation peaks at the GOIs.

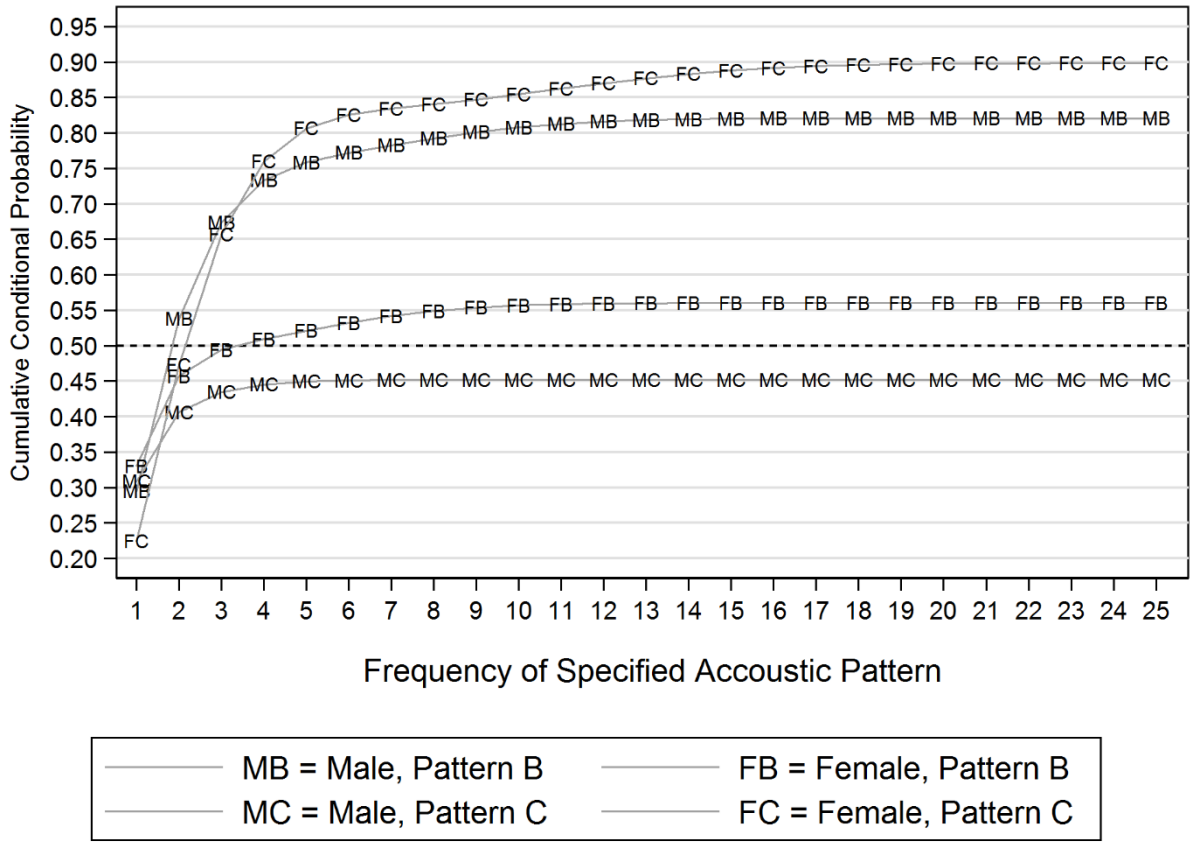

FIGURE 4. Margins plot of the cumulative conditional probabilities of occurrence of specific acoustic patterns of vocal fry by frequency level as predicted by gender. The effect of Gender on Pattern A is not statistically significant and, therefore, is not shown in this graph.

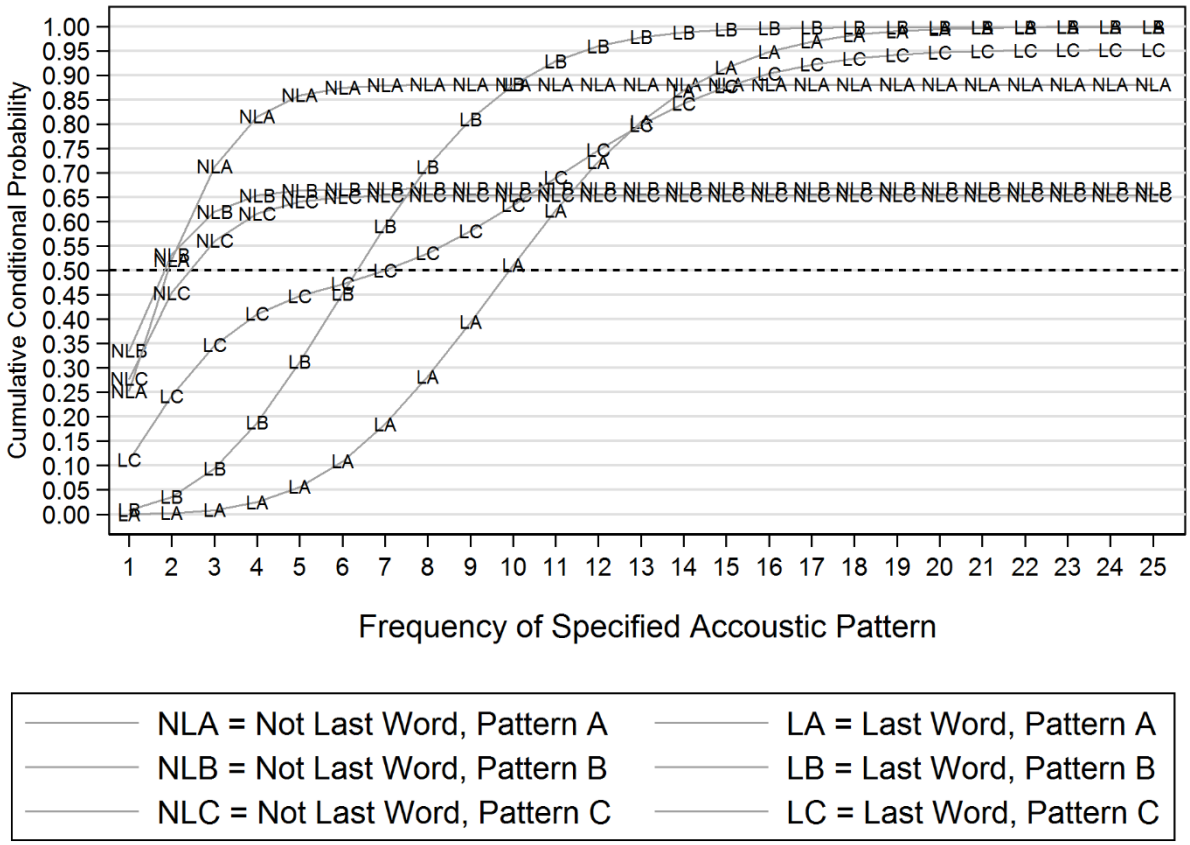

FIGURE 5. Margins plot of the cumulative conditional probabilities of occurrence of specific acoustic patterns of vocal fry by frequency level as predicted by the Word Position of the occurrence opportunity (i.e., Last Word vs. Not Last Word in sentence).

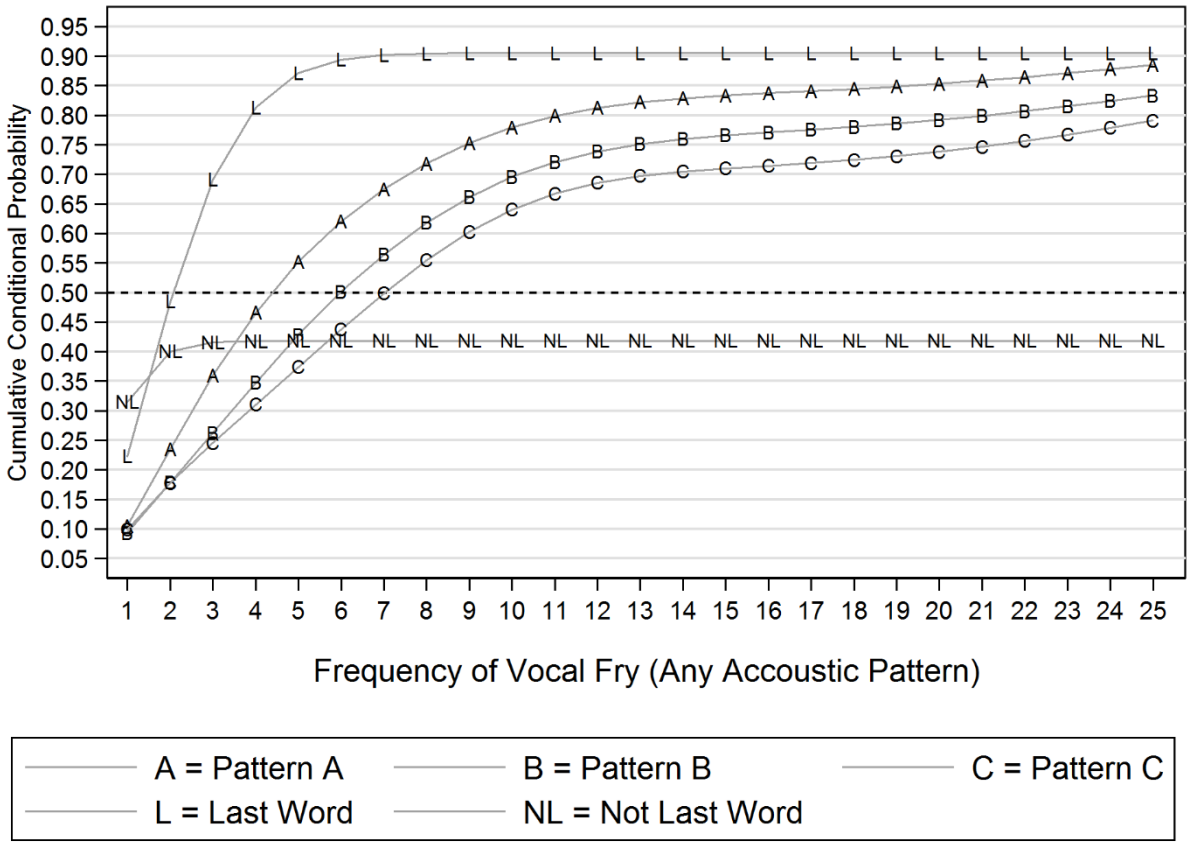

FIGURE 6. Margins plot of the cumulative conditional probabilities of occurrence of Vocal Fry Overall by frequency level as predicted by the Word Position of the occurrence opportunity (i.e., Last Word vs. Not Last Word in sentence) and specific acoustic pattern (i.e., A, B, or C).