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

*Nassima B. Abdelli-Beruh, †Thomas Drugman, and ‡R.H. Red Owl, *‡Brookville, New York; and †Mons, Belgium

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 American English (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 Kane-Drugman 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 Vocal 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 the present study showed that the occurrence frequencies of the three acoustic patterns and vocal fry overall (A + B + C) are greatest at the end of sentences 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.

Key Words: Acoustic patterns–Vocal fry–Automated detection of vocal fry–AE speakers–Gender differences.

INTRODUCTION

The present study is an extension of the work by Abdelli-Beruh et al,[1] Wolk et al,[2] and Drugman et al.[3] Abdelli-Beruh et al[1] and Wolk et 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] Using an automated detection algorithm based on the Kane-Drugman (KD) 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 findings of Drugman et al[3] regarding three acoustic patterns (A, B, C) associated with vocal fry. 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). The present study further addresses whether the frequencies of occurrences observed are consistent with the evidence of previous work based on perceptual data[1,2] that have indicated that vocal fry tends to be predominantly perceived at the end of sentences. Moreover, the present study 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 (ie, gender and word position) considered in the present study might be influenced by sentence length (operationalized as the number of words in each sentence), the analysis also incorporated sentence length as a covariate. The findings of the present study have implications for automated speech recognition programs and speech synthesis as they might help improve the naturalness of the synthesized tokens.[20]

Various terms coexist in the literature to describe this voice pattern (i.e., glottal fry). In the field of speech-language pathology, the prevalent terms are vocal fry, pulse register, creaky voice, and creak, whereas in the field of psycholinguistics, the terms "irregular phonation," "pulse phonation," "glottalization," or "laryngealization" are most frequently used.[1,2,7,16–18,21–27] It is not known whether there are substantial vibratory, acoustical, and auditory differences associated with these different labels but for Hollien,[23] pulse register is "undoubtedly" synonymous with vocal fry, glottal fry, creak, and strohbass (p. 2); Monsen and Engebreston[28] were of the opinion that vocal fry and creaky voice

Accepted for publication September 22, 2015.

From the *Department of Communication Sciences and Disorders, Post Campus at Long Island University, 720 Northern Boulevard, Brookville, New York 11548; †Faculté Polytechnique de Mons, TCTS Lab at University of Mons, Boulevard Dolez, 31, B-7000 Mons, Belgium; and the ‡Department of Educational Leadership and Administration, and Doctoral Program in Interdisciplinary Studies, Post Campus at Long Island University, 720 Northern Boulevard, Brookville, New York 11548.

Address correspondence and reprint requests to Nassima B. Abdelli-Beruh, Department of Communication Sciences and Disorders, Post Campus at Long Island University, 720 Northern Boulevard, Brookville, New York 11548. E-mail address: nassima.abdelli-beruh@liu.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2015.09.011

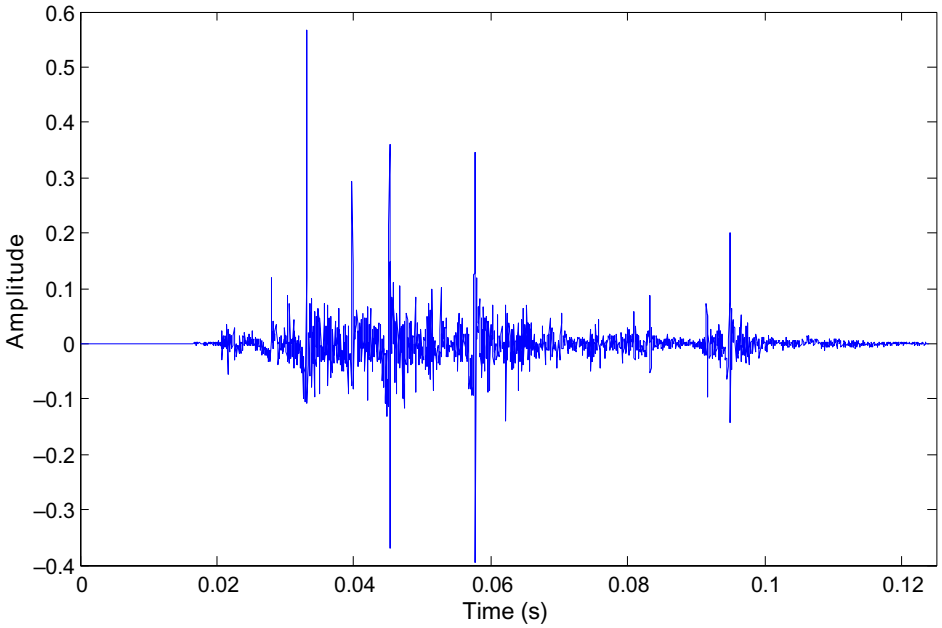

FIGURE 1. Shows a sample of an aperiodic pattern (Pattern A): Linear prediction residual for a typical Pattern A creaky voice segment. The waveform exhibits clear discontinuities with a highly irregular temporal structure. These peaks appear sporadically, and the inter-peak duration does not seem to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

can be used interchangeably; and Redi and Shattuck-Hufnagel[17] used glottalization and creak alternatively. Gordon and Ladefoged[29] used creaky phonation or vocal fry to describe the same vocal phenomenon. Imaizumi and Gauffin[30] considered creak to be a "special case" of vocal fry. Laver,[31] 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 the present study, the auditory criterion used in the KD approach is that of "rough quality with the additional sensation of repeating impulses."[32] The roughness voice quality and its concomitant multiple glottic pulses have been documented previously.[17,23,25,33,34]

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.[23,25,33,35–40] In addition to vocal fry's unique frequency range, Hollien[23] states that it also differs from modal and falsetto registers in its amplitude ranges (ie, lower than for the other registers) and in its frequency makeup as voice spectra show single or double pulses.[37] 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.[23,25,41]

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 (ie, glottal pulses irregular in duration from period to period); (2) creak (ie, glottal pulses of low F_0 accompanied by damping); (3) diplophonia (ie, glottal periods with systematic repetition in shape, duration, or amplitude); and (4) glottal squeak (ie, a sudden shift to relatively high sustained F_0 of very low amplitude) (p. 414). They reported that the latter type of glottalization (i.e., glottal squeak) occurred infrequently, while the former ones (aperiodicity, creak, diplophonia) have been previously documented by Huber.[42]

Drugman et al[3] analyzed excerpts of conversational speech samples produced by four AE, two Japanese, two Swedish, and two Finnish speakers. Using an automated detection program for creak based on features developed by Kane et al,[32] Drugman et al,[43] and Ishi et al,[44] Drugman et al 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 (Figure 1). In contrast to pattern A, pattern B is characterized by a more regular periodicity and the presence of two prominent excitatory peaks (Figure 2). The first peak is the Glottal Opening Instant 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 Glottal Opening Instant 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 GCI. Pattern C, like pattern B, is characterized by regular periods, which occur in modal voice, but it exhibits F_0 below 50 Hz. Unlike pattern B, pattern C does not present any secondary peak (Figure 3). Pattern C likely corresponds to what Ishi et al[44] 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.[17,20,44,45] They also reported that pattern A is the most frequently used by nine speakers out of 11, except for two male