Conversational Entrainment of Vocal Fry in Young Adult Female American English Speakers

***Stephanie A. Borrie and †Christine R. Delfino,** **Logan, Utah, and †Tempe, Arizona*

Summary: Objective. Conversational entrainment, the natural tendency for people to modify their behaviors to more closely match their communication partner, is examined as one possible mechanism modulating the prevalence of vocal fry in the speech of young American women engaged in spoken dialogue.

Method. Twenty young adult female American English speakers engaged in two spoken dialogue tasks—one with a young adult female American English conversational partner who exhibited substantial vocal fry and one with a young adult female American English conversational partner who exhibited quantifiably less vocal fry. Dialogues were analyzed for proportion of vocal fry, by speaker, and two measures of communicative success (efficiency and enjoyment).

Results. Participants employed significantly more vocal fry when conversing with the partner who exhibited substantial vocal fry than when conversing with the partner who exhibited quantifiably less vocal fry. Further, greater similarity between communication partners in their use of vocal fry tracked with higher scores of communicative efficiency and communicative enjoyment.

Conclusions. Conversational entrainment offers a mechanistic framework that may be used to explain, to some degree, the frequency with which vocal fry is employed by young American women engaged in spoken dialogue. Further, young American women who modulated their vocal patterns during dialogue to match those of their conversational partner gained more efficiency and enjoyment from their interactions, demonstrating the cognitive and social benefits of entrainment.

Key Words: Vocal fry–Conversational entrainment–Spoken dialogue–American women–Communicative success.

INTRODUCTION

Described perceptually as a "rapid series of taps like a stick being run along a railing,"[1(p98)] and originally considered a voicing characteristic associated with vocal pathology,[2] *vocal fry* has been touted as becoming increasingly prevalent in the conversational speech behaviors of young adult female American English speakers.[3] Although such increasing prevalence has yet to be empirically corroborated, there is certainly evidence that this voicing feature is frequently used this day and age by this population.[3–5] Sociocultural motives have been raised as a possible explanatory framework for the prevalence of vocal fry in the speech of young American women engaged in spoken dialogue[3]; however, the evidence regarding these motives is largely equivocal (eg, Refs. 3, 6). Here, we examine conversational entrainment, the natural tendency for people to modify their behaviors to more closely match their communication partner,[7] as one possible mechanism modulating the prevalence of vocal fry in the speech of young American women engaged in spoken dialogue.

Vocal fry—also commonly known as glottal fry, pulse or glottal register, laryngealization, or creaky voice—is typically defined as a series of discrete laryngeal excitations, with almost complete damping of the vocal tract between excitations.[2] The distinct vibratory pattern is generated with the arytenoid cartilages closely approximated,[8] and the resulting slow and aperiodic vibrations create a "creaking" or "popping" sound.[3,9] Vocal fry is a perceptually salient phenomenon, meaning that listeners can detect its presence with relative ease and with a high degree of accuracy. Michel and Hollien[10] reported that listeners were 95% accurate in distinguishing vocal fry from "harsh" phonation. Similarly, Blomgren and colleagues[11] reported that listeners were at least 95% accurate in categorizing speech samples as either vocal fry or modal (typical) phonation. Acoustically, vocal fry has been identified as occurring at the lower end of the fundamental frequency (F0) range.[2] In contrast to modal voice that occurs in the ranges of 85–180 Hz for men and 165–265 Hz for women,[12] vocal fry transpires around 7–78 Hz, a vocal range virtually identical for both men and women.[4,11,13] This gives rise to the notion that vocal fry has its own distinct vocal register—the glottal register.[13] Vocal fry has also been associated with increased measures of frequency and amplitude perturbation, termed jitter and shimmer, respectively.[4,14] In addition, x-ray data have revealed that during vocal fry, the vocal folds are very thick and relatively short,[15] and that airflow and subglottic air pressure is reduced.[11,16] Vocal fry can, therefore, be recognized perceptually, acoustically, and physiologically.

Recent studies have confirmed that the use of vocal fry is a woman-dominated trend in young American college students.[3–5] In an investigation of the effects of gender and nationality on the frequency of vocal fry in college students (aged 18–25 years), Yuasa[3] reported that the number of female American students who produced vocal fry (two thirds) during conversational speech with a same-sex interlocutor was significantly greater than that of both their male and their Japanese female counterparts. In the listener perception portion of the study, Yuasa identified that 78.9% of American college students (n = 175) reported they had heard creaky voice frequently used by women in the area where they resided—Northern California and Eastern Iowa. Similar to Yuasa, although using read passages as opposed to conversational speech, Wolk and colleagues[4] observed vocal fry use in approximately two thirds (n = 34) of female American speakers (also aged 18–25 years) and, in a follow-up study, the same research group found

Accepted for publication December 7, 2016.

From the *Department of Communicative Disorders and Deaf Education, Utah State University, Logan, Utah; and the †Department of Speech and Hearing Sciences, Arizona State University, Tempe, Arizona.

Address correspondence and reprint requests to Stephanie A. Borrie, Department of Communicative Disorders and Deaf Education, Utah State University, 1000 Old Main Hill, Logan, UT 84322-1000. E-mail: stephanie.borrie@usu.edu

Journal of Voice, Vol. 31, No. 4, pp. 513.e25–513.e32
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.12.005

that the rate of vocal fry use was approximately four times higher for female speakers than for male speakers.[5]

Sociocultural motives have been raised as a possible explanatory framework for the prevalence of vocal fry in conversational speech among young American women.[3] That is, that American women employ the creaky vocal quality in an attempt to project or convey a particular image of themselves. Yuasa,[3] in addition to reporting on the prevalence of vocal fry in relatively young, educated American women, collected listener perception data from 175 American college students regarding their subjective impressions of the use of vocal fry in this population. Using speech stimuli from a single speaker and a set of preselected rating characteristics, Yuasa reported that listeners identified speech produced with creaky voice as sounding fundamentally more educated, professional, genuine, and nonaggressive than speech produced with non-creaky modal voice. Dilley and colleagues[17] observed that female newscasters employed vocal fry more frequently than their male counterparts did. Although the study involved a small sample size (n = 5), the findings implicate that female newscasters may employ a creaky vocal quality in an attempt to evoke the authoritative connotations of masculinity (see Ref. 3 for a more detailed review). Collectively, these studies suggest that American women may use vocal fry in an endeavor to construct an identity that projects an educated and contemporary professional, capable of successfully competing with their male counterparts.

Opposing views of the impression of vocal fry, however, have also been reported. In a large-scale, nationwide study involving 800 listeners (18–65 years) and multiple voice exemplars, Anderson and colleagues[6] found that regardless of gender or age, vocal fry was interpreted negatively relative to a non–fry-speaking voice. American women exhibiting vocal fry were perceived as less competent, less educated, less trustworthy, less attractive, and less hirable. The authors concluded that the use of vocal fry may substantially damage a woman's job prospects. These findings are at odds with those of Yuasa[3] and others. Although differences in methodologies may contribute to the discrepancies noted, the existing data suggest that there is something more than just the desire to convey a particular image modulating the prevalence of vocal fry in the speech of young American women engaged in spoken dialogue.

Conversational entrainment describes the propensity for people to align their behaviors to more closely match those of their conversational partner. It transpires with no overt awareness[18] and has been evidenced in the alignment of both verbal (ie, acoustic-prosodic features, lexical choice, linguistic style) and nonverbal behaviors (eg, Refs. 19–22). For example, Levitan and Hirschberg[20] observed that communication partners entrained on a number of speech features including F0, intensity, jitter, shimmer, and speaking rate, whereas Louwerse and colleagues[21] reported behavioral alignment of facial expressions, manual gestures, and noncommunicative postures.

Recently, Borrie and Liss[23] demonstrated just how pervasive the entrainment phenomenon was, observing that healthy subjects unconsciously modified acoustic speech features to more closely match spoken stimuli, even when the features were pathologic in nature, as is the case with neurologically degraded speech, dysarthria. In Borrie and Liss's study, healthy subjects increased their rate of speech in response to productions from individuals with hypokinetic dysarthria (characterized by abnormally fast speech rate), decreased their rate of speech in response to productions from individuals with ataxic dysarthria (characterized by abnormally slow speech rate), and reduced their F0 (pitch) variation in response to productions from individuals with hypokinetic and ataxic dysarthria (both of which were characterized by abnormally reduced pitch variation). These findings suggest that the drive to entrain with one's communication partner is so ubiquitous, it transcends boundaries of typical norms—at least with regard to acoustic realizations of speech.

Indeed this pervasive pull to entrain during conversation is understandable when one considers the functional value of aligning behaviors with one's communicative partner. Perhaps most importantly, conversational entrainment has been shown to reduce the computational load of spoken processing and improve the effectiveness and efficiency with which information is exchanged.[24,25] Borrie and colleagues[7] observed that entrainment of acoustic-prosodic features of speech, including F0, intensity, and jitter, correlates with greater task success during a problem-solving task that required dialogue partners to work together, using verbal communication to solve; and Pickering and Garrod[26] suggest that coordinated language and behavior may facilitate mutual understanding and reflect a shared situational model between conversational dyads.

Entrainment has also been shown to regulate turn-taking dynamics[27] and is considered critical to the development of rapport, empathy, and intimacy between conversational partners (eg, Refs 18, 28). Lee and colleagues,[28] for example, observed that pitch entrainment predicts the likelihood of positive interactions in married couples discussing problems in their relationship, and Chartrand and Bargh[18] reported greater liking for a person who spontaneously mimics them. Gill goes as far as to comment that our ability "to synchronize with each other may be essential for our survival as social beings."[25(p111)] Indeed, conversational entrainment appears to function as a ". . .powerful coordinating device. . .to optimize comprehension, establish social presence, and create positive and satisfying relationships."[23(p816)] Thus, lack of entrainment, or inherent entrainment deficits, may result in conversational breakdowns.[7] Considering the negative ramifications of not entraining to one's communication partner, as well as that entrainment is realized even when the speech properties are disordered,[23] we postulate that all speech and voicing features are susceptible to this behavioral alignment phenomenon.

Here, we examine conversational entrainment as one possible mechanism modulating the prevalence of vocal fry in the speech of young American women engaged in spoken dialogue. To explore this proposed mechanism further, we also examine whether entrainment on this voicing characteristic affords functional communicative benefits in terms of more efficient[a] (ie, goal attainment in an accurate and timely manner) and enjoyable (ie, social connection and interaction satisfaction) conversation. Specifically, the following two key research

[a]We operationalize communicative efficiency using Duffy's definition in which communicative efficiency refers to "increasing the rate of communication without sacrificing intelligibility or comprehensibility."[34(p386)]

questions were addressed: (1) Do participants modify the frequency with which they use vocal fry depending on the level of vocal fry exhibited by their conversational partner? (2) Does vocal fry entrainment correlate with communicative success with regard to measures of efficiency and enjoyment? Based on robust models of conversational entrainment, it was hypothesized that participants would employ more vocal fry when interacting with a conversational partner presenting with substantial vocal fry than when interacting with a conversational partner presenting with minimal vocal fry. Further, in support of this mechanism, it was hypothesized that the more entrained a dyad was on their use of vocal fry, the more successful the dialogue would be in terms of efficiency and enjoyment.

METHODS

Participants

Twenty young, healthy females aged 18–29 years old (mean [M] = 20.61; standard deviation [SD] = 2.95) participated in the experiment. All participants were white American native speakers residing within the greater Phoenix area, who were assessed as speakers of Arizona dialect at the time of the investigation. As per self-report, participants presented with no history of speech, language, voice, or hearing problems. To obtain a participant pool that represented young adult female American English speakers, vocal quality was not accounted for in recruiting the participants. Participants were recruited from undergraduate and graduate classes at Arizona State University and were blinded to the specific purpose of the study. Institutional review board consent was obtained from all participants.

Speech stimuli

Speech stimuli were elicited from two young, healthy females, aged 23 and 27 years, blinded to the specific purpose of the study. Both individuals, termed *conversational partners*, were native speakers of American English, and like the participants recruited for the study, reported no history of speech, language, voice, or hearing problems. These conversational partners were explicitly selected based on their habitual vocal qualities, with one conversational partner presenting with substantial vocal fry (VF partner) and the other conversational partner presenting with relatively minimal vocal fry (NF partner)—validated by analysis of the percentage of vocal fry during a passage reading,[b] outlined below.

On separate occasions, the VF and NF partners were brought into the laboratory and seated in front of an industry-standard microphone (Shure SM58, Niles, IL) positioned at a mouth-to-microphone distance of 30 cm and connected to a digital audio recorder (TASCAM DR-40, Montebello, CA). The individuals were told that their task was to read aloud a standard passage reading, the Rainbow Passage,[29] using their typical speaking voice. This speech stimuli elicitation procedure resulted in the collection of two passage readings—one produced by the VF partner and the other produced by the NF partner. Audio recordings of the two passage readings were then transferred directly to a laboratory computer and labeled for subsequent analysis.

The presence of vocal fry was identified by perceptual analysis and established auditory criteria. The use of perceptual analysis was centered on the notion that entrainment involves perceptual feature detection,[30] that voice quality is largely an auditory-perceptual phenomenon,[31] and that listeners can perceptually detect the presence of vocal fry with relative ease and a high degree of accuracy.[10,11] A judge with extensive background in the study of voice and trained specifically in perceptual analysis of vocal fry annotated each of the passage readings for the presence of vocal fry, using *Praat* TextGrids.[32] The procedure required the judge to listen carefully to each audio recording and, guided by Blomgren and colleagues'[11] auditory criteria of (1) reduced and distinctly lowered pitch and (2) rough gravel-like quality, perceptually detect each episode of vocal fry, using boundary markers to mark the beginning and the end of each episode.

Once all vocal fry episodes were detected and labeled in the passage reading, the duration of each episode was calculated and transformed into a percent vocal fry (PVF) score by dividing the total time spent in vocal fry by the total time spent speaking (ie, total time taken to read the passage). The PVF scores for the NF conversational partner and the VF conversational partner were 2.39% and 18.42%, respectively. Thus, perceptual analysis confirmed that the speech of the VF partner was indeed characterized by substantially more habitual vocal fry than the speech of the NF partner.

Procedure

The experiment was conducted in a quiet research laboratory in the Speech and Hearing Department at Arizona State University. Participants were brought into the laboratory one at a time and seated in front of a microphone (Shure SM58) positioned at a mouth-to-microphone distance of 30 cm and connected to a digital recorder (Tascam DR-40). As per the same instructions given to the two speakers providing speech stimuli for the study, participants were asked to read aloud the Rainbow Passage using their typical speaking voice. Following the passage reading elicitation, the first conversational partner (VF or NF individual) was brought into the room. This partner was seated directly across from the participant, also in front of a microphone positioned at a mouth-to-microphone distance of 30 cm and connected to the same digital recorder. Stimuli and instructions for the dialogue task, the diapix task,[33] were then administered.

Each partner was given one of a pair of pictures and was instructed to hold their picture at an angle at which it would not be visible to the person sitting across the table from them. The pair of pictures depicted virtually identical scenes (ie, farm yard, beach trip), differing from one another by 10 small details (ie, color of boots, number of waves). The dyad was then told that their goal was to work together, simply by speaking to one another, to identify the 10 differences between the pair of pictures. They were instructed to complete the task as quickly and accurately as possible. No additional rules (ie, who could talk when) or roles (ie, giver, receiver) were given—dyads were free to verbally interact in any way they saw fit to achieve the

[b] A passage reading was used to control for linguistic structure and content, which has been shown to influence the use of vocal fry (eg, Refs. 3, 49, 50). Further, relative to other structured speech tasks (ie, sentence production), passage reading is considered to best approximate spontaneous speech (eg, Ref. 34).

instructed goal. The conversational partner was then thanked and asked to leave the room, at which point the participant was asked to complete a brief likeability rating scale with three questions pertaining to the enjoyment of the interaction (see Appendix). The second conversational partner (VF or NF individual) was then brought into the room. Identical setup and instructions were given to the new dyad; however, a different pair of pictures was used. Following this, the conversational partner was thanked and asked to leave the room, and the participant was asked to complete the likeability rating scale for the second interaction. The participant was then thanked and asked to leave the room.

Thus, five tasks—passage reading, dialogue with VF partner, likeability rating of dialogue with VF partner, dialogue with NF partner, and likeability rating of dialogue with NF partner—were elicited from each participant. The order of conversational partner was counterbalanced for the 20 participants so that 10 participants conversed with the VF partner first and 10 participants conversed with the NF partner first. The audio recordings of the passage readings and dialogues were then transferred to a computer and labeled for subsequent analysis accordingly.

Analysis

Vocal fry

The total data set consisted of 20 passage readings and 40 dialogues. The passage readings were annotated for the presence of vocal fry using the same auditory-perceptual procedure employed in the analysis of the conversational partners' passage readings (see Speech Stimuli section). A PVF score was generated for each participant's passage reading, yielding baseline data regarding their habitual level of vocal fry in that context. All participants displayed some habitual level of vocal fry in their passage reading, with a mean PVF score of 9.65%.

The audio recording of the dialogues were edited down to exactly 5 minutes of spoken dialogue and split into separate channels for the participant and the conversational partner. Using the same auditory-perceptual procedure used for analysis of the passage readings, each channel was annotated for the presence of vocal fry. The duration of each vocal fry episode was then tallied for the channel and divided by the total speaking time of the channel, resulting in a dialogue PVF score for each speaker. Thus, dialogue PVF scores were calculated separately for each participant and their conversational partner, resulting in 80 dialogue PVF scores.

Twenty percent of the passage readings and dialogues were remeasured by the original judge (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates regarding perceptual detection of vocal fry. Discrepancies between the remeasured data and the original data revealed that intra-judge and inter-judge agreement was high (all correlations $r > .95$), with only minor absolute differences.

Vocal fry entrainment score

A simple gross measure of entrainment was generated by subtracting the participants' dialogue PVF score from their conversational partners' dialogue PVF score. Thus entrainment, in this context, reflected the degree to which participants and their conversational partners employed similar amounts of vocal fry during a dialogue—the closer a score was to zero, the more entrained the dyad was. As there were 40 dialogues (20 with the NF partner and 20 with the VF partner), this resulted in 40 entrainment scores. These scores were then transformed to absolute values, labeled the *vocal fry entrainment score*, and used for statistical analysis.

Communicative efficiency score

The dialogue elicitation tool, the diapix task, grants us an objective measure of communicative efficiency[7,33] for each of the 40 dialogues. Recall that the diapix task required the conversational partner and the participant to work together as quickly and accurately as possible to identify the differences between a pair of pictures. Each pair of pictures contained 10 differences. Total number of differences identified in 5 minutes of spoken dialogue was then used as a simple, gross measure of communicative efficiency: relatively low and high numbers of identified differences indicate relatively high and low communicative efficiency, respectively. Thus, the *communicative efficiency score* is essentially an evaluation of how proficiently the dyad used verbal communication to collaboratively work through the demands of the dialogue task—in line with our operational definition in which communicative efficiency reflects "increasing the rate of communication without sacrificing intelligibility or comprehensibility."[34(p386)]

Communicative enjoyment score

A *communicative enjoyment score* was calculated for each of the 40 dialogues by summing the three ratings from the likeability rating scale performed by the participants following the dialogue task. The ratings included a rating of conversational partner likeability (ie, how much would you like to talk to X again?), social connection (ie, how well did you feel you connected with X, considering you had never met before?), and interaction satisfaction (ie, how satisfying was your interaction with X?). Thus, this score reflects how much the participant enjoyed the conversational experience, in terms of interaction satisfaction, partner likeability, and social connection.

RESULTS

Vocal fry use

Figure 1 reflects the mean PVF scores from the VF conversational partner (n = 20) and the PVF scores from the NF conversational partner (n = 20). An independent *t* test was used to confirm that during dialogue productions, the VF partner (M = 21.08, SD = 6.91) exhibited significantly more vocal fry than the NF partner did (M = 2.62, SD = .73); $t(38) = 11.89$, $P < 0.001$, Cohen d = 3.76.

Figure 2 reflects mean PVF scores for the participants when conversing with the VF partner (M = 15.01, SD = 4.67) and with the NF partner (M = 9.02, SD = 3.04). A paired samples *t* test revealed that during dialogue with the VF partner, participants exhibited significantly more vocal fry than during dialogue with the NF partner: $t(19) = 6.87$, $P < 0.001$, $d = 1.52$. Thus, participants used more vocal fry when conversing with a partner who

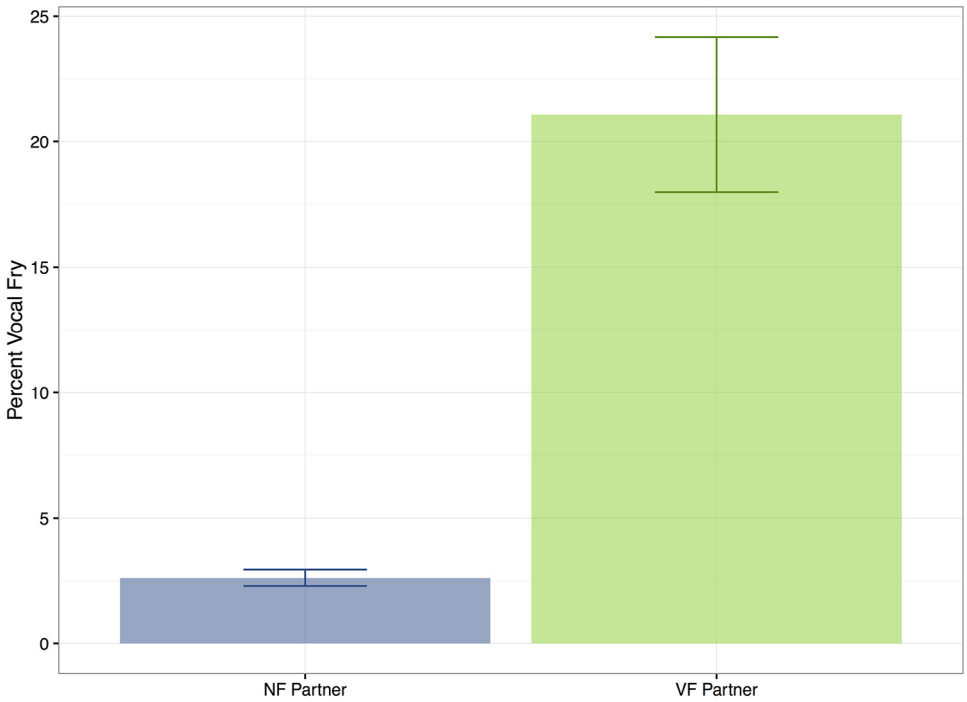

FIGURE 1. Mean percent vocal fry scores produced by two conversational partners—one who exhibited minimal vocal fry (NF partner) and one who exhibited substantial vocal fry (VF partner)—conversing with participants (n = 20).

exhibited substantial vocal fry than when conversing with a partner who exhibited minimal vocal fry.

Vocal fry entrainment and communicative efficiency

A Pearson product-moment correlation coefficient was computed to assess the relationship between vocal fry entrainment scores and communicative efficiency scores. The results revealed a significant negative correlation between vocal fry entrainment and communicative efficiency: $r(38) = -.48$, $P = 0.002$. This moderate relationship between the two variables of interest is also illustrated in Figure 3. Thus, these results demonstrate that conversational pairs who used similar quantities

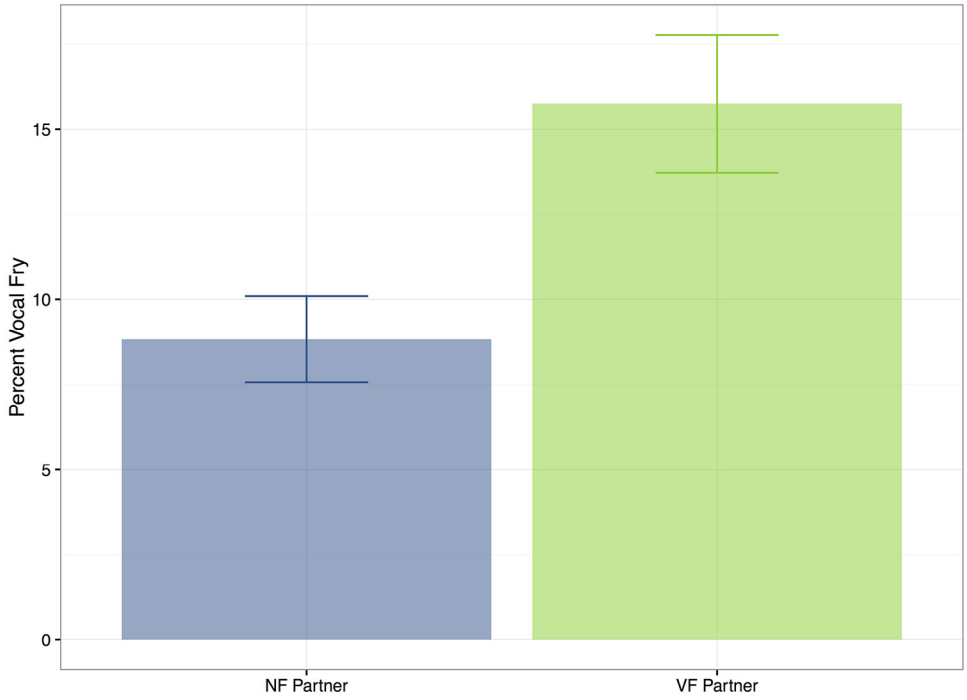

FIGURE 2. Mean percent vocal fry scores produced by the participants (n = 20) while conversing with two conversational partners—one who exhibited minimal vocal fry (NF partner) and one who exhibited substantial vocal fry (VF partner).

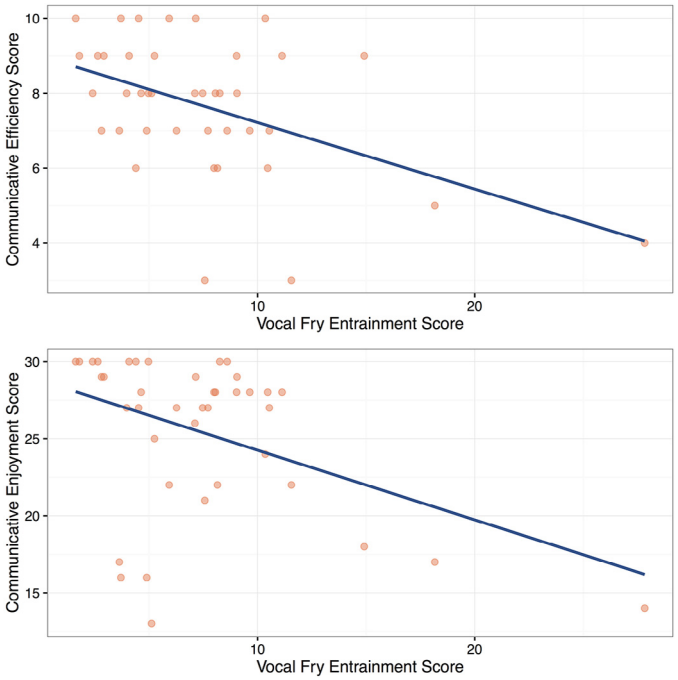

FIGURE 3. Vocal fry entrainment and communicative success. From top to bottom, the panels reflect significant correlations between vocal fry entrainment and indices of communicative efficiency and communicative enjoyment, respectively.

of vocal fry were more successful at using verbal communication to navigate the demands of the dialogue task.

Vocal fry entrainment and communicative enjoyment

A Pearson product-moment correlation coefficient was computed to assess the relationship between vocal fry entrainment scores[c] and communicative enjoyment scores. The results revealed a significant negative correlation between vocal fry entrainment and communicative enjoyment: $r(38) = -.43$, $P = 0.006$. This moderate relationship between the two variables of interest is illustrated in Figure 3. Thus, these results demonstrate that participants who more closely matched the levels of vocal fry used by their conversational partner experienced more enjoyment of the conversational experience.

DISCUSSION

Young American women attending college in the United States employed more vocal fry when interacting with a female conversational partner with substantial vocal fry than when interacting with a female conversational partner with quantifiably less vocal fry, a result evident in just five minutes of spoken dialogue with an unfamiliar conversational partner. This finding confirms our hypothesis that the pervasive communication phenomenon of conversational entrainment modulates, to some degree, the frequency with which the perceptually detectable creaky voice quality, vocal fry, is employed by young American women engaged in spoken dialogue. The idea that people modulate their speech behaviors to more closely match the speech behaviors of their conversational partner is certainly not new (eg, Refs. 18–28). However, corroborating entrainment of vocal fry is, and as such offers, a novel contribution to current perspectives regarding the prevalence of this voicing characteristic in the conversational speech behaviors of young American women.

Literature has reported on the widespread use of vocal fry among young American women, and sociocultural motives have been raised as the mechanism to explain this (eg, Refs. 3, 4). For example, it has been speculated that women employ a creaky vocal quality in an attempt to enhance their perceived desirability[36] or convey the image of a ". . .contemporary urban upwardly mobile woman."[3(p333)] As discussed, however, there is conflicting data on the image indexed by vocal fry. Although it certainly makes intuitive sense that sociocultural motives play some role in driving the prevalence of vocal fry in conversational speech behaviors, here we provide empirical evidence that the use of vocal fry in spoken dialogue involving young American women is modulated by the pervasive behavioral matching phenomenon of conversational entrainment. That is, female speakers unconsciously modulated the frequency with which they used vocal fry depending on the level of vocal fry displayed in the speech of their conversational partner. Accordingly, the behavioral alignment phenomenon of conversational entrainment should be considered, alongside other factors, to understand the use of vocal fry in conversational speech among young American women.

Further, the current study found that vocal fry entrainment among young American women tracked with an objective measure of communicative efficiency. Here, we observed that the more closely participants matched the level of vocal fry used by their conversational partner, the more efficient the pair was at using verbal communication to collaboratively work through the demands of the conversational task. This finding is consistent with previous literature reporting a relationship between conversational entrainment and task success in spoken dialogue involving both healthy (eg, Refs. 35, 37, 38) and clinical populations.[7,d] Literature in cognitive psychology suggests that linguistic alignment may be an external manifestation of a jointly built and understood situation model, a conceptual representation of relevant aspects of the situation under discussion,[24] and that mental model sharing can significantly improve communication efficiency (eg, Ref. 39). Lack of entrainment may reflect a breakdown in converging on a common situation model, reflected in less efficient communication.[40]

Vocal fry entrainment also tracked with a subjective measure of communicative enjoyment—a participant-based, social-index measure combining ratings of partner likeability, social connection, and interaction satisfaction. Here, we observed that the more closely a participant matched the level of vocal fry used by her conversational partner, the more the participant enjoyed the conversational experience. This finding is in line with the body of literature evidencing a link between entrainment and social outcomes, specifically that entrainment leads participants to like their

[c]Given that the entrainment score is the discrepancy between the participants' percent vocal fry score and their conversational partners' percent vocal fry score, numbers closer to zero reflect greater entrainment.

[d]Note, Borrie et al.[7] observed a positive relationship between entrainment and communicative success as their method for calculating entrainment, based on synchrony and proximity scores, was such that greater scores reflected greater entrainment.

conversational partners more,[18] experience enhanced feelings of connection,[41] and report a more satisfying interaction.[42] Communication accommodation theory (CAT), developed by Giles and colleagues,[43,44] is a theoretical framework to explain the social utility of entrainment. The basic premise of CAT is that entrainment—termed accommodation within this framework—serves as a means of regulating social differences between conversational participants. When conversational participants entrain to one another, social differences are reduced, and a more positive interaction ensues. Conversely, if conversational participants do not entrain to one another, social differences remain (and can increase), and the resulting interaction may be a less positive experience. Thus, according to CAT, entrainment facilitates social assimilation and a sense of belonging. The human need to belong has been evidenced as a fundamental motivation and driver of human behavior.[45] Indeed, the current finding, that vocal fry entrainment is associated with more enjoyable interactions, supports the notion that reducing differences between conversational participants promotes a more positive interactional experience.

Thus, in this study, we observed that vocal fry entrainment does occur and, further, it is predictive of communicative efficiency and communicative enjoyment—both, in their own rights, crucial elements of successful human communication. With all 20 participants in our study exhibiting some level of vocal fry while reading aloud a standard passage and engaging in spoken dialogue, we also validate earlier findings that vocal fry is indeed a prevalent voicing characteristic in the speech of young American women attending college in the United States.[3,4] This initial investigation into vocal fry entrainment sets the stage for a series of future experiments that include examining vocal fry entrainment within and across age, gender, and dialect categories. These studies are key given that young adult female American English speakers are reported to use significantly more vocal fry than their male counterparts,[3] that listener perception of vocal fry in regard to the perceived competence of the speaker is more negative in older adults relative to young adults,[6] and that dialectical variations in the rate of vocal fry use in American English have been observed.[46] Indeed, such investigations, in combination with the findings of the current study, will shed light on the complex interplay between the different drivers of behavioral change and whether, in some contexts, vocal fry may be immune to entrainment.

Another important area of investigation is whether an individual's perception regarding the image that vocal fry projects, as well as distinctive personality traits, modulates the extent to which the individual entrains his or her vocal fry use. Chartrand and Bargh[18] found that naturally empathetic individuals exhibited a greater degree of entrainment of body movements (ie, rubbing face and shaking foot), and Natale[47] demonstrated that an individual's social desirability, or "propensity to act in a social manner," is predictive of the degree that an individual will entrain to her conversational partner's vocal intensity. Future work might also conduct a more detailed entrainment analysis, investigating the accommodation and directionality of vocal fry over the course of the interaction. Given that dialogues were just 5 minutes in length, we simply calculated a global measure of entrainment. A local turn-by-turn analysis, however, would afford insight into the time course of vocal fry entrainment and whether, as postulated by others, entrainment increases over time.[48] Clinically, exploiting entrainment to support modification of voicing behaviors in patients may be warranted. These inquiries, among others, offer some future direction for continued work into this new avenue of vocal fry entrainment.

CONCLUSION

In sum, this study demonstrates that the frequency with which vocal fry is employed by young American women engaged in spoken dialogue is modulated, to some degree, by the ubiquitous tendency for people to align their behaviors to more closely match those of their conversational partner. This finding, therefore, adds an important dynamic to consider, alongside sociocultural motivations, when describing the prevalence of vocal fry in the conversational speech behaviors of young American women. Further, there is evidence that the more similar dialogue partners are in their use of vocal fry, the more successful their interaction is in terms of efficiency and enjoyment. Thus, this study also adds substantive support to the growing body of research demonstrating cognitive and social benefits of conversational entrainment, validating the behavioral alignment phenomenon as an essential element in successful human communication.

Acknowledgments

The speech samples used in this study were collected as part of a Master's thesis (second author) in the Motor Speech Disorders Laboratory (Liss) at Arizona State University, with support provided by the Neurological Foundation of New Zealand Grant 1119-WF and the National Institute on Deafness and Other Communication Disorders Grant R01 DC 6859. Ashely Hurst and LaeKin Clark, research assistants in the Human Interaction Laboratory at Utah State University (USU), are gratefully acknowledged for their assistance with the vocal fry analysis employed in the current study.

APPENDIX

Likeability Rating Scale

This is a completely anonymous rating scale. Its sole purpose is to collect subjective feedback about your most recent conversational experience (X = conversational partner). Your answers will be kept confidential and will be seen only by members of the research team. Please answer honestly and thoughtfully.

On a scale from 1 to 10, with 1 = *Not At All*, 5 = *Neutral*, and 10 = *Completely*, please rate the following questions regarding your most recent interaction.

(1) How much would you like to talk to X again?
1 2 3 4 5 6 7 8 9 10

(2) How well did you feel you connected with X, considering you had never met before?
1 2 3 4 5 6 7 8 9 10

(3) How satisfying was your interaction with X?
1 2 3 4 5 6 7 8 9 10

REFERENCES

1. Catford JC. *Fundamental Problems in Phonetics*. Edinburgh, SCT: Edinburgh University Press; 1977:98–103.
2. Hollien H, Moore P, Wendahl RW, et al. On the nature of vocal fry. *J Speech Lang Hear Res*. 1966;9:245–247.

3. Yuasa IP. Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women? *Am Speech*. 2010;85:315–337. doi:10.1215/00031283-2010-018.
4. Wolk L, Abdelli-Beruh NB, Slavin D. Habitual use of vocal fry in young adult female speakers. *J Voice*. 2012;26:e111–e116. doi:10.1016/j.jvoice.2011.04.007.
5. Abdelli-Beruh NB, Wolk L, Slavin D. Prevalence of vocal fry in young adult male American English speakers. *J Voice*. 2014;28:185–190. doi:10.1016/j.jvoice.2013.08.011.
6. Anderson RC, Klofstad CA, Mayew WJ, et al. Vocal fry may undermine the success of young women in the labor market. *PLoS ONE*. 2014;9:e97506. doi:10.1371/journal.pone.0097506.
7. Borrie SA, Lubold N, Pon-Barry H. Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Front Psychol*. 2015;6:1187. doi:10.3389/fpsyg.2015.01187.
8. Ladefoged P, Johnson K. *A Course in Phonetics*. 7th ed. Stamford, CT: Cengage Learning; 2015.
9. Hollien H, Wendahl RW. Perceptual study of vocal fry. *J Acoust Soc Am*. 1968;43:506–509.
10. Michel JF, Hollien H. Perceptual differentiation of vocal fry and harshness. *J Speech Lang Hear Res*. 1968;11:439–443. doi:10.1044/jshr.1102.439.
11. Blomgren M, Chen Y, Ng ML, et al. Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *J Acoust Soc Am*. 1998;103:2649–2658. http://dx.doi.org/10.1121/1.422785.
12. Titze IR. *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall; 1994.
13. Hollien H, Michel JF. Vocal fry as a phonational register. *J Speech Lang Hear Res*. 1968;11:600–604. doi:10.1044/jshr.1103.600.
14. Horii Y. Vocal shimmer in sustained phonation. *J Speech Lang Hear Res*. 1980;23:202–209. doi:10.1044/jshr.2301.202.
15. Allen EL, Hollien H. A laminagraphic study of pulse (vocal fry) register phonation. *Folia Phoniatr (Basel)*. 1973;25:241–250. doi:10.1159/000263709.
16. McGlone R, Shipp T. Some physiologic correlates of vocal fry phonation. *J Speech Lang Hear Res*. 1971;14:769–775. doi:10.1044/jshr.1404.769.
17. Dilley L, Shattuck-Hufnagel S, Ostendorf M. Glottalization of word-initial vowels as a function of prosodic structure. *J Phon*. 1996;24:423–444. doi:10.1006/jpho.1996.0023.
18. Chartrand TL, Bargh JA. The chameleon effect: the perception-behavior link and social interaction. *J Pers Soc Psychol*. 1999;769:893–910. http://dx.doi.org/10.1037/0022-3514.76.6.893.
19. Branigan HP, Pickering MJ, Cleland AA. Syntactic co-ordination in dialogue. *Cognition*. 2000;75:B13–B25. doi:10.1016/S0010-0277(99)00081-5.
20. Levitan R, Hirschberg J. 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In R. Pieraccini & A. Colombo (Eds.), Proceedings of Interspeech 2011. Brisbane: International Speech Communications Association.
21. Louwerse MM, Dale R, Bard EG, et al. Behavior matching in multimodal communication is synchronized. *Cogn Sci*. 2012;36:1404–1426. doi:10.1111/j.1551-6709.2012.01269.x.
22. Niederhoffer KG, Pennebaker JW. Linguistic style matching in social interaction. *J Lang Soc Psychol*. 2002;21:337–360. doi:10.1177/026192702237953.
23. Borrie SA, Liss JM. Rhythm as a coordinating device: entrainment with disordered speech. *J Speech Lang Hear Res*. 2014;57:815–824. doi:10.1044/2014_JSLHR-S-13-0149.
24. Garrod S, Pickering MJ. Why is conversation so easy? *Trends Cogn Sci*. 2004;8:8–11. doi:10.1016/j.tics.2003.10.016.
25. Gill SP. Rhythmic synchrony and mediated interaction: toward a framework of rhythm in embodied interaction. *AI Soc*. 2012;27:111–127. doi:10.1007/s00146-011-0362-2.
26. Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. *Behav Brain Sci*. 2004;27:169–225. http://dx.doi.org/10.1017/S0140525X04000056.
27. Wilson M, Wilson TP. An oscillator model of the timing of turn-taking. *Psychon Bull Rev*. 2005;12:957–968.
28. Lee C, Black M, Katsamanis A, et al., 2010. Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples. In INTERSPEECH (pp. 793–796).
29. Fairbanks G. *Voice and Articulation Drillbook*. 2nd ed. New York: Harper & Row.; 1960:124–139.
30. Phillips-Silver J, Aktipis CA, Bryant G. The ecology of entrainment: foundations of coordinated rhythmic movement. *Music Percept*. 2010;28:3–14.
31. Gerratt BR, Kreiman J. Toward a taxonomy of nonmodal phonation. *J Phon*. 2001;29:365–381. doi:10.1006/jpho.2001.0149.
32. Boersma P, Weenink D. 2016. Praat: doing phonetics by computer (Version 6.0.09) (Computer program). Available at: http://www.praat.org/. Accessed February 2, 2016.
33. Van Engen KJ, Baese-Berk M, Baker RE, et al. The wildcat corpus of native- and foreign-accented English: communicative efficiency across conversational dyads with varying language alignment profiles. *Lang Speech*. 2010;53:510–540. doi:10.1177/0023830910372495.
34. Duffy JR. *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. 3rd ed. St. Louis, MO: Elsevier Mosby; 2015.
35. Nenkova A, Gravano A, Hirschberg J. 2008. High frequency word entrainment in spoken dialogue. In Proceedings of the 46th annual meeting of the Association for Computational Linguistics on Human Language Technologies: Short papers (pp. 169–172). Association for Computational Linguistics.
36. Pennock B. 2005. The changing voice of women. In Actas del XXVII Congress Internacional de AEDEAN (pp. 407–415).
37. Reitter D, Moore JD. 2007. Predicting success in dialogue. In Proceedings of 45th Meeting of the Association of Computational Linguistics (pp. 808–815). Prague, Czech Republic: Association for Computational Linguistics.
38. Friedberg H, Litman D, Paletz SBF. 2012. Lexical entrainment and success in student engineering groups (pp. 404–409). Spoken Language Technology Workshop (SLT). Miami, FL: IEEE. doi:10.1109/SLT.2012.6424258.
39. Mathieu JE, Heffner TS, Goodwin GF, et al. The influence of shared mental models on team process and performance. *J Appl Psychol*. 2000;85:273–283. http://dx.doi.org/10.1037/0021-9010.85.2.273.
40. Branigan HP, Pickering MJ, Pearson J, et al. The role of beliefs in lexical alignment: evidence from dialogs with humans and computers. *Cognition*. 2011;121:41–57. doi:10.1016/j.cognition.2011.05.011.
41. Marsh KL, Richardson MJ, Schmidt RC. Social connection through joint action and interpersonal coordination. *Top Cogn Sci*. 2009;1:320–339. doi:10.1111/j.1756-8765.2009.01022.x.
42. Bernieri FJ, Reznick JS, Rosenthal R. Synchrony, pseudosynchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions. *J Pers Soc Psychol*. 1988;54:243–253. http://dx.doi.org/10.1037/0022-3514.54.2.243.
43. Giles H, Mulac A, Bradac JJ, et al. Speech accommodation theory: the first decade and beyond. In: McLaughlin CM, ed. *Communication Yearbook*. Newbury Park, CA: Sage; 1987:13–48.
44. Giles H, Coupland J, Coupland N. Accommodation theory: communication, context, and consequences. In: *Contexts of Accommodation: Development in Applied Sociolinguistics*. New York, NY: Cambridge University Press; 1991:1–68.
45. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull*. 1995;117:497–529. http://dx.doi.org/10.1037/0033-2909.117.3.497.
46. Byrd D. Relations of sex and dialect to reduction. *Speech Commun*. 1994;15:39–54. doi:10.1016/0167-6393(94)90039-6.
47. Natale M. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *J Pers Soc Psychol*. 1975;32:790–804. http://dx.doi.org/10.1037/0022-3514.32.5.790.
48. Jain M, McDonough J, Gweon G, et al., 2012. An unsupervised dynamic Bayesian network approach to measuring speech style accommodation. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 787–797). Association for Computational Linguistics.
49. Kreiman J. Perception of sentence and paragraph boundaries in natural conversation. *J Phon*. 1982;10:163–175.
50. Redi L, Shattuck-Hufnagel S. Variations in the realization of glottalization in normal speakers. *J Phon*. 2001;29:407–429.