Vocal fry occurs when an individual drops the fundamental frequency of their voice below its normal register and emits a low, growly, creaky tone of voice. Vocal fry, or creaky voice, has been widely acknowledged by media outlets as the generational voicing of millennial women. As a female-specific vocal pattern, it is often compared to valley girl speak (adding ‘like’ superfluously) or upspeak (assertion delivered with a question’s intonation). Like ‘like’-ing and upspeak, vocal fry is often derided by cultural critics, such as NPR’s Bob Garfield, who in a recent interview on Slate magazine’s language blog called the phenomenon “annoying, I mean, really annoying,” “vulgar,” and “repulsive.” [6] Garfield’s sentiments echo reactions discussed in other national media such as The Atlantic [14] and This American Life [7].

Critics of fry often emphasize that it is an exclusively female vocal pattern, and some, such as those calling recently to complain to This American Life about female reporters frying [7], say the voicing is so distracting that they cannot process the semantic content of the speech act associated with it. In this paper, we use sociolinguistics and contemporary philosophical theories of speech acts to develop a framework for understanding this pattern of reaction to vocal fry. We argue that when fry is heard as annoying and distracting, it is because the hearer interprets the speaker as echoing an utterance from a position of authority to which she is not entitled. We suggest that this reaction, coupled with the
tendency of complainants to attribute it only to female fryers, encodes sexist attitudes.

We begin in Section 1 by surveying the history of vocal fry as an acoustic and sociolinguistic phenomenon. We show that while it is neither a new voicing nor exclusively female, there is a sociolinguistic association, the frequency code, between female fry and imitations of masculinity. We argue that this imitative quality cannot be understood through existing accounts of injustice in speech acts, but that it can be explicated through the principles of echoic utterance as outlined in Sperber and Wilson’s [24] echoic account of irony. We construct a framework that explains not only why women are judged negatively for vocal frying, but why these particular negative judgments minimize asserted content from their utterances. We conclude with a few remarks on the unconscious gender bias embedded in the annoyance of some hearers of vocal fry utterances.

1 Vocal Fry in Sociolinguistics: A Historical Overview

While it has been getting new media attention of late, vocal fry has been studied as type of phonation for over 50 years. In this section, our historical survey demonstrates that fry is neither a novel phenomenon in the English-speaking world, nor is it exclusively produced by female speakers. Nonetheless, there is some evidence that a) it is becoming more frequent among some communities of young American women, and that b) female fry may be viewed as a different sort of speech act than male fry.

Some early research treated fry as a type of voice disorder, but it has been studied as a non-clinical voicing since at least the 1960s.[3, 12] Fry is produced by a slackening of the vocal cords such that they vibrate irregularly and lower the fundamental frequency of voice pitch. It is usually located on a spectrum of voicing types produced by voluntary changes in the anatomy of the vocal cords during speech. Such spectra (e.g. [3, 10]) vary in terminology and specificity, but generally identify at least five phonation types, arranged
from most constricted to least constricted vocal cords: whisper, breathy voice, modal (i.e. “normal”) voice, vocal fry, and full glottal stop (as in, the sound produced during the hyphenation in an utterance of “uh-oh”). Falsetto is sometimes included on such spectra between breathy and modal voices.

Early descriptions of the sound produced by vocal fry reveal unpleasant metaphorical associations that persist in contemporary complaints about the annoying quality of the voicing. Linguist J.C. Catford, for instance, writes of fry, “The auditory effect is of a rapid series of taps, like a stick being run along a railing.”[3, p. 32] More recently, definitions of fry are given in terms of acoustic mechanisms, as in “Pitch, which has been observed to be extremely low, is controlled by aerodynamic factors and not by varying the longitudinal tension. The [fundamental frequency] and amplitude variation of consecutive glottal pulses is further known to be very irregular. ‘Double pulsing’ is also a frequent characteristic of creak, where two pulses, with different amplitude and duration, occur within what appears to be one cycle.” [8, p. 13]

These descriptions are associated with the detection of fry by spectrograms, which measure patterns of vocal pitch frequency and intensity over time. Contemporary sociolinguistic and linguistic-anthropological analysis of fry in English-speaking subjects typically uses spectrogram measurements to detect instances of fry, then determines whether the speaker’s subject of discussion, speaking context, personal identity, or other features correlate with the presence or absence of frying. For instance:

- Esling [4] correlates higher social status with more frequent use of fry in Edinburgh speakers,
- Stuart-Smith [23] correlates it with male speakers in Glasgow,
- Fought [5] identifies it as a central feature of Southern California Chicano English,
Carpenter [2] association it with projections of masculinity in teenage males,

and Mendoza-Denton argues that it “assists [Chola/Chicana] gang girls with the construction of a hardcore persona” when telling fight narratives. [19, p. 266]

In an influential article linking phonation pitch to gender expression, Ohala [20] develops what he terms the “frequency code,” which links higher-pitched phonation types like falsetto with femininity and lower pitched phonation like frying with masculinity. The frequency code is implicit or explicit in many of the articles and books listed above, and recent work on gender and fry typically either reinforces or calls into question the associations of the frequency code. According to the frequency code, a woman speaking with creak will sound like a man; likewise, a man speaking in falsetto will sound like a woman.

The frequency code is often rationalized biologically, appealing to the idea that men typically have bigger larynxes and thus lower fundamental frequencies in their phonation. Sociolinguist Robert Podesva, whose research program centers around investigating associations between phonation types and personal identity markers, puts the point thus:

The creaky voice pattern may arise from iconic associations between creaky voice and masculinity — the low pitch characterizing creaky voice is interpreted as resembling masculinity, due to gross tendencies for men to have lower pitched voices than women. This indexical association can be recruited to link creaky voice to stances conventionally associated with men, like toughness, at higher orders of indexicality. . . . Similar ideological processes link falsetto, and its characteristically high pitch, to femininity.[21, p. 427]

The frequency code rationalizes certain features of the echoic account of female fry that we develop below. The misfiring that occurs in cases of dismissive reactions to female fry like the ones referenced in the introduction may originate in confusions of (sociolinguistic)
utterance indexicality. Masculine-indexed utterances issuing from female speakers present conflicted messages of what the speech act is, in ways that cannot be explained by discursive injustice alone.

It is worth noting that frequency-code explanations of female fry as masculine need not appeal to speaker intent; that is, a woman need not be consciously indexing masculinity to produce fry that is interpreted by an audience as echoing masculine voicing. A variety of recent studies tie fry not to masculinity but to other indexical and non-indexical utterance markers, including speech about emotional topics[17], increased intimacy of a verbal exchange[22], and demonstrations of authority by college women in Virginia[16]. Most of the sociolinguistic research on fry centers around the speaker’s utterances, rather than audience perception of those utterances, but a few studies (e.g. [9, 26, 17]) have investigated audience reactions. In these, female fry is correlated with a range of affects from association with specific emotions (boredom [9], non-aggression [26]) to better education [26]. In their [17], Loss and Zold argue that young women (under 40) use fry to communicate authority, while older women (over 40) use it to communicate about emotional topics. This corroborates Lefkowitz’s [16] study of college-age women in Virginia, for whom fry was a marker of authoritative speech. It is not discussed whether fry in these latter instances is used intentionally, nor whether young women fry users associate their use with sounding “like a man.”

2 Implications of the Frequency Code for Contemporary Female Fry

Many linguists believe that the reason why women vocal fry comes from a desire for authority or to be seen as authoritative.[25] Studies have shown that women who are a part
of a male dominated work place are more likely to vocal fry, because they are mimicking the deep voices of their male coworkers.[1] Vocal frying has been seen as an attempt made by women to gain authority with their peers by sounding more like them. Some critics of female fry say that women should change their voices to sound more professional, [25] while others say that female fry is disempowering and disadvantageous. [13] These negative judgments on female fry demonstrate that users are being judged by how they sound, rather than by the content of their speech.

The frequency code associates lower fundamental frequency with masculinity. But many hold additional expectations about how speakers of different genders should sound. For instance, Lowen [18] discusses the expectations that women should sound sweet, calming, and docile while men should sound authoritative, confident, and commanding. When women or men don’t sound as they are expected to, they are seen as violating their vocal genders.

Though there is a framework that says that women and men should sound a certain way, this does not provide a full explanation of what is going on in instances where the use of vocal fry is judged—the problem goes beyond a speaker merely breaking with the gender norm. In these instances, the audience interprets the frequency code violation as license to make certain types of judgments about the speaker that lead to dismissing any claims the speaker might make. With fry in particular, only women are judged negatively, although both men and women exhibit fry. [11] This reaction pattern, which is clearly demonstrated in the media attention to fry discussed in the introduction to this paper, shows a trend where women are not being judged by the content of their speech, but by their deviation from the frequency code and their failure to conform to expectations of how women should sound.

Utterances given as reactions to other utterances fall into two broad categories: content-
based and non-content-based. A content-based reaction is one that engages the semantic content or material of a speech act, where a non-content based reaction is one that does not engage the semantic content presented in a speech act, but engages in something else about the speech act (such as the speaker). For example, if a speaker says ‘my dog is probably the best dog,’ content-based responses might include, ‘Yes, he is definitely the best dog,’ or, ‘You know, I’m kind of partial to my own dog.’ A non-content based response to the utterance ‘my dog is probably the best dog,’ would be a response like, ‘I can’t hear you,’ or, ‘there is a spider on you!’

Non-content based responses can arise from a variety of conditions. The most innocuous generating condition for a non-content-based response is failure to hear or understand the original utterance, as when a respondent asks a speaker to repeat a question or phrase. Additionally, non-content-based responses can draw attention to extra-linguistic features of the conversation, such as in the case of the spider response above. However, non-content-based responses can also arise from respondent bias against a speaker, such as in cases where a speaker’s affect, accent, or other personal identity markers are the subject of the respondent’s response. For instance, journalist Jessica Grose writes about interviewing an older man for Businessweek who responded to her question by telling her she sounded like his granddaughter. Ms. Grose’s respondent uttered a non-content-based response.[11] The interviewee did not respond directly to the content of her question, but instead exclusively to her voice.

Cultural backlash against female fry is a pattern of non-content-based responses to violations of the frequency code and the norms of voice quality. In Section 4, we argue that ignoring the content a woman is trying to communicate because of the tone of her voice seems not only inexcusable, but chauvinistic. But in order to better understand how this pattern of non-content-based responses arises we first develop a framework to explain how
and why they occur.

3 The Echoic Account and Vocal Fry

When we engage in conversation, we are usually making content-based responses—directly engaging with what the speaker is saying. When we ask a question or make an assertion, the response, most of the time, is interacting with the semantic content of the speaker’s utterance. Though non-content based responses are usually not the most appropriate response to a speaker for the purposes of driving conversation forward, there are some situations when they are the most reasonable. When a speaker is engaging in oppressive speech, for instance, audiences may respond not to the content of the individual speech act but to the person’s values. For example, if someone says ‘all women are bad drivers,’ and the recipient of this utterance does not agree, they can (rightfully) decide to ignore the speaker and judge the speaker as unworthy of engagement. Content and non-content based responses can get confusing when recipients respond with content that is potentially spurred by the conversation. Using our example above, if the recipient of the comment responded by saying, ”are you racist, too?” this comment, though interacting with the content of the speech of the speaker, is not directly responding to the speaker’s claims. This would be considered a borderline case of content/non-content based responses. Ultimately, when determining whether a response is content or non-content based, it is best to take each instance pragmatically, case by case.

Another circumstance where this is a plausible reaction to a speaker is a scenario where the audience does not feel comfortable engaging the material put forth by the speaker, based on who the speaker is, rather than what they said in particular. If someone makes a racist, sexist, or generally derogatory claim, the audience can choose to ignore the comment and make a non-content based response to switch or ignore the topic. For example, if a
programmer approaches a philosopher of language and says, ‘isn’t all philosophy of language just the computational study of information,’ the philosopher has the ability to judge the speaker and refuse to engage. The philosopher may realize that regardless of the truth content of the speech act, no useful or friendly conversation will follow from this close-minded type of person. It is who is speaking that is being judged, not the content of the utterance. Contrast this with the same question asked by an undergraduate in her first philosophy class, who is more likely to be making a genuine request for information.

What is important to note from this contrast is that sometimes utterances of the same apparent semantic content serve as different speech acts and play different conversational roles to differing effect. If a respondent judges a speaker as willing to engage with her utterances, then she may think it worthwhile to continue a conversation. But, if she judges the speaker to be the kind of person who is closed off to discussion, she may not think it worthwhile to reason with them.

Among established accounts of judgment and misjudgment of speech acts in the philosophy of language literature, a few address the systematic misjudgment of speech act types on the basis of speaker identity. Here we consider one such account, Rebecca Kukla’s discursive injustice (DI). DI describes an observed phenomenon wherein the speech acts of women and minorities are routinely judged as less forceful than the speakers report to have intended.[15] Kukla gives the example of a female boss whose employees routinely misread her work demands as requests.

Instances of discursive injustice illustrate a type of incorrect judgment by a respondent to a speaker’s utterance, and it accounts for a number of the features of the annoyance response to female fry. In instances of DI against women, as in negative responses to female fry, the gender identity of the speaker influences the respondent’s response. Further, in both cases, the respondent takes the speaker less seriously than if the speaker were male.
However, DI typically does not leave room for non-content-based responses; while the respondent sees the speaker as speaking less forcefully than if she were male, the respondent still engages with the semantic content of the speaker’s utterances. In at least some cases of negative response to female fry, such as the one discussed in the previous section, the response is non-content-based.

Additionally, in DI, the speaker is not violating the norms of expected voice quality nor the frequency code. Respondents generating sexist DI do not react to the identity of the speaker because the utterance fails to conform to expectations of sweet, docile female voices, nor because it affirms those expectations, but simply because female victims of DI are universally assigned less than their share of authority. In contrast, respondents who react negatively to female fry confront the speech act’s violation of the frequency code, and the negative affect brought about by this violation is often implicit—and occasionally explicit—in their responses.

Since DI cannot fully account for the negative response to female fry, we look instead toward pragmatics. Broadly speaking, semantics is about what is said and pragmatics is what is meant, other than what is said. Pragmatics has the explanation we are looking for because it is in pragmatics that we find explanations for apparent misunderstandings, from the respondent, of what is meant by the speaker. In particular, we focus on the pragmatic account of echoic utterances from Sperber and Wilson. [24]

Sperber and Wilson state that an utterance is used to represent a thought of the speaker’s that it resembles in content. They establish that echoic utterances convey thoughts that are not directly about an actual or possible state of affairs, but about another thought that it resembles in content. [24] This is defined in contrast to descriptive utterances, which represent a thought about an actual or possible state of affairs. On their account, echoic utterances are attributive. An attributive utterance, “is not directly
about a state of affairs, but about another thought that it resembles in content, which the speaker attributes to some source other than herself at the current time."[24] An example that Sperber and Wilson use to attributive echoic utterance is:

a. John phoned his wife and told her that *the train was about to leave.*

In a, the italicized portion is the part of the sentence that is attributive. Perhaps John and his wife had made a previous plan to take a train, but they were about to miss it. John phoned her to inform her of this misfortune, but did not have to explain the entire situation; because his utterance was attributive, and his wife recognized it as such, the utterance made sense.

Sperber and Wilson argue that depending on the vocal tone of an utterance, the way that an utterance is delivered can influence the way that the utterance is understood by the audience. In the case, vocal tone can reign over the attributive utterance—the way that the attitude that is being referenced is understood by the audience. Vocal tone, according to Sperber and Wilson, can expose either a way that the speaker wishes the attitude to be understood as (optative) or mimic the attitude itself.[24] All of this is to say that the way that an utterance is delivered can influence, but not wholly determine, the meaning of the utterance.

For example, consider the exchange one of us had with our partner on Saturday morning:

J It’s such a nice day for a hike.

M Let’s go on a hike, then.

During the hike, it started pouring rain. So in the car on the way home, covered in mud, one of us said:
M’ It’s such a nice day for a hike.

In M’, both the utterance J and the attitude of that utterance are referenced to contribute to what is ultimately understood as sarcasm. In this case, M’ echoes J. In echoic ironic statements like this one, the attitude referenced in the echoic utterance is usually the opposite of what the speaker is trying to convey. In the case of the failed hiking trip, the attitude referenced is the positivity surrounding a hiking trip. The audience then derives that what this utterance and referenced attitude are meant to express—what kind of reaction to the previous utterance the speaker wants the audience to grasp from what they say—that the hiking trip has failed.

Sperber and Wilson devised the echoic account to explain ironical utterances, but we believe this account can be expanded to help understand negative reactions to female fry. Like in cases of irony, in cases of negative reaction to female fry, the respondent may see a reference to an earlier utterance. Though the speaker’s intent during a negative fry reaction is not analogous to the speaker’s intent in ironic echoes, the respondent hears an echo of masculine voicing and witnesses a reference to an attitude—in this case, a personal identity as a man—that the speaker does not intend to convey. Because of its violation of the frequency code, some respondents may judge any instance of female fry as inherently echoing male voicing.

In cases of negative response to female fry, we propose that when people react to women deploying vocal fry, the audience processes the speaker’s utterance as one that is echoic. The attitude that the audience sees being referenced could be that of an authoritative male figure. But, since a woman is speaking, the audience sees this utterance as mocking the male figure. Thus the speaker is not saying something worthy of engaging with, but is saying something worth commenting on in a different way.

The audience may hear female fry and identify it (consciously or not) as an echoic
utterance type, referencing an attitude the speaker does not possess. Though there may be notions at play that dictate how someone should sound, this may not be enough for them to think that the woman is trying to mock a person with authority. But, Sperber and Wilson mention that vocal tones may help guide how the interpretation of an attitude is meant to be taken. The audience is more influenced to think that the woman is mocking an authority figure because her tone of voice is conducive to that interpretation. The possibilities for what vocal fry may signify are vast and disjunctive, as discussed in Section 1. The audience now interprets this utterance as mocking because the tone, in correlation with the context and utterance type, conveyed by the speaker is one that elicits a mocking or facetious outlook.

As we discussed earlier, a non-content based response may be appropriate in instances when a respondent deems a speaker unwilling to converse on a particular subject, or if the respondent sees the speaker as close-minded. Not all cases of echoic utterances warrant this response. But, if a respondent sees a woman as mocking a male figure of authority, they may determine that this situation warrants a non-content-based response. What makes this reaction even more difficult to understand is that this non-content based response may or may not be consciously deployed. Though it may seem like an instinctual reaction to the respondent, what is really happening in this instance is a combination of gender norms dictating the way that the audience thinks women should sound like and a misguided reading of an echoic utterance.

As our sociolinguistic survey showed, women in fact deploy vocal fry in a variety of situations, many of which do not require echoic interpretation. The negative response interprets the assertions as echoic utterances, because they perceive women as ill-equipped to make the utterance that they are trying to make. Vocal fry is not a misfire of an intentional utterance—these women have the proper authority to make their utterance.
Many users of vocal fry are unaware that they are utilizing the vocal style when they speak and instead of accepting that a woman is attempting to make an assertion in a way that exudes authority, the audience is interpreting them as not having any authority whatsoever.

4 Conclusion

Female vocal fry may not be a novel linguistic phenomenon, but it has recently come under cultural scrutiny. Using the frequency code and echoic accounts of utterances, we have developed an framework for interpret this scrutiny. We argue that negative reactions to female fry obtain from tacit or explicit interpretations of the speaker as inappropriately echoing an authoritative-male voicing.

After giving examples of the negative reaction pattern in contemporary media, we gave an overview of vocal fry as a linguistic and sociolinguistic phenomenon in Section 1 and showed that female fry violates the frequency code, which associates lower pitches with masculinity and higher with femininity. In Section 2, we proposed the judgments women receive for vocal frying are non-content-based. In Section 3, we argued that the Sperber and Wilson’s pragmatic account of echoic utterances can help explain why women receive non-content based judgments for their vocal fry utterances. Women who speak in a vocal fry are seen as mocking authoritative males, which is unsettling to their respondents. Women are expected to sound a certain way, and when they don’t, they receive non-content based responses, which are often chastising.

Though our account may explain the position of the respondent, it is not an excuse for writing off negative reactions to female fry as a misunderstanding. The negative reaction to female fry is still rooted in normative expectations about how men and women should sound, and even though there is some biological foundation for the frequency code, it is outdated and sexist to hold that utterances violating gendered expectations about vocal
tone should automatically warrant non-content-based responses. This is to say that those who are using vocal fry when they speak aren’t usually aiming to reference this attitude of mocking, so non-content-based responses are often inappropriate. Women who experience them are subject to an injustice similar, though not identical, to discursive injustice: their voices, but not their ideas, are being heard.

References


