



A Study in Scarlett: Creaky Voice and Romantic Intention in Spike Jonze's *Her*

Joachim Jelle

ABSTRACT

Research indicates that women may use creaky voice to increase their own perceived romantic desirability. This auditory analysis of Scarlett Johansson's use of creaky voice related this specific use of non-modal phonation to her romantic intentions. This was done by looking at four scenes in Spike Jonze's *Her*, all of which suggested a different stage in the relationship between the two main characters (Johansson and Joaquin Phoenix). Creakiness in these scenes was assessed by counting the number of clauses in the four scenes as well as the number of creaks in the clauses. It was concluded that a positive correlation exists between the use of creaky voice and the attempt on part of the female speaker to sound desirable. The present study therefore supports the existing literature on the romantic signaling function of creaky voice.

Keywords: non-modal phonation, creaky voice, vocal fry, desirability, romantic effort, Scarlett Johansson, *Her*

Corresponding author: Joachim Jelle (joachimjelle1994@gmail.com)

Department of English, Aarhus University

Introduction

An important aspect of attraction is, of course, the visual appearance of the other person, but that is not the only relevant aspect. Many studies have engaged in the topic of vocal attractiveness and what makes a specific modal or non-modal phonation desirable (Liu and Xu 2011; Van Borsel, Janssens and De Bodt 2009; Pennock-Speck 2005). Indeed, the non-modal phonation types such as breathiness and creakiness are frequently touched upon in academia, but also in the mainstream media as the vocal fry has risen in popularity with e.g. the Kardashians, resulting in a clash between what is an attractive voice feature. This clash has caused numerous people to be very outspokenly opposed to the vocal fry deeming it annoying, and some studies suggest that it may be career threatening and unattractive (Anderson et al. 2014).

As mentioned, there are different phonation types, all of which can be placed into what Ladefoged calls a “continuum of phonation types, defined in terms of the aperture between the arytenoid cartilages, ranging from voiceless (furthest apart), through breathy voiced, modal voicing, and then on through creaky voice to glottal closure (closest together)” (Gordon and Ladefoged 2001, 1). This study will exclusively focus on the creaky phonation, which, according to Gordon and Ladefoged, “is typically associated with vocal folds that are tightly adducted but open enough along a portion of their length to allow for voicing” (2001, 4). Thus, it sounds like a “series of irregularly spaced vocal pulses that give the auditory impression of a ‘rapid series of taps, like a stick being run along a railing’” (Gordon and Ladefoged 2001, 4).

With their study of Scarlett Johansson from 2015, Crocker and Shaw investigated the connection between Johansson’s use of creaky voice and the nationality of her character i.e. British or American, and they found two interesting discoveries. Firstly, she uses creaky voice more frequently in her roles as a young American female than in her roles as a young British female. Secondly, they concluded that there was a connection between the type of character she played such as flirtatious, seductive and persuasive women and the level of creak in her voice, specifically deciding that it, indeed, was creakier (Crocker and Shaw 2015, 27).

The purpose of this study is to determine whether this application of a non-modal phonation enforces sexual attractiveness. I will do this by replicating Shaw and Crocker’s study of Scarlett Johansson, and therefore this study aims at extending their analysis to a more recent film, namely *Her* (2013) by Spike Jonze, by looking at the number of creaks Scarlett Johansson produces and determine a correlation between the phonation and the sexually romantic scenes, non-romantic friendly scenes and non-sexual romantic scenes.

Additionally, it is notable that the aforementioned types of scenes may differ slightly compared to the movies Crocker and Shaw discussed, as Scarlett Johansson has to rely merely on her voice to express her emotions, since her character is an Operating System (OS) in *Her* with an artificial

intelligence. In this movie, she falls in love with a man, Theodore, and engages in a relationship with him. Consequently, there is no visual imagery to suggest when she tries to appear sexy, flirtatious and attractive, which could indicate that her sexuality and attractiveness is cued auditorily, and perhaps even through her use of creaky voice.

It is relevant to look at the context in which these occurrences of the vocal fry appear to see if they are more frequent in sexual and romantic situations, in friendly romantic situations (excluding intercourse), in non-romantic friendly situations, or in sexual non-romantic situations. Therefore, I have picked scenes before and during the main characters' relationship exemplifying some of these different types of situations. Indeed, what is interesting here is that this study presents a distinguished perspective upon the category of attractiveness, as I look into seductiveness and flirtation in different types of relationships e.g. the romantic part and the sexual part, whereas other studies consider it as one big category. As life and the movie suggest, relationships between people are not just simple and straightforward things, and therefore relationships, be they friendly, romantic and/or sexual, should be considered in their context and not just as one huge category.

Background and Literature Review

As stated, Crocker and Shaw found that when Johansson portrayed young characters of American descent that she was more prone to use vocal fry, and more specifically when the character was supposed to be seductive or flirtatious, which fits the study of Yuasa (2010) and the study of Wolk et al. (2011; 2015, 26). Crocker and Shaw additionally found that the use of vocal fry also helped American women establish an identity that was indexed to their contemporariness, thus indicating creaky voice as an active stylistic feature, which is female dominated (2015, 26).

Interestingly, the voice quality of the vocal fry has been attached to both sexes. Shaw and Crocker (2015) suggest that, based on the work of Yuasa (2010), creaky voice was previously exclusive to men, which changed due to females adopting this voice quality as to associate themselves with the notion of power and masculinity (2015, 21). Additionally, they refer to the study of Pennock-Speck (2005), which shows that there is a "significance of an American nationality in the prevalence of creaky voice, concluding that creak "enhances [American women's] desirability" (Crocker and Shaw 2015, 21).

However, in their study of the attractiveness of the female voice, Liu and Xu found that creaky voice was always the least attractive out of the three voice qualities; normal, breathy and creaky, as perceived by a male speaker (Liu and Xu 2011, 1275). Furthermore, they suggested that it may be related to perceived body size of the female and they explored this through an evolutionary approach, which may be difficult to relate to my study, as the main character is only a voice from an Operating System (Liu and Xu 2011, 1277). There are, furthermore, additional studies that suggest creaky voice

to be unattractive and undesirable. The study of Anderson et al. shows that “young adult female voices exhibiting vocal fry are perceived as less competent, less educated, less trustworthy, less attractive, and less hireable (2014, Discussion)”. The listeners were divided into three age groups: 18–33 (N=265, 133 women and 132 men), 34–50: (N=268, 134 women and 134 men), and 51–65 (N=267 133 women and 134 men) (2014, Discussion). This study then suggests that both men and women dislike the use of vocal fry.

Nevertheless, Pennock-Speck’s study of three female actors (Gwyneth Paltrow, Reese Witherspoon, and Renee Zellweger) who, like Scarlett Johansson, have portrayed both American British characters suggested a correlation between the vocal fry, young American women and a wish to appear more desirable (2005, 413). This is in line with what Crocker and Shaw found, but clearly not in line with the studies of Liu and Xu (2011) and Anderson et al. (2014), exemplifying the clash between the correlation of creaky voice and desirability.

Method

To obtain the data I need, I am using an auditory analysis, thus excluding the process of acoustic analysis due to time restraints. Consequently, I will not be defining the level of creakiness through an acoustic spectrogram, but rather determine the level of creakiness by quantifying how many instances in a scene exhibit the use of vocal fry. By going with the assumption that there is a correlation between level of creakiness and sexual attraction, I picked one scene that displayed sexually romantic interaction, one that displayed non-sexual romantic interaction, and lastly two scenes that showed non-sexual, non-romantic friendly interaction. Notably, I contextualized these scenes with qualitative comments that explicitly states why the specific scene is one of the aforementioned types and what the effect of the mood or atmosphere appears to be like.

I decided to reduce the units to analyze from the larger sentences due to the large difference between amount of talking, and to some extent subject of discussion, in the sexually romantic and non-romantic scenes, which is why I chose clauses.

I defined a clause to be the subject and finite verb plus e.g. the direct and/or indirect object, prepositional phrase and proper name like the main characters’ names. Furthermore, to be clear, I decided to include interjections and questions as clauses, since they do not exclude the occurrence of creaks.

Having quantified the number of clauses that contain creakiness, I calculated the percentage of creaks out of the number of clauses for easier comparison between the different types of scenes.

Results

In the first scene, Samantha and Theodore have been romantically involved for some time even exploring the sexual possibilities of their unorthodox relationship, which is why I defined this scene as being sexually romantic. Here, they have decided that there will be a surrogate body for Samantha, but she will still provide the voice and directions, which results in very flirtatious and somewhat authoritative behavior from her.

I have chosen to cut the scene after approximately three minutes, since the mood changes drastically from being sexually romantic to apologetic and distinctly non-romantic. Consequently, Scarlett Johansson here has 35 short clauses during the three-minute span, whereas thirteen of these are imperatives commanding Theodore's actions. Additionally, there was 11 instances where Scarlett Johansson used creaky voice, which can be seen in figure 1, and some of these, I would argue, were more intense than others in the sense that they were easier to identify. Most these creaks appear before the shift from the largely flirtatious and seductive nature to the sexual nature of the scene.

The second scene I picked is partly the last part of the previous scene where the mood changes, as Ted gets confused during the romantic time with the surrogate. The mood becomes apologetic, sad and confusing causing a total removal of the sexually romantic atmosphere. Additionally, I picked the next scene in which they say good-bye to the surrogate, and then proceed to have a talk about their relationship, which is still largely affected by the previous somber mood and lack of romance, leaving exclusively a friendly scenario.

During these four minutes, Johansson produces 45 clauses, many of which are of interrogative nature due to the uncertainty of the situation. In this non-romantic scene, there are merely three instances where Johansson produces creaky voice. One of these clauses is addressed to the surrogate and the other two are addressed to Ted in the shape of a question and an apologetic statement.

Moving on to a non-sexually romantic scene, unlike the surrogate scene, Ted and Samantha share a Sunday on the beach the day after having had sexual intercourse for the first time. The mood of this scene is romantic and warm, as the Ted and Samantha resemble teenagers who just fell in love. The nature of their conversation may not be overtly sexual, yet they are talking about the human anatomy, which gives the impression of some hidden interest in this non-sexual romantic scene.

Despite the scene being shorter than the rest, as figure 1 below shows, Johansson creaks 10 times out of the 23 clauses. Interestingly, this number of creaks is higher than that of the sexually romantic scene. She excludes creaking in interjections and surprisingly in a question of explicitly sexual nature.

In the last scene I investigated, we begin with the first meeting of Ted and Samantha when Ted, literally, turns her on for the first time. Therefore, the atmosphere appears to be polite and quite inquisitive, since Ted is uncertain as to how this female operating system works. With a couple of

light, sarcastic comments, they are establishing the fundamental feature of friendship in their relationship, which gives the impression that this is not necessarily love a first hearing.

As the scene fades out, I ended up on 56 clauses in which Johansson used creaky voice in 11 of these, as shown in figure 1. A lot of the sentences where she does not creak, she is helping Ted organize his computer and in the beginning of their conversation, which gives the scene a certain formal atmosphere albeit with the occasional joke. Interestingly, the few times she creaks, it is when she is saying her own name, explaining Ted how she came to be, and when she lightly takes jab at him.

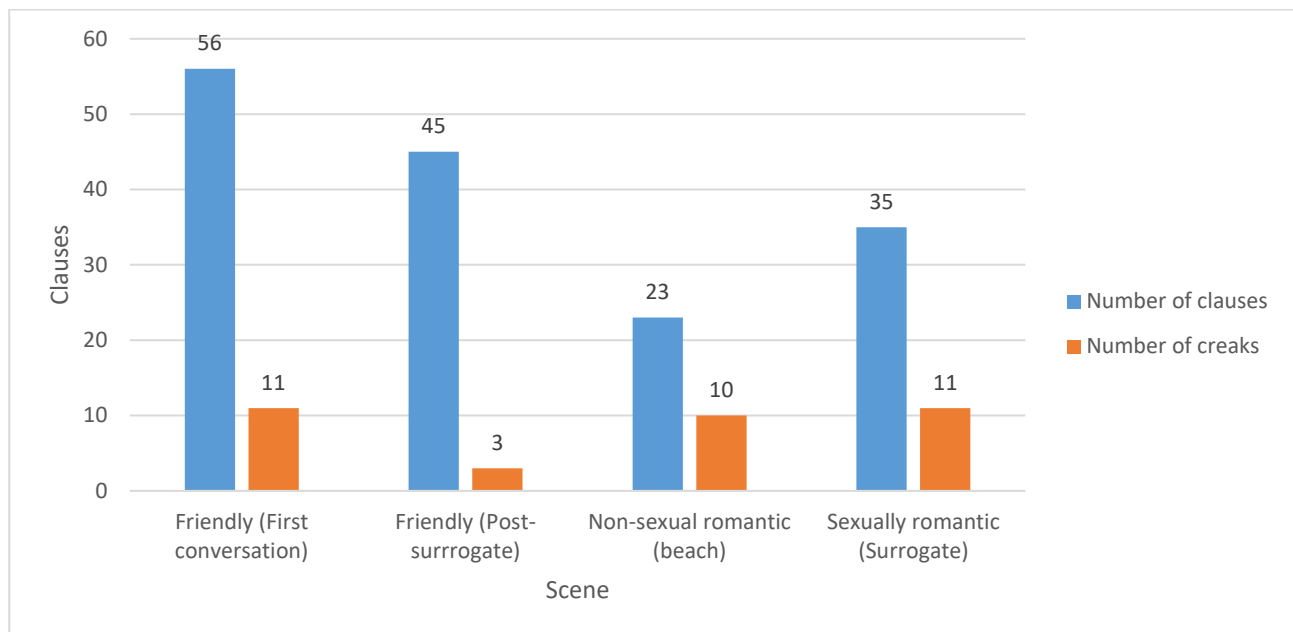


Figure 1: number of creaks (orange) to number of clauses (blue) per type of scene.

Discussion

With assumption that there is a correlation between desirability and the level of creakiness, the findings of the scene with the surrogate are to some extent as expected. As described, this scene is sexually romantic and quite explicit in its way of doing so, which would suggest that Johansson would exhibit a higher level of creakiness compared to the other scenes. Unsurprisingly, this scene displayed a high amount of creak per clause, which additionally was higher than the two least romantic scenes. Interestingly, the level of creakiness did not exceed that of the playfully romantic scene on the beach, as she in this sexually romantic scene provides 31,43 % creaks out of all her clauses, meaning that almost every third sentence contain a vocal fry.

With the findings put forth by Crocker and Shaw (2015) that Scarlett Johansson produced more creaky voice in the roles that were more seductive and flirtatious than in the roles that were not, the findings here fit very well with the aforementioned. Interestingly, it is almost all of the creaks that

appear before the shift from seductive to sexually explicit. This may be due to the character of Johansson switching from being consciously flirtatious to more affected by the moment and enjoying the time with her partner. With that being said, I am speculating merely on part of one scene. To get a more in-depth look, one would have to look at some of the other more explicitly sexual scenes, which, indeed, were in the movie, but were excluded due to difficulty in establishing the mood and atmosphere of the scene.

Yuasa (2010) suggested that the non-modal phonation of creakiness was previously perceived as a trait of masculinity and authority. By following this notion, one could argue that this fits well with the fact that Johansson's character speaks 32 clauses, whereas 13 of these are imperatives, directing the partner Ted. She has to consciously appear more authoritative and commanding, and that may to some degree cause a higher level of creakiness in her voice.

By choosing the scene immediately after the surrogate scene, we get an interesting contrast between these two scenes, since the relationship is at the same stage but the atmosphere and mood has changed. Following the hypothesis, the sexually romantic scene should provide a higher amount of creaks than this non-romantic scene, which indeed is also the case. There are merely three instances in this scene that Scarlett Johansson produces a vocal fry, and only two of these creaks are produced in conversation with Ted. In comparison, Johansson merely creaks in 6,5 % of all the clauses compared to e.g. the other exclusively friendly scene in which 19,6 % of the clauses contain a creak. Therefore, it seems reasonable to suggest that the context of a situation also plays a part in the way we choose to portray ourselves through speech when it comes to attraction. Indeed, Bell (1989) mentions in relation to audience design that "style shifts according to topic or setting derive their meaning and direction from the underlying association of topics or settings with typical audience members (1989, 247)". This may back up the claim that the change of setting affects the amount of creakiness produced.

Additionally, this scene also helps contrast the point above that creakiness may be related to a perception of authority. Consequently, this scene fits well under that notion with the atmosphere here being confusing and frustrating. They do not know the current state of the relationship, and thus they both appear as victims of this fight without any of them being very commanding and authoritative, which very well could affect the outcome of creakiness.

As shown in figure 1, there are fewer clauses in the non-sexual romantic scene on the beach compared to the other scene. It does, however, have the highest level of creakiness at 43,5%, which is even higher than that of the sexually romantic scene. Consequently, this result is to some extent surprising, since the scene with the highest level of creakiness was expected to be the sexually romantic scene. Nevertheless, this finding provides further evidence to the claim of Crocker and Shaw (2015) that the scenes in which Johansson's character had to appear flirtatious and seductive will contain a higher level of creakiness. Furthermore, it fits well with the assumption above that when

the scenes become more sexually explicit, the higher amount of creakiness may decay, as the setting of the scene is just romantic and flirtatious.

An aspect of the creaky voice I have neglected to mention is its placement in the clauses. Several papers have mentioned that creaky voice was more likely to be found in syllables at end of sentences than elsewhere (Redi and Shattuck-Hufnagel 2001, Wolk et al. 2011). Interestingly, in the non-sexually romantic scene, most the creaks Johansson produced are placed in the end of the clauses. Looking at the other results, this does indeed seem to be the most popular place for vocal fries to appear.

Following my hypothesis, I would have presumed the conversation in this exclusively friendly scene to contain a very low level of creaks like the post-surrogate scene, and the result was to some degree quite surprising. The reason for this presumption was that it is the couple's very first conversation, which arguably should not contain a high amount of flirtation, but rather an exchange of information. Nevertheless, the results show that 19,6 % of the 56 clauses contain, which is considerably higher than the post-surrogate scene with 6,5%. The reason for this increase is difficult to determine, yet it may, unlike the surrogate scene, be due to a more unconscious attempt at appearing desirable, since they, firstly, do not know each other, and secondly, she is supposed to be an Operating System with artificial intelligence with the purpose of organizing his life. However, the setting of this scene compared to the post-surrogate scene is very different, since the conversation in this scene is more upbeat and light with the occasional sarcastic comment. As a result, I would suggest the reason for this is due to the desirability of the friendly relationship and not necessarily that of the romantic relationship as it would be too early in the relationship to establish.

A point of discussion may very well be the methodology behind this study, as it is somewhat simple, even though it is to some extent a replication paper of a more time-consuming study that uses more complex analytical tools. Crocker and Shaw's use of acoustic analysis and spectrograms is ideally what would have been preferable, if I were not under time restraints, as it could have confirmed the phonation with certainty. However, I used an auditory analysis which could potentially result in my mishearing, yet this method allowed me to analyze more speech and scenes due to its being less time-consuming. Additionally, I could have used sentences instead of reducing it to clauses, which would have given me larger units to analyze. This approach, however, would not have considered the larger sentence difference there would be in the sexually romantic scenes with more interjections and the exclusively friendly scenes with longer, explanatory sentences. Ultimately, this does, nevertheless, still mean that there is a somewhat higher number of clauses in the scenes with more explanation such as the exclusively friendly scenes. I could have cut the scenes earlier, but that may have tampered with the contextualization and to some extent the atmosphere.

I must also note that it is specifically my qualitative description and contextualization of these scenes, and these interpretations may not be in line with what everyone interprets. However, I do

believe I provided adequate description for these scenes even though it was difficult to place some of the scenes into the categories of the different types of relationships, e.g. the post-surrogate scene, as they are still in the same place in their relationship, yet the mood and atmosphere have changed the setting and their feelings for each other.

Lastly, it is notable that the subject of the research is an Operating System, and is thus merely portrayed through her voice, which causes the conversations between Ted and Samantha to be through either speakers or headphones. Therefore, the latter may result in potential whispering or lowering of the voice, to not come off speaking too loudly, which in return could potentially affect the amount of creaky voice.

Conclusion

By extending the study of Shaw and Crocker to the film *Her*, I have sought out to investigate their finding of a correlation between level of creakiness and desirability or seductiveness. The present study has found evidence that supports the assumption that a higher level of creakiness is correlated with the speaker's intention of being flirtatious, seductive, desirable, and thus sexual attraction.

More specifically, in scenes that of a sexually romantic or just romantic nature, the number of creaks produced by Scarlett Johansson is significantly higher than in the scenes of friendly nature. However, a finding suggests that scenes of friendly nature with a sad and somber mood are those, which contain the lowest level of creaks, since there is such a large gap between the two friendly scenes.

Lastly, for further studies, it would be interesting to investigate the use of breathy voice, since there are some linguists who suggest this feature to be more desirable. One could replicate this study and exchange creakiness for breathiness. Additionally, it would be interesting to investigate other character's voiced by Scarlett Johansson to see if this finding is consistent, e.g. *The Jungle Book* and *Sing*, since these movies are aimed at kids and young adults, which could be a lack of creakiness, going by the present findings.

References

- Anderson, Rindy C., Casey A. Klofstad, William J. Mayew, and Mohan Venkatachalam. 2014. "Vocal Fry May Undermine The Success of Young Women in The Labor Market." *PLOS ONE* 9 (5): e97506. doi:10.1371/journal.pone.0097506.
- Bell, Allan. 1989. "Language Style as Audience Design." *Sociolinguistics, A Reader and Coursebook*, edited by Nikolas Coupland and Adam Jaworski, 240-50. New York: Palgrave.
- Gordon, Matthew, and Peter Ladefoged. 2001. "Phonation Types: A Cross-Linguistic Overview." *Journal of Phonetics* 29 (4): 383-406. doi:10.1006/jpho.2001.0147.
- Liu, X. and Xu, Y. (2011). "What Makes a Female Voice Attractive." The 17th International Congress of Phonetic Sciences, Hong Kong: 1274-1277.
- Pennock-Speck, Barry. 2005. "The Changing Voice of Women." In *Actas XXVIII Congreso Internacional AEDEAN*, ed. J. J. Calvo García de Leonardo, J. Tronch Pérez, M. del Saz Rubio, C. Manuel Cuenca, B. Pennock Speck, and M. J. Coperías Aguilar, 407-414. academia.edu/2471714/The_changing_voice_of_women.
- Redi, Laura, and Stefanie Shattuck-Hufnagel. 2001. "Variation in the Realization of Glottalization in Normal Speakers." *Journal of Phonetics* 29 (4): 407-429. doi:10.1006/jpho.2001.0145.
- Shaw, Francesca, and Victoria Crocker. 2015. "Creaky Voice as a Stylistic Feature of Young American Female Speech: An Intraspeaker Variation Study of Scarlett Johansson." *Lifespans and Styles* 1 (0): 21. doi:10.2218/ls.v1i0.2015.1180.
- Wolk, Lesley, Nassima B. Abdelli-Beruh, and Dianne Slavin. 2012. "Habitual Use of Vocal Fry in Young Adult Female Speakers." *Journal of Voice* 26 (3): 111-116. doi:10.1016/j.jvoice.2011.04.007.
- Yuasa, Ikuko P. 2010. "Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women?" *American Speech* 85: 315-337.