Creaky Voice as a Stylistic Feature of Young American Female Speech: An Intraspeaker Variation Study of Scarlett Johansson

Francesca Shaw and Victoria Crocker
chess.shaw@hotmail.co.uk
victoria.crocker.1989@googlemail.com
Creaky Voice as a Stylistic Feature of Young American Female Speech: An Intraspeaker Variation Study of Scarlett Johansson

Abstract
This study examines the stylistic use of ‘creaky voice’ in a single speaker: the American actress Scarlett Johansson. Recently, there has been a marked increase in both media and academic interest in creaky voice, with work by Yuasa (2010) and Wolk et al. (2011) confirming the prevalence of this feature among young American female speakers. Our study was directly motivated by the work of Barry Pennock-Speck (2005), who took a qualitative approach to analyzing the speech of three American actresses for stylistic modulation of their voice quality. The present study focuses on only one American actress (Johansson), who was chosen as she is an established, successful young American female (at time of research) and therefore was an appropriate subject to represent the social group we are discussing. Our materials included six of Johansson’s films that were developed whilst she was between the ages 18–24. This age range falls in line with previous work on creaky voice (Wolk et al. 2011) who defined their age bracket of study as 18–25 years old. We contrasted American and British character roles and noted the level of creak present through both quantitative and qualitative analysis of six films: three in which she played an American and three in which she took on an English (UK) accent. Acoustic data evaluation involved coding for creak on syllabic nuclei and carrying out a statistical analysis to determine significant influences on the pattern we observed. Our qualitative analysis covers the following variables: character traits and personality, time period in which the film is set, and the age of Johansson’s character. Results showed that there was significantly more creak in Johansson’s speech while she was performing in an American role, in line with the study previously conducted by Pennock-Speck. Our qualitative findings suggest that creak is modulated at an additional level, indexing seductiveness and intimacy with the interlocutor.
Creaky Voice as a Stylistic Feature of Young American Female Speech: An Intraspeaker Variation Study of Scarlett Johansson

Francesca Shaw and Victoria Crocker

1 Introduction

Towards the end of the 20th century, creaky voice was considered a marker of male speech and it was posited that women were beginning to adopt this voice quality in order to evoke the authoritative connotations of masculinity (Yuasa 2010). Barry Pennock–Speck’s 2005 work has shown the significance of an American nationality in the prevalence of creaky voice, concluding that creak “enhances [American women’s] desirability” (Pennock-Speck 2005:413).

Physiologically, creaky voice is the positioning of the vocal folds such that “the arytenoid cartilages are tightly together, so that the vocal folds can vibrate only at the anterior end” (Ladedefoged and Johnson 2010:150). The resulting slow vibrations create a low ‘creaking’ sound with the apparent strain on the voice resulting in this quality being referred to as glottal or vocal ‘fry’.

Creaky voice (also known as vocal fry) is phonologically contrastive within several languages, including Hausa which has creaky-voiced /b d j/ contrasting with their modal counterparts (Laver 1994:195–197). Sounds with a glottal place of articulation can cause the surrounding phonetic context to be produced with creak, and some speakers simply have more creak in their voices inherently. It may also be present in speech as a result of environmental and contextual factors such as tiredness and age, or as the result of a speech disorder (Wolk et al. 2011). Pragmatic use of creaky voice has previously been observed within the Tzeltal Maya speech community in South America: Brown and Levinson (1978) and Sicoli (2010) describe the use of creak in this group when seeking commiseration, and reciprocally to show sympathy.

More recently, there has been an observed trend in the prevalence of creaky voice among a particular social group: young American women. Media coverage of ‘The Creaky Voice Craze’ (Quail 2009) has been extensive, particularly with respect to celebrities whose speech features this trait. This trend provides the motivation for our study, as we conducted an intraspeaker variation study of the patterns of creaky voice in the speech of the American actress Scarlett Johansson to investigate the hypothesis that creak indexes membership of the social group ‘contemporary young American female’. For the purposes of this paper, we take contemporary to mean young, educated, urban-oriented, and upwardly mobile, informed by Yuasa (2010:316).

In his 2005 paper, Barry Pennock–Speck investigated the way in which creak was seemingly being used stylistically in young American female speech. His study looked at the levels of creaky voice in three American actresses — Gwyneth Paltrow, Reese Witherspoon, and Renee Zellweger — when they took on British and American character roles. Pennock–Speck noted that creak appeared more frequently in the American character roles, and concluded that “desirability depends on the cultural setting — creak is desirable in America but less so in Britain” (Pennock–Speck 2005:413). His study aimed to show the importance of creak in constructing and maintaining a desirable identity among young American females. He did this by examining the manipulation of creak with respect to the accent, and thus the identity of, the characters that the actresses embodied. Pennock–Speck proposed that the presence of creaky voice in the actresses’ speech may go some way to explaining why creak is a desirable trait in the population more generally, given their status as role models (Pennock–Speck 2005:412). His hypothesis was that “creak is an important component of contemporary female speech in the United States and not in Britain and this should be borne out if creak is not found in their portrayal of British women” (Pennock–Speck 2005:411).

Our study is structured in a similar way to Pennock–Speck (2005). We selected a different American actress than the ones in his study, Scarlett Johansson, and employed both a quantitative and qualitative analysis of speech data drawn from three American character roles and three British (RP accent) character roles. Creak is usually noticeable auditorily, which is what allowed Pennock–Speck (2005) to conduct his analysis without spectrographic analysis. For our analysis we chose to make use of the acoustic information available; creak is visible on a spectrogram due to its characteristically low F0 and irregular glottal pulses, compared with modal voicing. In line with Pennock–Speck’s (2005) conclusions, our study finds that our speaker has higher levels of creak in her speech when playing American character roles compared with British roles, and posit that creak is being used to index contemporariness in the context of being a young American female. The geographical aspect of this contextual social meaning explains our finding that Johansson’s constructed British English speech does not feature creak to the same extent.

Scarlett Johansson is a famous Hollywood actress, having starred in blockbusters and art-house films alike. Johansson was born in New York City, but has a General American (GA) accent. We considered Johansson to be an ideal candidate for our investigation of whether creak indexes ‘contemporariness’, which we take to mean educated, urban-oriented, and upwardly mobile. Additionally, our selection of Scarlett Johansson was motivated...
by her body of work in both American and British roles between the ages of 18-24, providing us with sufficient data for both American and British accented speech.

Table 1 outlines each of the six films from which we drew our speech data. The first three are films where Johansson appeared as a British character with an RP accent. The second three are her American character roles with a GA accent. None of the characters in these films used a non-standard regional or class-based accent.

<table>
<thead>
<tr>
<th>FILM (abbreviation)</th>
<th>TIME PERIOD in which the film was set</th>
<th>CHARACTER NATIONALITY; ACCENT</th>
<th>ESTIMATED CHARACTER AGE</th>
<th>S.J.’S AGE at time of filming</th>
<th>RELEASE DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE PRESTIGE (PRESTIGE)</td>
<td>19th Century</td>
<td>British; RP</td>
<td>18–25</td>
<td>21</td>
<td>2006</td>
</tr>
<tr>
<td>THE OTHER BOLEYN GIRL (OQB)</td>
<td>16th Century</td>
<td>British; RP</td>
<td>20</td>
<td>23</td>
<td>2008</td>
</tr>
<tr>
<td>GIRL WITH A PEARL EARRING (GWAE)</td>
<td>1665</td>
<td>British; RP</td>
<td>18</td>
<td>19</td>
<td>2003</td>
</tr>
<tr>
<td>HE’S JUST NOT THAT INTO YOU (HJNTIY)</td>
<td>21st Century (present day)</td>
<td>American; GA</td>
<td>mid–20s</td>
<td>24</td>
<td>2009</td>
</tr>
<tr>
<td>VICKY CRISTINA BARCELONA (VCB)</td>
<td>21st Century (present day)</td>
<td>American; GA</td>
<td>mid–20s</td>
<td>23</td>
<td>2008</td>
</tr>
<tr>
<td>LOST IN TRANSLATION (LIT)</td>
<td>21st Century (present day)</td>
<td>American; GA</td>
<td>20–25</td>
<td>18</td>
<td>2003</td>
</tr>
</tbody>
</table>

Table 1: Speech data sources.

Although it was possible to confirm Scarlett Johansson’s age at time of filming, the age of the character was not disclosed in any of our chosen films. We have been able to estimate the approximate age of the characters in three of the six films, using film context and time references — for example, in *Lost in Translation* Johansson’s character is a recent college graduate, and *The Other Boleyn Girl* is based on a real woman: Mary Boleyn. It is clear to see in all films that Johansson is playing a character that is close to her actual age during filming, which we consider to be a satisfactory control for the effects of perceived age.

2 Background and Literature Review

Yuasa (2010) investigated the effects of gender and nationality on the frequency with which creak appears in the speech of young individuals. The subject group focuses on ‘young’ speakers, comprising college students aged from 18 to 25. Their findings confirm the observation that rising use of creak is a female-dominated trend. Two-thirds of female American college students (female) produced creaky voice, a much higher result than that of their male or Japanese counterparts.

Wolk et al. (2011) also studied creaky voice patterns in ‘young adults’, analysing around 34 subjects, again with an age bracket of 18 to 25. In line with Yuasa (2010), they found that two-thirds of American female speakers displayed habitual usage of vocal fry (creaky voice), noting that creak was significantly more likely to appear at the end of sentences, which informed our decision to include ‘utterance position’ as a variable in this work.

Popular media have also picked up on the trend for young American women to display creak, with a focus on current celebrities (at the time of research) who are noted for the trait. Frequently cited examples include socialite Kim Kardashian, pop singers Ke$ha and Britney Spears, actress Zooey Deschanel, and the subject of our study, Scarlett Johansson. The articles often take a negative standpoint, asserting that creaky voice is unpleasant to listen to. An extreme example of this is the blog post *Your Vocal Fry is Making Me Hate You*, in which the author echoes the sentiment found in many other similar posts: “I have never met a woman who speaks with vocal fry who isn’t actually a fucking annoying person” (Borden 2013). This seems to sit at odds with Pennock-Speck’s conclusion that, in the correct context, creak is a desirable trait. We posit that there is a disparity between perception of those with high levels of creak in their speech and the identity that the ‘creakers’ themselves aim to construct.
Our qualitative analysis looks at what meaning might be indexed by creak further to contemporariness, and examines why these meanings might make creak a desirable feature, despite negative associations made by some listeners.

3 Method

Our study consisted of two investigations into the patterns of creak demonstrated in Scarlett Johansson’s speech. The first was a quantitative analysis of the number of tokens of creak present in our data, looking at the effect of the following variables: duration, utterance position, syllable nucleus, film, and character nationality. The second was a qualitative analysis of our data, considering what creak may be indexing socially, through examination of contextual effects, such as character traits and her attitude towards her interlocutors.

Ideally, we would have analysed continuous speech however there was not an adequate length of continuous speech available in any of the six films. In most cases we were able to control for context and interlocutor(s) by taking the entirety of each speech sample from one scene; however, in Girl With a Pearl Earring there was simply too little dialogue in any one scene to gather a sufficient amount of speech data from a single scene. Our solution was to compile several scenes, but keep Johansson’s interlocutor (Colin Firth) consistent. In our analysis of all the films we were careful to control interlocutor in this way to avoid confounding our results with interlocutor as an additional variable. Thus our data analysis is based on composite sound files created by extracting from each film those moments that consisted solely of Johansson’s speech. The range of these speech samples was 19secs to 44secs with a mean of 32secs. Speech from her interlocutors was removed from each clip.

Creak can be present anywhere there is voicing, but in order to focus our study we limited the envelope of variation to syllable nuclei: this is typically the vowel in the middle of a syllable, although it may also include syllabic consonants [l], [r], [m], [n] or [ŋ]. We began by orthographically transcribing all of the dialogue of every participating character in every film extract, as a starting point for more fine-grained acoustic analysis and annotation. A thorough analysis of the dialogue with all participants considered would inform our qualitative analysis of interlocutor effects.

We then segmented and phonetically transcribed each syllable nucleus in Johansson’s speech, in order to test the potential effects of duration and segment type. We used the acoustic analysis software Praat (Boersma and Weenink 2011) to produce spectrograms of our speech data, which we then used to inform our boundary placement and code our tokens as either creaky or modal. Our strategy for coding used auditory and acoustic information: when creak is present it is often clearly audible, but where it was not possible to make a confident judgement, we examined the spectrogram and waveform for presence of creak by looking for low fundamental frequency and irregular striations and glottal pulses. Where creak was present in Johansson’s speech, the nucleus was labeled as “1”; all modal nuclei were labeled “0” (see spectrogram Figure 1 below).

Three members of the research team worked on one speech sample initially and calculated our inter-rater agreement at 78%. Given this high margin for inconsistencies in coding, all three team members coded all syllable nuclei within the data set.

Figure 1 (below) exemplifies the transition from creaky voice to modal voice in the speech of Scarlett Johansson. Tier 4 on Figure 1 shows our phonemic transcription, tier 5 shows the coding for presence of creak (1=present, 0=absent), and tier 6 highlights the nucleus position (N = non-final). The vertical blue lines through

![Figure 1: Creak (al) versus modal voice (a), from Scarlett Johansson in Vicky Cristina Barcelona.](image-url)
these transcription tiers show our phoneme boundary placement, informed by the patterns on the spectrogram. The horizontal blue plotted line on the spectrogram shows that the striations in the glottal pulses become regular after an initial period of irregularity — a visual indication of the transition from non-modal (creaky) to modal voicing.

Knowing that creak is associated with utterance-finality (Henton and Bladon 1988 in Riebold 2010:44), we also coded for utterance-final and non-final portions of speech. Again, decisions about what constituted an utterance were made based on mutual agreement, relying on our perception of prosodic breaks. For each utterance the last three nuclei would be deemed "final" (marked as ‘F’) and anything else marked as ‘non-final’ (‘N’). In utterances containing three or fewer syllable nuclei we marked all the tokens as ‘non-final’.

Because our category of nucleus phoneme contained many singleton tokens, we grouped them together into subcategories by position — high-front, high-back, low-front, low-back, and mid-back (schwa). Diphthongs were grouped separately, as were non-vocalic nuclei (Johnson 2009).

4 Results

We used the statistics program Rbrul (Johnson 2009) to determine the statistical significance of each independent variable as a predictor of creak. A step-up/step-down multiple regression analysis indicated what the best model for predicting creak in these data would be, and the results of this can be seen in Table 2 (below).

Film, as a factor, is a subset of nationality, and was therefore not considered separately in our statistical analysis, but we do consider the effects of the films themselves in our qualitative analysis.

<table>
<thead>
<tr>
<th>Input Probability</th>
<th>0.33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total N</td>
<td>582</td>
</tr>
<tr>
<td>Deviance</td>
<td>712.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nationality</th>
<th>F.W.</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>0.56</td>
<td>p&lt;0.00817</td>
<td>314</td>
</tr>
<tr>
<td>British</td>
<td>0.44</td>
<td></td>
<td>208</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Utterance Position</th>
<th>P&lt;0.00894</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-final</td>
<td>0.44</td>
</tr>
<tr>
<td>Final</td>
<td>0.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nucleus phoneme</th>
<th>Not significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low front</td>
<td>[0.583]</td>
</tr>
<tr>
<td>Low back</td>
<td>[0.523]</td>
</tr>
<tr>
<td>High front</td>
<td>[0.48]</td>
</tr>
<tr>
<td>High back</td>
<td>[0.452]</td>
</tr>
<tr>
<td>Mid-back</td>
<td>[0.397]</td>
</tr>
<tr>
<td>Diphthongs</td>
<td>[0.533]</td>
</tr>
<tr>
<td>Other</td>
<td>[0.532]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Segment Duration</th>
<th>Not significant</th>
</tr>
</thead>
</table>

Table 2: Step-up/step-down multiple regression results to show significance of variables.

The results in Table 2 show that the best predictor of creak in our data is the nationality of the character, in line with our predictions and those of Pennock-Speck (2005). Utterance position is also a good predictor of creak for these data, with utterance-final position being 11% more likely to feature creak than utterance non-final according to our results. This is again in line with what we expected to find given that utterance-final position has already been shown to be a good predictor of creak (Wolk et al. 2011). Due to the low fundamental frequency at the ends of utterances, the presence of creak in this position is likely to be a result of phonetic factors rather than being manipulated stylistically. However, our study did not measure the intensity of the creak, which may have indexed some stylistic information and thus would make a good candidate for further investigation. The phonemic identity of the nucleus was found to be statistically insignificant, as was duration of the segment.

A comparison of the number of tokens of creak present when grouped by character nationality finds that there was a near-equal proportion of modal tokens collected for both the British and American speech data groupings, in contrast with the number of creak tokens in our American speech data (114 tokens), which was markedly higher.
than found in the British speech data (70 tokens). This supports our central hypothesis that Scarlett Johansson uses more creak in her American roles than in her English-accent roles, suggesting that creak is a stylistic feature — lending support to the wider hypothesis that creak is used stylistically by the general young American female population.

Finally, Figures 2.1 and 2.2 (below) show the variation of creak by film. While we were unable to test this effect statistically for reasons of colinearity, a descriptive analysis of differences between films points to some suggestive influences on the realization of creak. In Figure 2.1 we can see that the American film that contained the highest level of creak was *Lost in Translation* (39.5%); however, there is a relatively small difference between this and the lowest level of creak in the American film *He’s Just Not That Into You* (34.2%).

In comparison we can see that there are fewer instances of creak within the British films as shown in Figure 2.2. *Girl with a Pearl Earring* and *The Other Boleyn Girl* have low creak presence with 13.6% and 14.3% of coded tokens, respectively. However, *The Prestige*, a film in which Johansson assumes a British accent, has the second-highest percentage (39.4%) of creak out of all films considered (American and British), contrary to the predictions of our main hypothesis and the results for nationality. With this in mind we look at *The Prestige* and film as a variable in our qualitative analysis and consider that factors other than nationality can affect the stylistic use of creak.

![Figure 2.1: ‘Creak’ to ‘non-creak’ tokens by film (American).](image)

![Figure 2.2: ‘Creak’ to ‘non-creak’ tokens by film (British).](image)

## 5 Discussion

Our hypothesis that more creak would be present in speech data taken from American character roles compared with the levels found in British roles is supported: nationality was the strongest predictor of creaky voice. This is in line with the previous findings, including those of Pennock-Speck (2005), Yuasa (2010), and Wolk et al. (2011). The argument put forth in these studies is that creak might be a desirable feature of young American female speech,
and in addition, that it indexes a valuable contemporary persona. In this section, we consider what other social meaning creak may carry.

Our quantitative analysis returned a seemingly anomalous result for The Prestige (where levels of creak were found to be relatively high for a British character role). To look further into this finding, we returned to the scene in The Prestige from which we took our speech sample and looked at its contextual setting in terms of interlocutor and character type. In this scene, Johansson’s character is trying to be persuasive, acting seductively towards her male interlocutor in order to try to get what she wants. In fact, when looking at all the scenes across all six movies, we found creaky voice to be more prevalent in scenes of intimacy or seduction, or in roles where sexiness is a central facet of the character. All of the American roles we sampled included deliberate provocativeness and flirtation as key components of Johansson’s character. Creak may, therefore, be a social marker of overt sexuality and seduction as well as indexing contemporariness and femininity. It would be advantageous to study this further with a wider sample of films that allowed for a comparison between nationality and character traits. In the present study, excluding The Prestige, the British characters were meek and timid, and were pursued by male characters, whereas all three American characters were pursuing the male characters.

Our analysis is consistent with the ‘Speaker Design’ theory (Schilling-Estes 2008a), under which an individual modulates aspects of their speech in order to construct a socially meaningful persona. It is plausible that Scarlett Johansson is constructing identities at several levels — primarily, and least autonomously, at the level of ‘contemporary American female’, and secondarily at the level of the characters she portrays and the extent to which they are overtly sexual and flirtatious. The fact that she is an actress and we have collected speech data from her in character roles implicitly means that the speech data obtained will be constructed rather than spontaneous and will involve the creation of a persona. This fits with Natalie Shilling-Estes’s ‘Speaker Design’ theory: “[t]he desire to project a particular type of persona, maintain or bring about a particular type of relationship between oneself and one's interlocutors, and/or position oneself with respect to wider social groups or societal values or norms” (Schilling-Estes 2008b:975).

A possible contextual confound is that all three of the British-accent films were historical (to varying degrees, see Table 1). Within this study, this means that if Johansson is using creaky voice as a newer, emergent stylistic feature (indexing being contemporary), we cannot actually separate this from her use of creaky voice as an American feature (indexing national identity). It would be informative to do further research looking at American historical films and comparing the actress’s use of creak to those set in the modern day.

It would also be useful to look at more fine-grained divisions of the relatively broad label ‘contemporary’. For example, a division between educated, urban-oriented older women compared with younger women with similar traits may illuminate the importance of youth in the prevalence of creaky voice.

Furthermore, the ‘husky’ qualities of Johansson’s voice are often noted in the popular press. The Harvard Crimson said the following about her in a 2006 article: “It makes you wonder if, as a youth, Scarlett’s parents took a big strip of sandpaper and ground that voice from a perfect diamond down into a rough-cut masterpiece” (Chung 2006). Though ‘huskiness’ and creaky voice are two distinct features, we should consider the possibility of speaker-specific influences on our results. Johansson may be manipulating and exaggerating traits beyond social indexicality in recognition of her husky voice as an iconic attribute. Further study of Johansson’s speech outside of her character roles would improve understanding Johansson’s ‘base line’ level of creak.

This would also allow us to investigate any potential exaggeration or caricaturing effects when playing a role. As Auer (2007:32) notes, in creating a character “the portrayed person is rendered in a streamlined and exaggerated way […] the stylistic features used in stylizations are usually only a subset of the features that make up a social style: those that are salient and easily recognized by outsiders”.

Indeed, the salience of creak as a marker of contemporariness may explain why it not only persists, but is on the rise among young American women, despite a widespread distaste for this phonetic feature as indicated by media coverage of the ‘creaky voice phenomenon’. It is an easily reproducible and thus efficient way for a young woman to show that she is a part of a desirable social group, the attractiveness of which outweighs negative feedback from those outside of this social group.

6 Conclusion

This investigation has found quantitative evidence to support the hypothesis that creaky voice is used stylistically by young American females in order to construct certain aspects of their identity, with creak found to index contemporariness in particular. This is in accordance with ‘Speaker Design’ theory (Schilling-Estes 2008a), giving evidence that a single speaker can manipulate aspects of their speech in order to construct and maintain an identity.

The results of our quantitative analysis of intraspeaker variation in Scarlett Johansson’s use of creaky voice support our main hypothesis: that there would be proportionally more creak in her American roles than there would in roles that required her to affect an English accent. Utterance position is also a significant predictor of creak; our study took a phonetic interpretation of this, as opposed to stylistic.

We have argued that the variability of creak is a result of the construction of identity at multiple levels.
Johansson is using creak to indicate membership of the social group ‘contemporary American female’, but in addition to this, creak levels are influenced by the construction of specific character roles. Johansson shows more creak in her voice when playing seductive and persuasive characters, suggesting that creaky voice is a salient marker of these traits for young American females.

We have suggested several directions for further work, one being an investigation in greater depth into the qualitative contextual influences we have seen in this study, particularly in the case of historical and period films. We also suggest a comparison between speech taken from films and comparatively ‘natural’ speech in which Johansson is conveying only her own identity.

References


chess.shaw@hotmail.co.uk
victoria.crocker.1989@googlemail.com