2023. *Proc Ling Soc Amer* 8(1). 5548. https://doi.org/10.3765/plsa.v8i1.5548.

"He sounds like a bureaucrat": Variation of voice quality in stylistic performances

Robert Xu*

Abstract. Voice quality is a vital but understudied sociolinguistic feature in constructing style and personae. This study examines stylistic performances of character types in Beijing Mandarin, to understand how voice quality varies and comes to conventionally associated with social meanings. Acoustic analysis of recorded performances shows that voice quality is a dynamic variable that should be examined from the lens of different linguistic and discursive levels. Combined with meta-discursive analysis, the findings show that voice quality not only helps to typify a persona, but also contributes to forge relationships in frames of social interactions.

Keywords. voice quality; sociolinguistic style; persona; character type; prosodic variation; Beijing Mandarin

1. Introduction. Voice quality can be broadly defined as the characteristic auditory coloring of a speaker's speech. It results from the configuration of both the larynx, including but not exclusive to the mode of the vocal folds, and the upper vocal tract, from the pharynx to the oral and nasal cavity (Laver 1980). While some regard voice quality essential and therefore "semipermanent" (Abercrombie 1967) to a speaker's identity because of its intimate relationship with the body, it is highly malleable and manipulatable, changing from moment to moment, presenting a speaker's social membership and standing. It is also an iconic representation of a speaker's affective stance, signaling the changes of their physical and emotional states (Podesva and Callier 2015). Podesva (2007) explores the use of falsetto in the construction of a diva persona in certain social interactions. Starr (2015) showed that a sweet voice plays a crucial role in constructing certain type of femininity in Japan. Together, these studies have shown that voice quality is a key stylistic element in the construction of personae.

In this study, I use character type as a vehicle to elicit stylistic performances that are rich in voice quality variation. Character types, or characterological figures, are abstractions of salient and performative social images of personhood (Agha 2005), like "Valley Girl" in the American context. They are enregistered with a set of linguistic and other semiotic features. They are an important and useful site to study style because they establish the most conventional association between variation forms and social meanings, through iterations of meta-discursive comments and discussions in the public discourse. They live in dynamic dialogic social contexts and evoke specific actions, therefore mediating social dynamics through the affects they often display. This paper is part of a larger project where I explore prosodic variability of character types in Beijing, where a social dynamics distinct from typical English-driven sociolinguistics studies meets a prosodically complicated language.

In this paper I will highlight the results of three character types in my study – Bureaucrat, Angry Woman, Childish Girl. They will also be compared against two other character types - Neighborhood grandma and policing grandma. By investigating the voice quality of these character types, I examine how voice quality varies in dynamic and structured ways, and how it takes on locally-salient social meanings.

* Thanks go to the Stanford Center at Peking University for funding and assisting this study, and to Penny Eckert, Rob Podesva, and Qing Zhang for comments and advice. Author: Robert Xu, Stanford University (robxu@stanford.edu).

2. Methods.

2.1. DATA COLLECTION. Based on an experimental pilot study (Xu 2019), where I surveyed 18 character types in Mainland China, and my fieldwork in Beijing, I chose 12 character types that are universally familiar and recognizable in Beijing. They are frequently labeled and widely commented on in local discourse regarding their conventional use of linguistic resources. In this study, speakers can easily perform them linguistically and bodily without a script and much preparation.

An interactive game was designed and participated by 62 Beijing Mandarin speakers. In this game of three[1], each speaker was given a set of cards with random 4 of the 12 character types. On the cards there were only the character type labels, and no scripts were provided. One speaker performed the given character types while the other two listened blindfolded. The listeners then guessed and labeled the performances by writing down three labels they associated the performance in order of possibility, which were graded with different weights later. The performances became the speech data I analyzed in this study, and the guessing results provided a rough perceptual evaluation of the performances for the initial processing of the data.

The groups of three speakers also participated in a guided focus group discussion, in which they commented on each other's performances, what they thought about these character types, and the broader social and linguistic landscape of Beijing. For this study, I use these discussions as meta-discursive data to understand how these character types become salient, how they are voiced linguistically, and how they are evaluated in the local context. All the performances and discussions were audio taped with body microphones and videotaped.

2.2. DATA ANALYSIS. Five character types stood out in both how prominent and relevant speakers placed them in their life, and how successfully these types were performed during the game. These types are:

> (1) Bureaucrat (*lingdao*): A senior and old-fashioned official that always gives a longwinded and tedious speech in routine meetings.
>
> (2) Angry Woman (*pofu*): An unreasonable woman who created a scene by yelling in the public over something trivial.
>
> (3) Childish Girl (*sajiao*): A young woman acting in a childish manner to win over favors from others, in particular the boyfriend.
>
> (4) The Grandmas: Two types of grandmas were elicited for the speakers. Neighborhood Grandma (*dama*) is a nosy, gossipy, self-righteous older woman next door. The Policing Grandma (*jiedao dama*) is one that is incorporated in local administration in charge of cleanliness and safety of the neighborhood. The two types do not distinguish themselves enough to be elaborated for their voice quality variation. In the follows I will mention them as one group, the Grandmas, in this paper.

The successful performances of these character types in the game, judging from the grading of the guessers' labeling, were transcribed and forced aligned. Manual corrections were made to adjust the forced aligned boundaries and to generate different annotation tiers for different

[1] One speaker did not show up for their scheduled group. As a result, this group consisted of two Beijing speakers and one Northern Mandarin speaker. During the game for this group, the non-Beijing speaker only participated in guessing the labels, while each Beijing speaker performed six labels.

prosodic boundaries. This paper reports the results of the quantitative analysis of performances for which I have completed this process. That includes 18 performances of Bureaucrats, 8 of Angry Woman, 11 of Childish Girl, 15 of Neighborhood Grandma, and 14 of Policing Grandma.

To quantify voice quality, the audio recordings of the correctly labeled performances were analyzed acoustically and auditorily. Corrected H1-H2 (written as H1*-H2* in the follows) and Smoothed Cepstral Peak Prominence (CPPS) were measured at the midpoint of every voiced segment. H1-H2 measures the degree of constriction of the vocal folds, the result of which potentially places voice on a spectrum of creakiness at the low end and breathiness at the high end. CPPS measures the regularity of vocal folds vibration, higher being more regular.

Following Laver's discussion on Frøkjaer-Jensen and Prytz's α-score as a measurement for vocal tension (Laver 1980, Frøkjaer-Jensen & Prytz 1976), I measured tension in the voice by dividing average intensity at higher frequency (1500-3000 Hz) by intensity at the lower frequency (0-1500 Hz). The idea is that tense voice has stronger upper harmonics than lax voice, meanwhile a relaxed pharynx could provide resonance for the fundamental and lower harmonics. Therefore, the higher the intensity ratio is, the tenser the voice might sound.

In addition, I measured Noise/Harmonic Ratio (NHR), and nasality with A1-P0. However, NHR for additive noise and vocal jitteriness did not provide interpretable results for larynx settings and auditory effects. A1-P0 did not generate significant results, It is also not a useful measurement for inter-speaker comparison. I will not report these results in this paper.

Finally, voice qualities that are auditorily significant but acoustically difficult to measure, like falsetto and nasality, were auditorily coded.

3. Results. In this section, I will first report the global voice setting in the performances of these character types, followed by how H1*-H2* vary at different domains, including the utterance level and throughout the performance, or the discursive level.

3.1. GLOBAL VOCAL SETTINGS. Figure 1 and 2 show the general pattern of H1*-H2* and CPPS for these five character types.

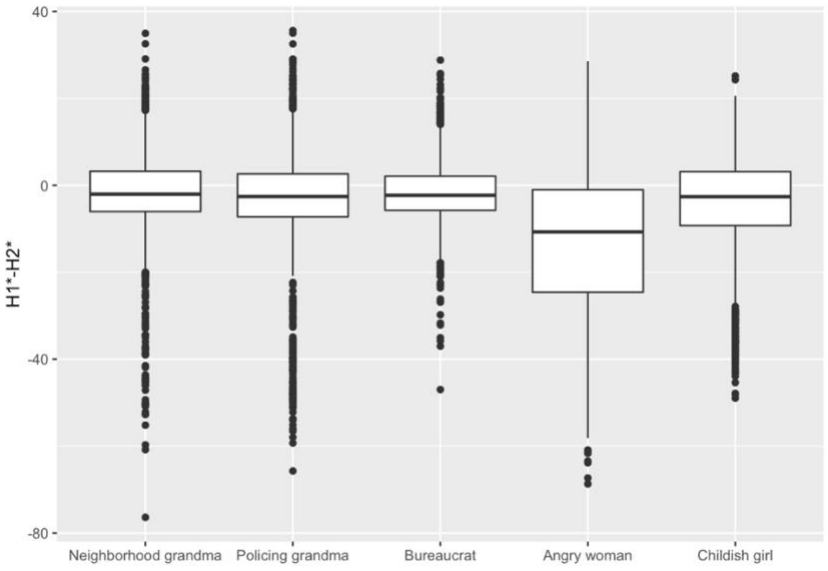

Figure 1. Mean H1*-H2* across character types

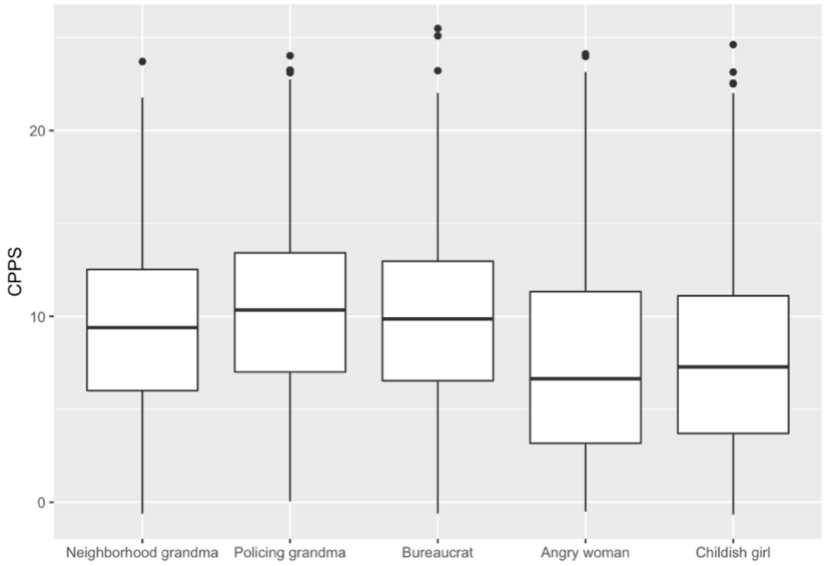

Figure 2. Mean CPPS across character types

In comparison to the other character types, Angry Woman has significantly lower H1*-H2* and CPPS. This indicates that speakers performed Angry Woman with a more constricted vocal setting, and less regular vocal folds vibration. Auditorily, Angry Woman's performances do not sound creaky despite the low H1*-H2*, but has a forceful and harsh quality. In another study addressing pitch variation of the same set of data, it was found that Angry Woman's voice was associated with significantly higher pitch than the other character types (Xu, forthcoming), which could in part explain why the high constriction did not result in creaky voice auditorily. The high pitch is in part realized by the use of falsetto in some performances. The forceful quality of Angry Woman is also evident in Figure 3, where Angry Woman is shown to have significantly higher voice intensity than all the other character types.

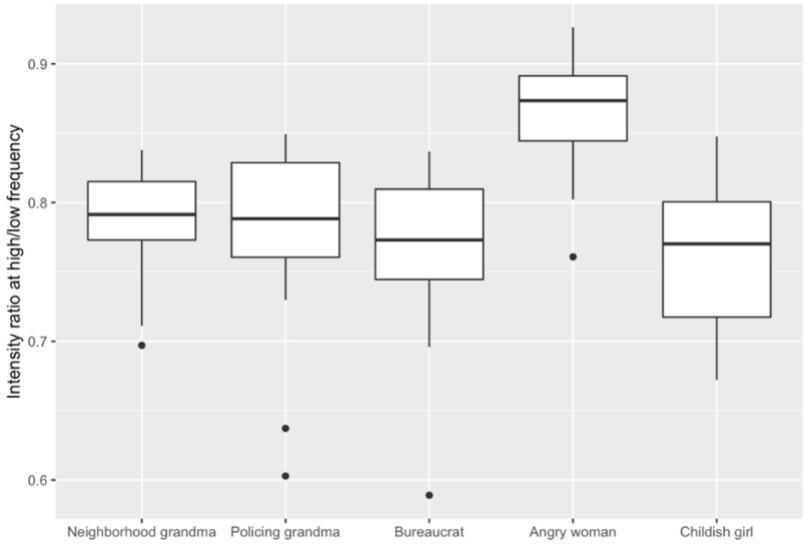

Figure 3. Mean vocal intensity across character types

Also standing out from the other types is Childish Girl, which uses a voice with less regular vocal folds vibration as shown by its significantly lower CPPS than Bureaucrat and the Grandmas, as seen in Figure 2. Moreover, the Childish Girl sound more nasal in performances by most speakers. Some speakers, especially some male speakers, also used falsetto to portray Childish Girl throughout their performances.

Figure 4 shows how H1*-H2* vary across speakers. Speakers used highly variable H1*-H2* when performing Angry Woman. In contrast, they used the least variable H1*-H2* when performing Bureaucrat, either because the conventional voice of Bureaucrat was closest to their modal setting and therefore easiest of perform, or because Bureaucrat's voice quality was so conventionalized that speakers converged to a similar voice setting during their performances.

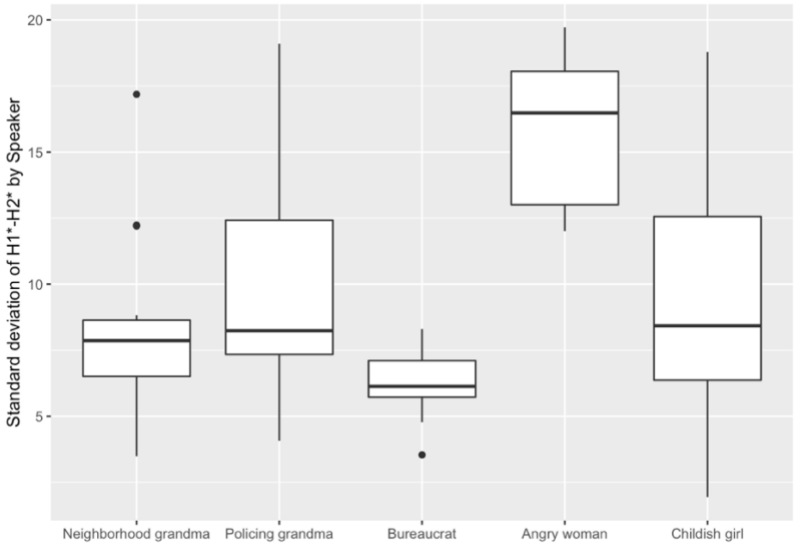

Figure 4. Cross-speaker voice variation in performing the character types

3.1 H1*-H2* VARIATION IN TIME. This section focuses on the variation of H1*-H2* in time at the utterance and discursive level. Figure 5 shows how H1*-H2* vary along the course of an utterance, and Figure 6 shows how H1*-H2* vary along the course of an entire performance.

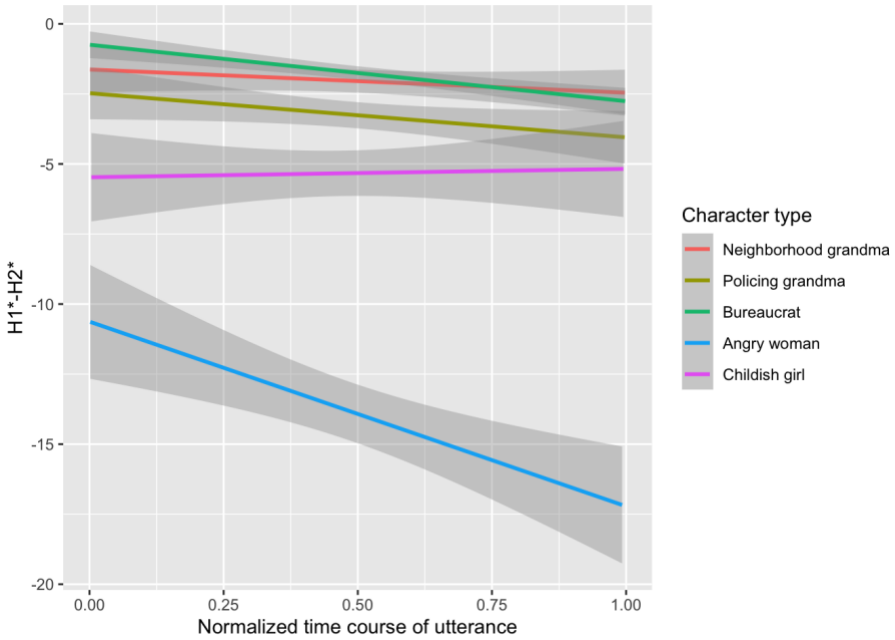

Figure 5. H1*-H2* variation along the course of an utterance

As shown in Figure 5, while generally H1*-H2* lowers towards the end of an utterance, the declination of H1*-H2* for Angry woman is extremely drastic, indicating that the vocal folds setting becomes even more constricted than the other character types at the end of an utterance. Childish Girl does not seem to have a declination of H1*-H2*, unlike the other types, suggesting a consistent voice quality throughout the course of an utterance.

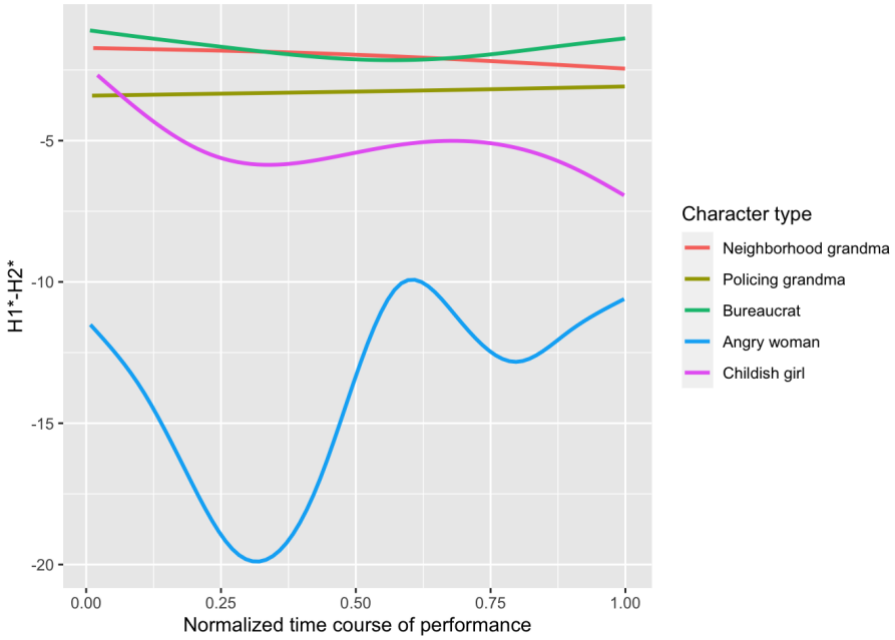

Figure 6. H1*-H2* variation along the course of a performance

As shown in Figure 6, H1*-H2* also varies to different degrees throughout the entire performance. A large degree of variation is found in *Angry Woman* and *Childish Girl*, despite

Childish Girl's lack of variation at the utterance level, suggesting that voice variation can operate at different linguistic levels.

4. Discussion. The results show that voice quality vary in various dimensions. First, speakers performed some character types, such as Angry Woman and Childish Girl, with distinct global voice settings. The attention-seeking, in-your-face Angry Woman, whose performances often center around conflicts in the public, has significantly lower H1*-H2* and CPPS, though sounding mostly pressed instead of creaky, with occasional falsetto. This voice has a forceful, harsh, sharp quality, with higher intensity at higher frequency. This tensity of the voice co-occurs with harsh swear words like *cao nimade* "fuck your mother" and a standing posture – many speakers chose to perform Angry Woman while standing and other types sitting down - with extensive pointing and body movement (Figure 7). The use of tense voice and these other semiotic resources together iconically resembles the perception of this character type in the meta-discursive comments made by participants. The action of Angry Woman was described as *majie* "yelling in the street" or *sapo dagun* "spreading the anger, rolling all over". The affect Angry Woman often displays was described as *duoduobiren* "pushing her agenda in a forceful way" and a form of "vocal fighting". The sharp yelling voice of Angry woman is doubly purposed: it is directed towards the person she has trouble with, with strong affective force, but it also summons attention to set up a stage to create a scene.

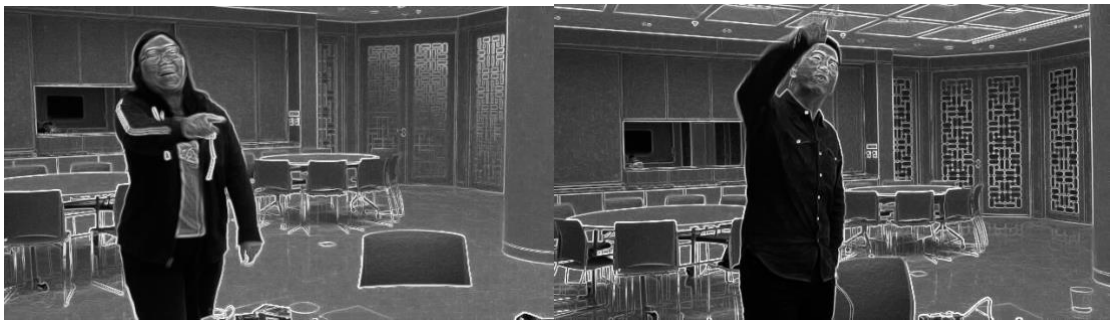

Figure 7. Stills of performances of Angry Woman

Childish girl, whose strategic flirting also aims at influencing the interlocutor, uses a significantly less regular vocal vibration as shown by its CPPS, in comparison to the Bureaucrat and the Grandmas. In addition, some speakers used falsetto to mimic the voice of a young woman, who during the interaction was imitating a child's voice. Many speakers also performed Childish Girl with a nasal voice. One speaker, as seen in Figure 8, even manually pinched his nose to have a nasal sounding voice throughout the entire performance. This is also reflected in the meta-discursive comments. The Childish Girl's voice was described as *ruanbulada* "soft sounding", therefore sounding "hyper feminine". It was also described as *diasheng diaqi*, or *naisheng naiqi* "nasally and childlike", as a way of "attention seeking purposely for certain demands from the boyfriend". The "cute" voice of Childish Girl, much like how a young woman in South Korean engaging in the *aegyo* (Moon 2017), enables the woman who does it to put on a docile and subordinate guise to enter a negotiation, where the childlike weakness turns into a persuasive power (Yueh 2012).

Figure 8. A still of a speaker performing Childish Girl

Interestingly, the nasality of this kind of voice is more ideologically constructed than articulatorily realistic. As demonstrated by this speaker who pinched the nose, as long as its quality has something to do with the nose, it does not matter if you are rejecting air flow in the nasal cavity.

The results also show that speakers seem to converge to a certain vocal setting for Bureaucrat, which has the least inter-speaker variability. This suggests that types like *Bureaucrat* are either more ideologically conventionalized in voice, or they are more modal and require less effort to perform. The lack of variability of voice at different linguistic levels and among individual speakers is reflected in the meta-discursive comments. Bureaucrat's speech was regarded *qianpianyilv* "boring", "monotone", and "always sounds the same". The content of the speech was full of *konghua* "all style no substance". The function of the speech is to show *shangxiaji guanxi* "the hierarchical power dynamics" and *fayanquan* "the power of speaking". This kind of invariable voice is coupled by low voice pitch (Xu forthcoming) and "many pauses and dragging of sentences". All the speakers performed the Bureaucrat with a rigid sitting position and limited body and hand movements. The bureaucrat exhibit authority by posing and speaking in a motionless set up. Requiring attention without using variable voice and body movement is a power move.

Figure 9. A still of a speaker performing Bureaucrat

Even though types like the grandmas do not show extreme tendencies in voice quality in my measurements, it does not mean they do not have idiosyncratic qualities. An auditory impression of the Grandmas shows that they are frequently performed with more nasality and sprinkles of falsetto. The acoustic correlates explored in this paper might not be able to cover some voice settings that the Grandmas typically sound like. More importantly, voice quality measurements should not be conceptualized as a spectrum of high and low values, like some other phonetic features like F0 and timing, where the extremes are more prominent and has larger potential in carrying social meaning. Voice quality of middling measurements might still be colorful enough to be recognizable, and further bricolage with other linguistic resources to construct a style.

The results also show that voice quality vary at different linguistic levels. Voice becomes more constricted generally towards the end of an utterance. But Angry Woman shows more variability at both utterance and discursive levels, while Childish Girl is more variable only at the discursive level. This indicates that variation of voice quality is multi-dimensional. Voice quality is not only variable depending on the general qualia of the character types, but it also varies to different degrees depending on what kind of action a person engages in. A natural next step is to investigate the correlation of the speech action and the varying voice quality, to explore how affect unfolds through voice quality in time.

5. Conclusion. In conclusion, these findings highlight the dynamic nature of voice quality. Variationists have benefited a lot from Laver's works and the Laryngeal Articulator Model (LAM) proposed by Esling et al. (2019) as this model bridges articulation and auditory quality. However, it tends to classify voice quality as types, instead of treating them as varying gradients (Garellek 2022). Examining voice quality as dynamic variables enables us to avoid labels that have become strongly associated with variationist studies focusing on English, and allows us to understand its variability at different linguistic and discursive levels.

Meta-discursive analysis further supports that character type, as a critical site to examine the rhematization of linguistic and social differences (Gal & Irvine 2019), is brought into being in part through the complex use of voice quality. By varying the voice at different levels and uniting multiple qualities, speakers can index different stylistic elements of personhood that is abstracted from dialogical interactions in everyday life. The degree of inter-speaker convergence indicates ease of performance and degree of enregisterment. Finally, Voice quality not only helps to typify a persona contextualized by the body, but also contributes to forge a specific relationship and power dynamics to the addressee in the particular social interaction that character type conventionally engages in, by indexing affects and positioning.

References

Abercrombie, David. 1967. *Elements of general phonetics* . Edinburgh: Edinburgh University Press.

Agha, Asif. 2005. Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15(1). 38–59. https://doi.org/10.1525/jlin.2005.15.1.38.

Esling, John H., Scott R. Moisik, Allison Benner & Lise Crevier-Buchman. 2019. *Voice quality: The laryngeal articulatory model*. Cambridge: Cambridge University Press.

Frokjaer-Jensen, Børge & Svend Prytz. 1976. Registration of voice quality, *Bruel and Kjaer Technical Review* 3. 3–17.

Gal, Susan & Judith T. Irvine. 2019. *Signs of difference: Language and ideology in social life.* Cambridge: Cambridge University Press.

Garellek, Marc. 2022. Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics* 94. 101155. https://doi.org/grbbtd.
Laver, John. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
Moon, Kyuwon. 2017. *Phrase final position as a site of social meaning: Phonetic variation among young Seoul women.* Stanford, CA: Stanford University dissertation.
Podesva, Robert. J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11(4). 478–504. https://doi.org/d4srtk.
Podesva, Robert. J. & Patrick Callier. 2015. Voice quality and identity. *Annual Review of Applied Linguistics* 35. 173–194. https://doi.org/10.1017/S0267190514000270.
Starr, Rebecca L. 2015. Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society* 44(1). 1–34. https://doi.org/10.1017/S0047404514000724.
Xu, Robert. 2019. Placing social types through prosodic variation: An investigation of spatial meanings in Mainland China. *Proceedings of the Linguistic Society of America (PLSA)* 4(42). 1–12. https://doi.org/10.3765/plsa.v4i1.4540.
Xu, Robert. forthcoming. Polyphonous and meaningful: Pitch variation in stylistic performances. *Penn Working Paper in Linguistics.*
Yueh, Hsin-I. 2012. *The tactic of the Weak: a critical analysis of feminine persuasion in Taiwan*. Iowa City, IA: The University of Iowa dissertation.