

Social Meaning of Voice Quality in Mandarin Chinese: Stance, Dialect, and Gender Variations

Qiuyuan LI¹, Peggy MOK²

^{1,2}*The Chinese University of Hong Kong*

Voice quality - characteristics such as breathiness, creakiness, and modal voice - arises from vocal folds configuration and is a rich source of linguistic and social meanings [1]. Cross-linguistic studies show that voice quality can encode lexical contrasts, mark utterance boundaries, or convey emotional states [2-4]. While voice quality does not convey linguistic meaning in Mandarin Chinese, it plays a fundamental role in the phonological systems in certain Southern Chinese languages. Breathy voice was involved in the tone split process in Wu Chinese and is preserved as a cue to tonal contrast [5]. These differences highlight the complex relationship between voice quality and the meanings it might take on, particularly in bilingual contexts where Mandarin interacts with local dialects. This study takes stance, the speaker's attitude toward what is being said or toward the interlocutor, as an entry point into this relationship. It asks whether speakers use voice quality to signal stance in spontaneous interaction, and whether such use varies by dialect background and gender.

The dataset consists of spontaneous conversational recordings in Mandarin from 40 speakers (20 Beijing Mandarin natives, 20 Wenzhou Wu natives), aged 18-35, balanced for gender. Utterances were annotated for voice quality and stance dimensions (Figure 1): Affect (attitude toward the topic, 1-5 with 5 being the highest), Investment (commitment/certainty, 1-5), and Alignment (alignment with the interlocutor, 1-5) [6].

Generalized Additive Models (GAMs) were fitted to predict the probability of modal phonation as a smooth function of each stance dimension, with dialect background and gender as fixed effects and random intercepts for speaker. Stance robustly predicts phonation across models (Figure 2). Affect shows a U-shaped pattern: strong evaluations are more likely to be produced with modal voice than neutral evaluations. Alignment and Investment show near-monotonic decrease and increase, respectively, indicating more modal phonation as speakers become less aligned with their interlocutor or more committed to their own utterance. Social factors also matter: Wenzhou speakers are 17.4% relatively more likely to use modal phonation than Beijing speakers, and male speakers are 20.5% relatively more likely than female speakers.

Acoustic measurements were normalized within speaker and reduced with Principal Component Analysis (PCA), yielding a set of acoustic dimensions that largely correspond to (i) breathiness/aperiodicity, (ii) energy/harmonicity, and (iii) source and spectral-peak configuration. Group-level GAMs on the resulting PC scores largely converge with the auditory analysis: stance systematically predicts movement in acoustic space consistent with "more modal" configurations at higher Investment and at extreme Affect. By contrast, Alignment has a less unified pattern across groups. The stance dimensions show dissociable effects across PCs, suggesting that speakers recruit multiple, partially independent phonatory-acoustic strategies to index stance (Table 1), instead of relying on a single voice-quality cue.

In sum, stronger Affect and higher Investment lead to more modal phonation, and Wenzhou speakers as well as male speakers are more modal than their counterparts. This study provides novel evidence that speakers actively manipulate voice quality to index stance, and that this practice is conditioned by dialectal background in a Mandarin-Wu bidialectal setting. Crucially, the dialect effect aligns with a typological asymmetry: voice quality is primarily paralinguistic in Mandarin but lexically contrastive in Wu, suggesting that speakers' phonological experience can shape how voice quality is deployed for sociopragmatic meaning even when they speak Mandarin. More broadly, the results show that social meanings expressed via voice quality are shaped by language-specific phonological systems as well as gender.

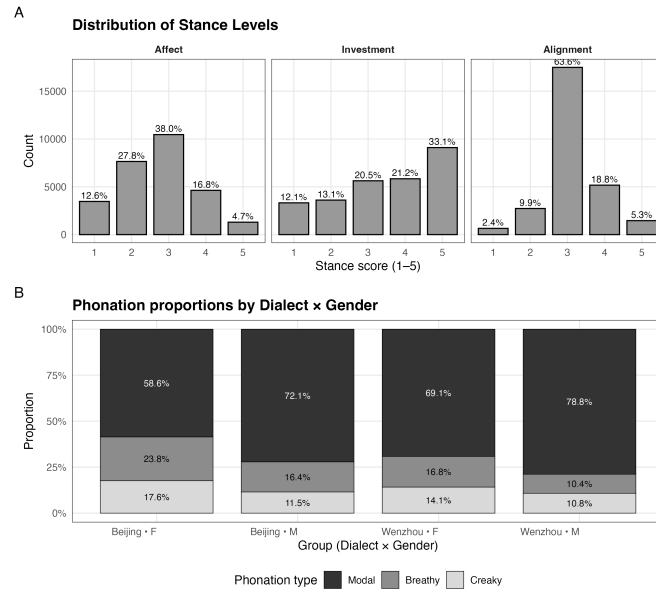


Figure 1. Distribution of stance levels and voice quality types.

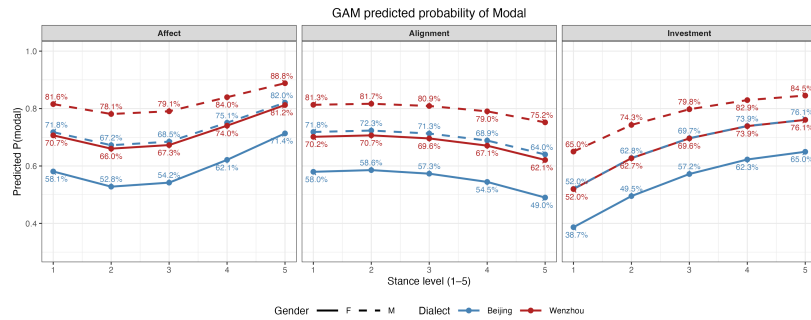


Figure 2. Predicted probability of modal voice across stance levels and dialect-gender groups.

Table 1. Group-level GAM smooths for stance effects on Principal Components 1-3. *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$; “N.S.” indicate non-significant smooths ($p \geq .05$).

Stance	PC	Beijing Female	Beijing Male	Wenzhou Female	Wenzhou Male
Affect	PC1	N.S.	N.S.	N.S.	N.S.
Affect	PC2	***	***	***	N.S.
Affect	PC3	N.S.	***	N.S.	N.S.
Investment	PC1	N.S.	*	*	N.S.
Investment	PC2	***	***	***	*
Investment	PC3	***	***	**	**
Alignment	PC1	N.S.	N.S.	N.S.	N.S.
Alignment	PC2	***	***	**	***
Alignment	PC3	***	N.S.	*	N.S.

References

- [1] Garellek, M. (2019). The phonetics of voice 1. In *The Routledge handbook of phonetics* (pp. 75-106). Routledge.
- [2] Podesva, R. J., & Callier, P. (2015). Voice quality and identity. *Annual review of applied Linguistics*, 35, 173-194.
- [3] Matar, N., Portes, C., Lancia, L., Legou, T., & Baidier, F. (2016). Voice Quality and Gender Stereotypes: A Study of Lebanese Women With Reinke's Edema. *Journal of Speech, Language, and Hearing Research*, 59(6), S1608-S1617.
- [4] Yanushevskaya, I., Gobl, C., & Ni Chasaide, A. (2018). Cross-language differences in how voice quality and f contours map to affect. *The Journal of the Acoustical Society of America*, 144(5), 2730-2750.
- [5] Ge, C., Xu, W., Gu, W., & Mok, P. P. K. (2023). The change in breathy voice after tone split: A production study of Suzhou Wu Chinese. *Journal of Phonetics*, 98, 101239.
- [6] Kiesling, S. F. (2011, April). Stance in context: Affect, alignment and investment in the analysis of stancetaking. In *iMean conference* (Vol. 15).