Sounds of Personality: Inference from Voices by Non-Native Speakers

Bin Li*, Yihan Guan*, and Si Chen†
*City University of Hong Kong, China
E-mail: binli2@cityu.edu.hk, yihanguan2@cityu.edu.hk
†The Hong Kong Polytechnic University, China
E-mail: sarah.chen@polyu.edu.hk

***Abstract*—People listen to get information in oral communication, and may consciously or unconsciously draw inferences on speakers based on their voices. A variety of phonetic cues have been found relevant in perceiving personalities from speech. This correlation has been documented in research involving native speech perception, though perceptual judgements on personality parameters were reported varying across cases. The tendency and sensitivity in voice perception has also attracted attention on non-native communication, where both language proficiency and language-specific features are considered influential as well. It is thus interesting to examine if similar or different sets of phonetic cues may affect non-native listeners' assessment of voices and personality.**

This study examined English-as-a-second-language (ESL) speakers' listening comprehension and perception of personality when listening to speech samples in American English. Speech extracts were modified to include variations in temporal and spectral dimensions. Results show that modification to pitch seemed beneficial to improve ESL listeners' comprehension accuracy. The modification also resulted in more favorable judgement of personality traits. Changes to the speaking rate yielded similar positive correlation with comprehension and personality judgement.

I. INTRODUCTION

Speech can be interpreted by human beings as an index of personality [1]. The interpretation relies not only on the overt content, but also on the vocal delivery [2][3]. Much inference or preference could thus be derived from voices, or even prosody only [4][5]. Over the past decades, acoustic analysis on native speakers' production and perception has revealed certain correlation between phonetic characteristics of speech and perceptual judgement of personality and emotion [6][7][8].

Characteristics such as pitch and speech rate are identified as correlates of physiological states and thus signal emotional status [3]. For example, extraversion as the best established dimension of personality is found correlated with raised pitch, faster speech, and fewer pauses [2]. But there has been mixed evidence regarding how pitch fluctuation is correlated with affective dimensions such as valence and potency [3]. There is also studies on personality judgement based on speech rate that present results from various aspects. For example, increased speech rate is found correlated with lower degree of agreeableness, but with higher evaluation of consciousness and openness to experience [9], as well as a higher level of competence [10]. However, correlates of personality and acoustic cues can be language and culture dependent, for instance, speakers with longer pauses are judged to be quiet and introvert in English, but anxious in German [11].

Much has been examined regarding personality judgement of native speakers on native speech, but research involving non-native or cross-language perception is scarce. As non-native speech perception is subject to more variables such as cultural and linguistic backgrounds, judgement of non-native personality traits may involve adjustment in expectations as to what to hear in a speaker of a different linguistic identity [12][13]. Following this direction, we find it worthy to explore whether non-native listening and perception of such kind would be more susceptible to language-specific features, and idiosyncratic features of speakers?

II. STUDY DESIGN

A. Research objective

We aimed at examining phonetic parameters' effects on perceptual judgement of personality traits and content comprehension by Chinese learners of English at advanced levels.

B. Speech stimuli

We selected and modified speech samples produced by a male, middle-aged, native speaker of American English. The samples were extracted from an online TED talk on the history of the English language [14]. The speaker is an accredited linguistics professor. We selected this video for its academic merits, speaking clarity, and appeal to the general public (YouTube statistics suggested that by August 2019, it has been viewed 323,239 times and liked by over 5,600 viewers). The total duration of the talk is 801 seconds with 2219 words in 141 utterances. First, as this was a public talk with high expected liveness and interaction with the audience, there were 134 seconds of long non-speech parts (rhetorical pauses, applause, laughter), which were manually removed from the

recording. Secondly, we calculated the number of words and morphemes in each clause (6.86 words or 8.64 morphemes per clause), which served as a reference for us to select and extract two sections of similar duration (160.25 seconds containing 64 clauses and 179.84 seconds containing 72 clauses) and of integrative discussion on subtopics. We used the audio of the extracted sections as bases for the manipulation of the two phonetic cues. As identified in previous research, the fundamental frequency and speech rate are two major prosodic correlates relating to and influencing personality perception. Our speaker in the chosen speech spoke at a mean rate of 2.75 words or 3.47 morphemes per seconds, and with a pitch range from 86-150 Hz (mean F0 at 145 Hz ±13.5 Hz). Our manipulation of speech samples was all completed using Praat [15] and tested for naturalness and clarity by naïve listeners including 2 advanced learners of English and 2 native speakers of American English.

We intended to modify the mean F0 in two directions to assess their effects: to raise and to lower the means by a semitone. Our preliminary test suggested that one semitone up, which brought the mean F0 to 153 Hz, resulted in an unnatural/unreal rendition of the voice sample. According to a summary of F0 statistics [16], our male speaker's pitch falls at the higher end towards acting and with greater fluctuations than those of an average adult male. So, only the other type of modification to pitch, 1 semitone down from the original (yielding a mean F0 at 136 Hz), was used for the study.

In addition to F0 modification, we also manipulated the speech rate: to raise and to lower the average rate. Our preliminary test showed that Chinese advanced learners of English found that the speaker talked quite fast in the original speech, and that his speed of speaking required much of their attention and efforts to accommodate. Listening to fast speech can be demanding and stressful, which would not serve our research purpose very well. Therefore, we lowered the speech rate by lengthening the samples by 1.2 times, which sounded natural to all naïve listeners.

In sum, two modified speech stimuli were generated by manipulating the pitch (F0) and speech rate (SR). Two original bases were also included in the experiment. Altogether, we had six speech stimuli in three sets, i.e., each set containing two speech samples in three conditions.

C. Assessment of Personality Traits

The Big Five personality model was adopted [17], which is also a commonly applied model in research on voice perception. It includes five broad dimensions to describe human personality and psyche as follows: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. We prepared ten statements, each two of which aimed to access one of the five personalities in the Big Five model. At the end, one statement was used to elicit assessment on teacher quality. Chinese translation of the statements was provided during listening.

Listeners were instructed to make a judgement of each statement using a 5-point Likert scale ranging from "Strongly disagree" to "Strongly agree" (1-5 points). The scores corresponding to each trait were obtained by numerical calculations of the points.

D. Listeners

Seventy-five university students of non-language related majors were recruited as listeners. They were all native speakers of Hong Kong Cantonese and had learned English since kindergarten age. Their English proficiency levels were comparable according to their self-reported scores in the entrance examination and performance in academic English courses. None reported speech or hearing issue. Students were randomly divided into three groups of 25, and each listened to one set of speech samples only.

E. Task

Listeners were asked to complete two tasks: listening comprehension and personality assessment. They were instructed to listen carefully to a short clip from a public talk on the history of the English language, and then to answer questions on the content and the speaker. The comprehension questions included a simple and direct one on the content, and another involving simple synthesis of information. This design was to ensure that our listeners paid attention during listening while being kept as innocent as possible to our research aim on voice perception.

After the comprehension questions, listeners were asked to complete a short questionnaire containing the 11 statements (Table 1) and to indicate their assessments of the speaker's personality traits on a 5-point scale.

TABLE 1: STATEMENTS IN PERSONALITY QUESTIONNAIRE.

Statement	Personality
1. The speaker is full of energy and passion.	Extraversion
2. The speaker is willing to convey what he knows to others.	Extraversion
3. The speaker is tolerable/ forgiving to others' mistakes.	Agreeableness
4. The speaker tends to embrace different opinions.	Agreeableness
5. The speaker has a clear logic.	Conscientiousness
6. The speaker is competent in his field.	Conscientiousness
7. The speaker is composed.	Neuroticism
8. The speaker seems approachable to me.	Neuroticism
9. The speaker has innovative ideas.	Openness to experience
10. The speaker is an ingenious and deep thinker.	Openness to experience
11. I would like to have him as a teacher at my school.	Teacher quality

III. RESULTS AND DISCUSSION

This study set out to examine correlation between voices and perceptual judgments of English speaker's personality by advanced ESL speakers. We used comprehension tasks as controls. Preliminary processing of data revealed that 29

listeners did not complete the whole experiment or failed to pass the 50% accuracy rate in listening comprehension. So, only responses from the remaining 46 listeners were included in data analysis (Table 2).

TABLE 2: MEAN % ACCURACY IN LISTENING COMPREHENSION. (Orig.: original samples; F0: stimuli with pitch lowered; SR: stimuli with speaking rate decreased)

	Sample A		Sample B	
	Q1	Q2	Q1	Q2
Group 1 (19)	Orig.: 87.5	Orig.: 31	Orig.: 64	Orig.: 37.5
Group 2 (16)	F0: 89.5	F0: 37	SR: 80	SR: 47
Group 3 (11)	SR: 91	SR: 48	F0: 72	F0: 62.5

Our results show that participants performed better in Question 1 than in Question 2, which was anticipated as the nature of Question 1 was factual while answering Question 2 required synthesis and interpretation of information. Secondly, accuracy rates in listening comprehension were higher when both speech samples were modified in F0 or speech rate. But pitch lowering led to better information gathering only (higher accuracy rate in Q1), whereas decreased speech rate seemed more beneficial as it improved both information gathering and information synthesis as well. This is not surprising. When a native speaker spoke more slowly, the ESL listeners may have gained more time and could allot more cognitive load for higher-level processing of information.

Personality assessment was calculated by extracting the mean of every two statements corresponding to each trait. For example, the score for extraversion was the mean of Statement 1 and Statement 2.

To assess whether Speech sample A and Speech sample B were comparable in terms of elicitation of personality judgement, a linear regression model was built based on the performance of participants in Group 1 where they all listened to the original version of Speech sample A and Speech sample B. Scores (five personality mean scores + one teacher quality score) were taken as the dependent variable, two speech samples and six dimensions (five personalities and one teacher quality), and their interaction, were treated as the independent variables. The results show that there was, in general, no difference between two samples (t = 1.34, p = 0.18), a pairwise comparison also indicated no difference between two samples across all six dimensions (all $|t| < 1.3$, ps > 0.18). This further proved that although two speech samples had different contents, they were comparable in eliciting judgements, and therefore could be combined as one reference in subsequent analysis and comparison.

To assess whether manipulated samples were comparable in eliciting responses, two linear regressions models like the previous one were built with one on F0 lowering and the other on speech-rate decrement. Results show no significant difference between two samples (both $|t| < 1.90$, ps > 0.06), and pairwise comparisons indicate similar results (all $|t| < 1.90$, ps > 0.06). This allowed us to combine data of two samples for integrative analysis.

Fig. 1 below presents average scores on six dimensions across manipulation conditions. Our listeners gave ratings at above 3 out of 5 for all statements, indicating that they were very positive regarding their assessments of the speaker. Statements on agreeableness received the lowest scores on average and those on conscientiousness the highest. So, our listeners thought the speaker sounded very conscientious, enthusiastic, composed and open-minded, but only just-above-average friendly and agreeable. And, our listeners would like to have the speaker as their teacher.

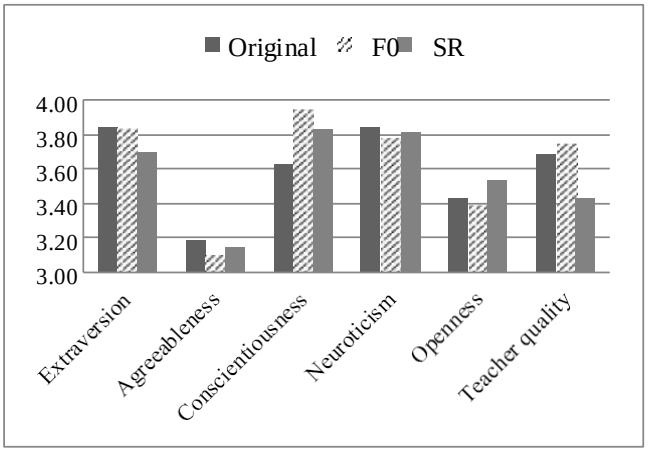

Fig. 1. Mean scores for all personality traits across manipulation conditions.

A linear mixed-effects regression model was built to examine and compare the scores in different manipulation conditions and along different personality dimensions. Mean scores were the dependent variable, and the independent variables include six dimensions (five personalities and one teacher quality), three manipulations (one original, one F0 manipulation and one SR manipulation) and their interaction. Random variables include listener difference (46 listeners) and group difference (three groups). Post-hoc pairwise comparisons using the Tukey method would be carried out when significance was shown on certain variables.

First, manipulation on F0 (lowering) elicited highest ratings on conscientiousness, followed by scores of extraversion, neuroticism, teacher quality, openness, and agreeableness. Across manipulation conditions, F0 condition (M=3.98, SD=0.59) correlated with a higher (but not significant) rating than the original ones (M=3.61, SD= 0.90) in the conscientiousness traits (pairwise t = 0.80, p = 0.68). As mentioned earlier, our speaker was a middle-aged male with mean F0 at 145 Hz and greater pitch fluctuations than average. The lowering of his pitches by a semitone brought F0 down to 136 Hz, which is still relatively high among average adult male speakers of American English. Higher pitches have been found among native listeners

associated with greater competence and activation, which actually belongs to conscientiousness and extraversion in our personality dimensions. It is reasonable, thus, that our listeners would rate the speaker very highly on traits relating to conscientiousness and extraversion, but not so on the opposite traits such as agreeableness or openness. This shows that the correlation between high pitches and traits relating to conscientiousness is not only present in native listening but also among non-native listening as well.

Secondly, the SR manipulation (speaking rate decrease) received the highest scores in conscientiousness and neuroticism, followed by extraversion, openness, and lastly agreeableness. Compared with the scores in the original condition, the most significant positive change lies in the ratings on conscientiousness (modified: M=3.83, SD=0.55 vs. original: M=3.61, SD=0.90) and openness (modified: M =3.53, SD=0.54 vs. original: M=3.46, SD=0.54); while the negative change in extraversion (modified: M= 3.70, SD=0.82 vs. original: M=3.82, SD=0.94), although no significance was shown (all $|t| < 1.1$, ps > 0.52). The positive correlation has been reported in literature on native English listening, that slow speaking (not far from the normal rate) is perceived as more competent (or, conscientious in our term) and ambitious, but inversely correlated in a U-shaped curve with benevolence such as being sincere and polite [18]. Our results from ESL listeners are interesting: compatible with and varied from major findings on native listening. Our speaker in the study when slowing down would sound, to non-native listeners, more conscientious yet more open but less extravert.

Lastly, we asked the listeners to evaluate on teacher quality. In general, they all found the speaker well suitable as a teacher. But, interestingly, lowering the F0 only slightly increased the rating with no significance revealed (t = -0.5, p = 0.86), while lower speech rate decreased their ratings with marginal significance only (t = -2.2, $p < 0.05$). As we had expected earlier, non-native listening involves more variables such as cultural expectation. Our ESL students who found slower speaking more composed and conscientious may have expectations or general impression regarding native speakers of English (NSE): NSE teachers are energetic and passionate. Therefore, they may conclude that for teacher candidates, an English person who speaks slowly though clearly is not as appealing as one that speaks faster.

IV. Conclusion

This paper presents a preliminary experiment on non-native perception and inferences of personality as a result of phonetic modification to voices in English. Prosodic correlates of voices were manipulated to examine if they would affect content comprehension and perceptual judgement of personality. Our participants improved in the listening comprehension tasks where the speaker's voice was modified in F0 and speech rates. The results show that lower rates in speaking have stronger influences on improving accuracy as the change may allow more allocation of cognitive loads in non-native listening. For perception of personality traits, lowered pitch is found associated with higher ratings on conscientiousness, and so is the speech rate. Speech rates also show a positive relationship with openness, and an inverse pattern with extraversion. For non-native listeners, being conscientious and composed is not as desirable or appealing for an NSE teacher as being passionate and energetic.

Our findings from Chinese ESL learners are compatible with those from previous studies on native English perception, which suggests universal effects of prosodic cues on personality perception and on voice analysis across languages and cultures. There are also unique findings from our listeners who seemed to have imposed their anticipation of an English speaker on their general judgement of personality. The adjustment in speech perception arising from cultural stereotypical expectation should be taken into consideration in intercultural and cross-linguistic research examining the relationship between voices and personalities.

V. Future Direction

This study offers preliminary empirical evidence that warrants further exploration of speech prosody and social or cognitive norms. Follow-up research could find experimental potentials from the following aspects. Firstly, there could be more diverse speech samples including both gender and age as social variables. Future research could also add on the variability of talker occupation and speech topics to examine how cultural and social concepts are phonetically encoded in our speech, and to assess effects of socioeconomic variables on personality perception. Secondly, a full version of the Big Five could be adopted for more comprehensive analysis of personality traits and to assess the prosodic correlations in more depth.

References

[1] E. Sapir, "Speech as a Personality Trait." *American Journal of Sociology*, vol. 32, pp892-905, 1927.

[2] M. Argyle, *Bodily Communication.* 2nd Ed. New York: Routledge, 1988.

[3] J. Harrigan, R. Rosental, K. Scherer, *New Handbook of Methods in Nonverbal Behavior Research*. New York: OUP. 2008.

[4] W. Apple, L. A. Streeter, R. M. Krauss, "Effects of pitch and speech rate on personal attributions". *Journal of Personality and Social Psychology*, vol. 37, pp715-727, 1979.

[5] J. J. Ohala, "An ethological perspective on common cross - language utilization of F0 of voice". *Phonetica* vol. 41, pp1−16, 1984.

[6] B. L. Brown, W. J. Strong, and A. C. Rencher, "Perceptions of personality from speech: Effects of manipulations of acoustical parameters". *The Journal of the Acoustical Society of America*, vol. 54(1), pp29-35, 1973.

[7] M. Zuckerman, and K. Miyake, K. "The attractive voice: What makes it so?". *Journal of Nonverbal Behavior*, vol. 17(2), pp119-135, 1993.

[8] K. B., Zellner, "Prosodic styles and personality styles: are the two interrelated?". *Proceedings of Speech Prosody 2004*. (pp383-6).

Nara, Japan; 2004.
[9] T. Uchida. "Impression of Speaker's Personality and Naturalistic Qualities of Speech: Speech Rate and Pause Duration". *The Japanese Journal of Educational Psychology*, vol. 53(1), pp1-13, 2005.
[10] B. L. Smith, B. L. Brown, W. J. Strong, and A. C. Rencher. "Effects of Speech Rate on Personality Perception". *Language and Speech*, vol. 18(2), pp145-152, 1975.
[11] K. Scherer. "Personality markers in speech". *Social markers in speech*. Cambridge: Cambridge University Press. pp147–209, 1979.
[12] C. D. Aronovitch. "The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker." *The Journal of Social Psychology*, vol. 99(2), pp207-220, 1976.
[13] P. Lukkarila, A. M. Laukkanen, P. Palo, P. "Influence of the intentional voice quality on the impression of female speaker." *Logopedics Phoniatrics Vocology*, vol. 37(4), pp158-166, 2012.
[14] J. McWhorter, "Txting is killing language". Retrieved from online on Jan. 2, 2018 https://www.youtube.com/watch?v=UmvOgW6iV2s
[15] P. Boersma, Praat, a system for doing phonetics by computer. Glot International vol. 5:9/10, pp341-345, 2001.
[16] H. Traunmüller, A. Eriksson, A. "The frequency range of the voice fundamental in the speech of male and female adults." Technical Report. Linguistics Department, University of Stockholm. Stockholm, Sweden, 1994.
[17] J. S. Wiggins, ed. *The Five-Factor Model of Personality*. Guildfor Press, 1996.
[18] B. L. Brown, "Effects of Speech Rate on Personality Attributions and Competency Evaluations". *Language: Social Psychological Perspectives*, H. Giles, W. P. Robinson, and P. M. Smith eds. Oxford, UK: Pergamon Press. pp293-300. 1980.