I. The Production Experiment

1.1 Participants

A total of 26 students whose mother tongue is Mandarin participated in this study, including 13 males and 13 females. Their average age was 20 years old, and the standard deviation was 1.61. All participants were undergraduate students in UESTC.

1.2 Tools

The main tools of this study include a PC using MacOS[①] system, RStudio[②] and Microsoft Excel[③].

1.3 Statistical analysis

We ran a series of linear mixed-effect models using the "lmer"[④] function of the lme4 package in R[⑤]. We take gender, attitude, speaker, item, meanf0 (average fundamental frequency), f0Span (the pitch span), meanIntensity, articulationRate, f0difference (the difference of pitch between deferential and non-deferential), f0Change (the group separated by f0difference), H1_H2 (the first and second harmonic difference) as dependent variables. The models are started with a full model which includes attitude (deferential, non-deferential) as a within-subject fixed factor and gender (male, female) as a between-subject fixed factor. We fit the data through the lmer function, and one of the models is as follows:

lmer(meanf0~attitude+gender+(1+attitude|speaker)+(1+1|item), data=RD, RMEL=F)

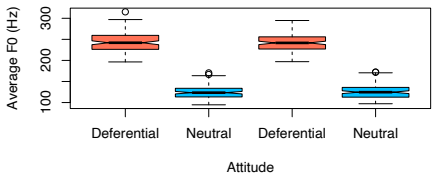
During the modeling process, according to the "avona" function in R, we remove the predictors from the model in turn from the least significant predictors until all the remaining predictors are statistically significant. During model comparison, the *p*-value was obtained by likelihood ratio test.

According to our observation, the best fitting model is fit_ 5：

lmer(meanf0~gender+(1+attitude|speaker)+(1+1|item), data=RD, REML = F)

II. Descriptive Statistics

2.1 Average fundamental frequency



① Apple. MacOs Download. From: https://support.apple.com/kb/DL2091?viewlocale=en_US&locale=en_US
② IDE. About RStudio. From: https://rstudio.com/about/
③ Microsoft. Excel. From: https://www.microsoft.com/zh-cn/microsoft-365/excel
④ D. Bates, M. Maechler, and B. Bolker, "lme4: Linear mixed-effects models using S4 classes," retrieved 1 May 2016. From: http://cran.r-project.org , 2006.
⑤ R Development Core Team. R Language Definition. From: https://cran.r-project.org/doc/manuals/R-lang.html

Figure2.1 Average fundamental frequency in different attitudes

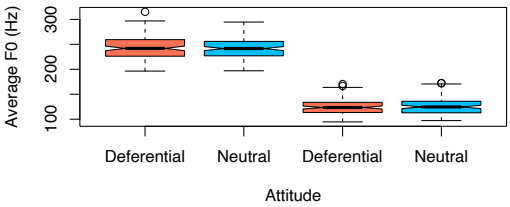


Figure2.2 Average fundamental frequency in different genders

According to figure 2.1 and figure 2.2, there is a significant difference of pitch between men and women in the same attitude: the median difference is large, and there is almost no overlap in the fluctuation range. The reason for this difference may be the congenital difference of biological human body structure; Otherwise, between the same gender, there are also differences in the pitch of expressing the deferential speech and the neutral speech. When expressing the respectful attitude, the pitch is significantly higher than that of expressing the neutral: there are differences in the median, and the upper quartile of the neutral is close to the lower quartile of the deferential.

2.2 Pitch span

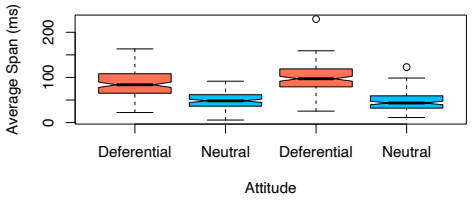


Figure2.3 Pitch span in different attitudes

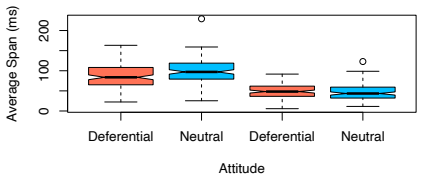


Figure 2.4 Pitch span in different genders

According to figure 2.3 and figure 2.4, there is no significant difference in the average pitch span between different sexes. According to boxplot, we can observe that the span median of speakers of different genders is close, and the fluctuation ranges of the upper and lower quartiles are more coincident; However, under different attitudes, the situation has changed. According to the difference between the median and the upper and lower quartiles, we can infer that there are differences between men and women when expressing words with different attitudes.

2.3 Mean intensity

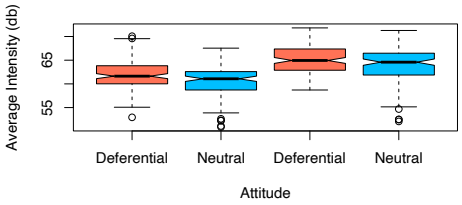


Figure2.5 Mean intensity in different genders

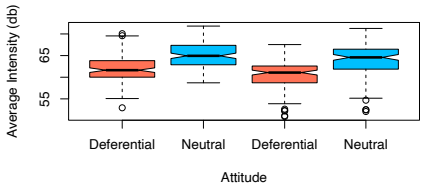


Figure2.6 Mean intensity in different attitudes

According to figure 2.5 and figure 2.6, there is no significant difference between the expression of respect and neutral speech between different genders: the median digits are close and the coincidence of the range of sound intensity changes is high; However, if we only look at the change of attitude, the intensity of the voice when expressing the deferential speech is significantly lower than that when expressing the neutral speech, which is also in line with common sense. When expressing respect, the tone is gentler, and the tone intensity is reduced.

2.4 Articulation rate

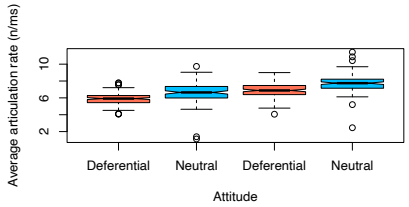


Figure2.7 Articulation rate in different genders

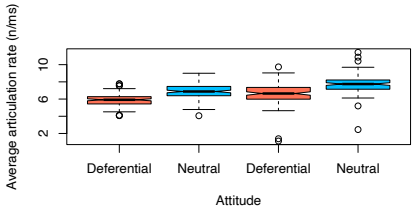


Figure2.8 Articulation rate in different attitudes

According to figure 2.7 and figure 2.8, the median of the data is very close between different sexes and different attitudes, and the most intuitive feeling of the entire boxplot is that it is deflated. Therefore, it can be concluded that there is no difference in the article rate between different genders and different attitudes in the case of specific speech and neutral speech, and the fluctuation range of the value is very small, indicating that the

article rate is a stable variable.

2.5 Breathiness

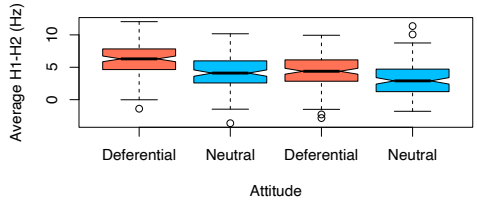


Figure2.9 Breathiness in different genders

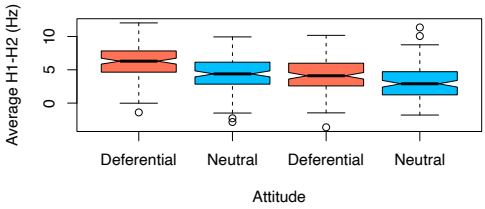


Figure2.10 Breathiness in different attitudes

According to figure 2.9 and figure 2.10, the speaker will have more obvious breathing sound than the neutral speech when expressing the deferential speech; In addition, women breathe more loudly than men. In other words, women breathe more clearly when talking.

III. Inference Statistics

According to the results of inference analysis, f0, f0Span and articulationRate of deferential and neutral speech are significantly different, and the *p*-value of their fitting model is less than 0.001. In addition, according to fit_full and fit_reduce model, f0 cannot be used to predict whether the speech is deferential because its *p*-value is greater than 0.05; f0Span can be used to predict, and the accuracy is more than 95%; meanIntensity, articulationRate and H1_ H2 can be used for prediction with an accuracy of more than 99.9%.

IV. Discussion and Conclusions

The main purpose of this study is to answer two questions. 1. Prosodic features of deferential speech in Mandarin. 2. What features can predict deferential speech and their accuracy.

According to the above results, we know that when expressing deferential speech, speakers usually use faster speech speed, higher average pitch, and more obvious breathiness voice than neutral speech. According to the results of inferential statistics, we can use the pitch span, the mean intensity, the articulation rate and the first and second harmonic difference to predict whether the speaker is deferential. The prediction accuracy of pitch span is more than 95%, and that of other variables is greater than 99.9%.

References

[1] Apple. MacOs Download. From: https://support.apple.com/kb/DL2091?viewlocale=en_US&locale=en_US

[2] D. Bates, M. Maechler, and B. Bolker, “lme4: Linear mixed-effects models using S4 classes,” retrieved 1 May 2016. From: http://cran.r-project.org , 2006.

[3] IDE. About RStudio. From: https://rstudio.com/about/

[4] Microsoft. Excel. From: https://www.microsoft.com/zh-cn/microsoft-365/excel

[5] R Development Core Team. R Language Definition. From: https://cran.r-project.org/doc/manuals/R-lang.html