# Communication Effectiveness and Respect:
# Analyzing the Impact of Acoustic Features on Voice Preferences of Elderly Adults

高敏瑄

Kang Chiao International School Xiugang Campus 11th-grade

Keywords: Elders, Voice Preference, Praat, Acoustics

Instructor: Su Hung Kuan

September 2024

# Chapter 1 Introduction

People always seek the most "effective" way to communicate with elders. Noted by Lowery (2013), communication with elders, the so-called "elder speech", is characterized by simplified wordings, exaggerated intonation, slower speech rates, or increased volume. While it is important to know the most effective way to communicate with older adults, the most effective approach is not always the most respectful or pleasing one. Ryan, et al. (1995) pointed out that purposely differentiating elders from the younger generation was sometimes perceived as patronizing, which hinders intergenerational communication. Despite this, few researchers have addressed the issue of the voice preferences of elders.

The present study aims to examine these preferences, which are represented by different acoustic features, particularly focusing on the acoustic features that impact elders' perception of voices and how different factors influence each other. The research questions are as follows:

1) What types of voices are preferred by older adults?
2) How do different acoustic cues influence elder's voice preferences?

The present study involves recording stimuli and modifying different acoustic features on Praat to create various experimental groups. By manually adjusting stimuli, each factor can be strictly controlled while minimizing the risk of the voices sounding unnatural or computer-generated. In addition, multinomial logistic regression, one of the generalized linear models, is used to assess the ranking of each acoustic feature and their interactions. The present study hopes that the findings could provide the public with a better understanding of elders' voice preferences, enhancing communication with older adults.

# Chapter 2 Literature Review

This section reviews key acoustic features relevant to the current study.

Firstly, fundamental frequency (F0) is the lowest frequency of a periodic waveform. Reby et al. (2012) found that men typically prefer higher-pitched women's voices, though not exceeding 300 Hz, while women's preferred pitch is around 96 Hz, close to men's natural pitch. Hollien et al. (1991) found no consistent differences in voice preferences across age or gender.

Secondly, intensity refers to the power carried by sound waves (Filippi et al., 1999). Ilie and Thompson (2006) observed that louder sounds tend to be perceived as less pleasant and more tense, with similar acoustic cues between speech and music.

Thirdly, speech rate will be analyzed. Janicki and Szczypiorski (2015) found that slower speech improves intelligibility for older adults. Lessa et al. (2013) showed that elders, regardless of hearing impairment, recognize speech more effectively when delivered at slower rates, implying a negative correlation between speech rate and comprehension.

Fourthly, jitter, or irregular vocal vibrations, will also be considered. Rozsypal & Millar (1979) determined that jitter is crucial for a natural-sounding voice, though high jitter levels are seen as pathological (Suire et al., 2019). Pinto et al. (2014) linked higher jitter values with smokers and heavy drinkers.

Finally, HNR (Harmonic-to-Noise Ratio), which measures vocal breathiness, will be evaluated. Hear (2023) and Ting (2021) indicated that breathiness, particularly in men, is associated with attractiveness (Sebesta et al., 2017). However, Borsel et al. (2009) found that high breathiness in male voices is often perceived as abnormal.

Regarding these acoustic features and elder speech, research suggests that slower speech rates and simpler sentences can improve communication with older adults (Kemper & Harden, 1999; McGuire et al., 2000). However, over-accommodation, such as exaggerated pitch changes, can be perceived as disrespectful (Ryan et al., 1995; Lowery, 2013).

# Chapter 3 Methodology

## 3.1 Materials

Two positive sentences ("你的心穩了，一切的事情都會穩" and "心中有陽光，照到哪裡都溫暖") were recorded by four native Mandarin speakers (2 males, 2 females), aged 30–40, in a soundproof environment, ensuring neutral intonation without any background noise (Lin, 2021). Each recording lasted between 2 to 4 seconds. The recordings were manipulated using Praat software, altering the pitch by +/- 50 Hz and speech rate by +/- 0.3x, as per the scale from Hollien et al. (1991). In total, 72 stimuli were generated (4 speakers x 2 sentences x 3 pitch levels x 3 speech rates).

## 3.2 Procedure

The study recruited 31 elderly participants (12 males, 19 females), divided into three age groups: 65–70, 70–75, and 75+. All participants were native Taiwanese or Mandarin speakers with no hearing disorders. Participants rated the 72 randomized audio stimuli on a scale from 1 to 4 (1: disfavored, 2: slightly disfavored, 3: favored, 4: very favored). The experiment took place in a quiet room, free from distractions, lasting around 15 minutes. When more than one participant was present, a partition was used to prevent distraction or influence between participants. The acoustic features, including pitch, intensity, duration, jitter, and HNR, were analyzed using Praat. Pitch and intensity were measured in Hz and dB, respectively, while the speech rate was calculated in syllables per second.

### 3.3 Factor Analysis

The dataset will be analyzed using multinomial logistic regression, a Generalized Linear Model (GLM) suitable for datasets with multiple categorical dependent variables. This method will assess the influence of various acoustic features on participants' voice preferences. The Akaike Information Criterion (AIC) will be used to compare the quality of different statistical models. AIC is an estimator that assesses the goodness of fit of a model while penalizing for the complexity of the model. Lower AIC values indicate a model that balances fit and simplicity, which is especially useful when comparing multiple models. This helps in selecting the most efficient model with the least prediction error (Klee, 2008).

## Chapter 4 Results and Analysis

### 4.1 Stepwise Selection

In this section, the dataset adopts a method to compare models and eliminate insignificant factors. Factors with less significance can be observed with a reduced AIC value, marked with the * sign after each value. The factor with the lowest AIC value will be eliminated from the model one at a time in each selection round. This step will be repeated until all AIC values are greater than the original AIC (AIC value obtained without any factors being removed) and no further reductions can be made. Tables 1 to 4 exhibit the process of a step-wise feature selection.

The first step of variable selection is demonstrated in Table 1. The original model with all predators had an AIC value of 2939.18. Removing certain variables reduced the AIC: *pitch mean* (-1.34), *intensity mean* (-1.96), *HNR* (-1.75), and *rater's first language* (-0.79). -1.96 means that when *intensity mean* was removed, the AIC was lowered by 1.96 to the value of 2937.22 (see third row in Table 1). If a variable gets the lowest negative value, the variable is considered to be least influential because removing the variable can improve the model prediction the most. In this case, *intensity mean* was removed due to the lowest AIC value. Variables including *pitch range*, *intensity range*, *speech rate*, *jitter*, *gender*, and *age* were retained as their removal increased AIC values, which means that the removal of these variables would decrease the accuracy of the model prediction.

Table 2 displayed the model after removing *intensity mean*, lowering the AIC value to 2937.22. In this model, *pitch mean*, *HNR*, and *rater's first language* were candidates for removal due to lower AIC values when excluded. Finally, *pitch mean* was removed in the next model fit due to its lowest contribution.

Table 3 demonstrated the model after removing *pitch mean*, further reducing the AIC to 2935.85. *HNR* and *rater's first language* were next likely to be removed due to better model fit upon their exclusion. Thus, *HNR*, which lowered the AIC the most after removal (AIC from 2935.85 to 2934.35), was removed from the model.

Table 4 exhibited the model after removing *pitch mean*, reducing the AIC to 2934.35. In this iteration, *rater's first language* was the most likely to be removed, leading to the best model (Table 5) with an AIC of 2933.55. The remaining features including *pitch range, intensity range*, *speech rate*, *jitter*, *gender*, and *age* (AIC = 2950.01) are included in the final model due to their higher AIC value after removal.

Last but not least, the best model (see Table 5) was obtained after removing *rater's first language* from the previous model, with an AIC value of 2933.55. For the remaining parameters, since the elimination of these would increase AIC values, they pose a relatively significant effect on the model fit. The factors include *pitch range* (AIC = 2952.66), *intensity range* (AIC = 2943.71), *speech rate* (AIC = 3042.76), *jitter* (AIC = 3014.90), *gender* (AIC = 2936.52), and *age* (AIC = 2948.31).

*Table 1. Feature Selection: Step 1*

| Removed factor | Df | AIC |
|---|---|---|
| None | 0 | 2939.18 |
| Pitch_mean | 1 | 2937.84* |
| Pitch_range | 1 | 2955.39 |
| Intensity_mean | 1 | 2937.22* |
| Intensity_range | 1 | 2948.03 |
| Speech_rate | 1 | 3048.60 |
| Jitter | 1 | 2980.02 |
| HNR | 1 | 2937.43* |
| Gender | 1 | 2942.90 |
| Age | 1 | 2954.86 |
| Rater_1lg | 1 | 2938.39* |

*Table 2. Feature Selection: Step 2*

| Removed factor | Df | AIC |
|---|---|---|
| None | 0 | 2937.22 |
| Pitch_mean | 1 | 2935.85* |
| Pitch_range | 1 | 2954.22 |
| Intensity_range | 1 | 2948.10 |
| Speech_rate | 1 | 3046.27 |
| Jitter | 1 | 2984.88 |
| HNR | 1 | 2936.03* |
| Gender | 1 | 2940.94 |
| Age | 1 | 2952.90 |
| Rater_1lg | 1 | 2936.43* |

*Table 3. Feature Selection: Step 3*

| Removed factor | Df | AIC |
|---|---|---|
| None | 0 | 2935.85 |
| Pitch_range | 1 | 2955.02 |
| Intensity_range | 1 | 2946.11 |
| Speech_rate | 1 | 3045.03 |
| Jitter | 1 | 2982.91 |
| HNR | 1 | 2934.35* |
| Gender | 1 | 2939.57 |
| Age | 1 | 2951.52 |
| Rater_1lg | 1 | 2935.06* |

*Table 4. Feature Selection: Step 4*

| Removed factor | Df | AIC |
|---|---|---|
| None | 0 | 2934.35 |
| Pitch_range | 1 | 2953.46 |
| Intensity_range | 1 | 2944.51 |
| Speech_rate | 1 | 3043.62 |
| Jitter | 1 | 3015.74 |
| Gender | 1 | 2938.07 |
| Age | 1 | 2950.01 |
| Rater_1lg | 1 | 2933.55* |

*Table 5. Feature Selection: Step 5*

| Removed factor | Df | AIC |
|---|---|---|
| None | 0 | 2933.55 |
| Pitch_range | 1 | 2952.66 |
| Intensity_range | 1 | 2943.71 |
| Speech_rate | 1 | 3042.76 |
| Jitter | 1 | 3014.90 |
| Gender | 1 | 2936.52 |
| Age | 1 | 2948.31 |

After feature selection, we used the summary() command from R to produce the model-fitting results (see Table 6). We can compare the estimated coefficient and *p*-value to identify which feature is significant. The one with the greatest estimated coefficient poses the most significant effect on the model and is thereby the most influential factor. With an alpha value of 0.001, the most influential factor was *speech rate* (z(1) = 10.303, $p$ < 2e-16), with a high estimated coefficient of 0.514123. Next, *jitter* (z(1) = -8.871, $p$ < 2e-16) was also an influential factor whose absolute value of coefficient was the highest (-1.451567). *Pitch* *range* (z(1) = -4.566, *p* < .001) was considered the third important factor with an estimated coefficient of -0.005011. Then, *intensity range* (z(1) = 3.478, *p* < .001) was the fourth with an estimated coefficient of 0.046912. Furthermore, with an alpha value of 0.01, *age3* (z(1) = 2.997, *p* < .01) was also tested significantly essential to voice preference, with the estimated coefficient of 0.384413.

*Table 6. Acoustic Contribution Ranking*

| | **Estimate Coefficient** | **Standard error** | **z-value** | **Pr(>\|z\|)** | **Significant Codes** |
|---|---|---|---|---|---|
| Pitch_range | -0.005011 | 0.001097 | -4.566 | 4.97e-06 | *** |
| Intensity_range | 0.046912 | 0.013489 | 3.478 | 0.0005051 | *** |
| Speech_rate | 0.514123 | 0.049900 | 10.303 | < 2e-16 | *** |
| Jitter | -1.451567 | 0.163625 | -8.871 | < 2e-16 | *** |
| Gender_m | -0.206752 | 0.092877 | -2.226 | 0.026008 | * |
| Age2 | -0.193180 | 0.097830 | -1.975 | 0.048307 | * |
| Age3 | 0.384413 | 0.128279 | 2.997 | 0.002729 | ** |

*p<.05; **p<.01; ***p<.001

4.2 The Influence of Acoustic Features on Voice Preference

The following section discusses parameters that obtain alpha values of .001, including *speech rate, jitter, pitch range,* and *intensity range.*

As mentioned in the previous section, *speech rate* was the most influential factor in elders' voice preference (preferred: 3.62 syl/sec; unpreferred: 3.36 syl/sec) with a positive estimated coefficient (see Table 6). That is to say, with elders favoring faster *speech rates* over slower ones, a positive correlation was shown between the rater's preference and *speech rate*. The present study aligns with Ryan, et al. (1995), who argue that intentional adjustments like slower speech rates can make elders feel patronized. However, since this area hasn't been widely investigated, we don't know the normal range of speech rate for Chinese. The present study can only extrapolate that a speech rate of 3.36 syllables/sec is too slow for older adults to understand and prefer.

*Jitter* is the second most contributing acoustic feature on voice preference (preferred = 1.82%, not preferred = 1.97%). Likewise, with a negative coefficient yielded by *jitter*, a negative correlation was demonstrated between the rater's preference and voice disturbance.

A study by Pinto et al. (2014) examined the influence of smoking on voice quality and discovered that the absence of jitter in human voices correlates with one's smoking behavior (estimated coefficient: 0.679; SD: 0.1283; *p*: 0.003). Further, Suire et al (2019) also suggested that jitter posed negative effects or indications of one's impression. Therefore, the present study aligned with previous research that the high presence of jitter was not preferred in voices.

As shown in Table 6, *pitch range* yielded a negative coefficient, indicating that the higher the pitch range, the less likely the elderly are to prefer the voice. This finding can related to Reby et al. (2012) in which they claim that a pitch that is either too high (men: 300 Hz; women 600 Hz) or too low (men: 50 Hz; women: 100 Hz), is perceived as unnatural by participants. A wider pitch range might run the risk of going beyond the normal range and thus might not be preferred. However, it does not imply that a smaller pitch range necessarily leads to a greater preference among the elderly. It should be noted that pitch is perceived as unnatural for both high and low, similar to jitter.

For *intensity range*, a positive coefficient was obtained. In the acoustic measurements in the present study (see Section 3.3), with a similar maximum intensity value for all stimuli, the minimum value posed a greater influence on *intensity range*. That is, a lower minimum intensity value indicates a higher intensity range. Similar to a study by Ilie and Thompson (2006) that suggested a lower intensity was perceived as more pleasant, less energetic, and less tense, the present research aligns with the previous paper that a positive correlation is obtained between the rater's preference and *intensity range*. It is worth noting that there is little research on the perception and preference of intensity and pitch range. Further exploration into how these ranges are perceived could yield interesting results. The present study serves as a preliminary research and future study can be done to help us understand this phenomenon more.

4.3 The Influence of Participant Background on Voice Preference

Several background information about participants were recorded during the questionnaire, including age, gender, and mother tone (see Section 3.2.1). Based on row 3 in Table 7, elders over 75 years old preferred a relatively higher *pitch range* (226.40 Hz), lower *intensity range* (39.77 dB), slower *speech rate* (3.43 w/s), and lower *jitter* (1.66%). The present study aligned with previous research that older adults tend to prefer a lower speech rate and pitch with moderate level (Edwards & Noller, 1993; McGuire et al., 2000). *Jitter* was not preferred in human voices and was often perceived as a pathological or negative character (Suire et al., 2019). For *intensity range*, the present study focused on the discussion on minimum *intensity* (see Section 4.2). According to Ilie and Thompson (2006), a higher *intensity* was rated as tenser, more energetic, but unpleasant. Similar to that, the present study concluded that a lower *intensity*, in general, was preferred rather than a higher one.

*Table 7. Preferred Acoustic Features from Different Age Groups*

| **Age** | **Pitch range** | **Intensity range** | **Speech rate** | **Jitter** |
|---|---|---|---|---|
| 65-70 | 221.82 | 39.88 | 3.73 | 1.86 |
| 70-75 | 223.80 | 40.16 | 3.58 | 1.80 |
| 75+ | 226.40 | 39.77 | 3.43 | 1.77 |

# Chapter 5 Conclusion

The present study investigates the acoustic features in voice preferred by older adults. Results show that pitch range, intensity range, speech rate, and jitter are influential factors contributing to elder voice preference. That is, the present study concludes that acoustic features preferred by older adults include a lower *pitch range*, higher *intensity range*, higher *speech rate*, and lower *jitter*. In addition, by comparing preferences based on age groups, older adults demonstrate a more consistent result with the trend discovered by the present study. Thus, the present study not only supports previous research about human voice preference but also recognizes the slight variation in preference among participants from different age groups.

However, the present study acknowledges several limitations in participant backgrounds. For the contribution of different age groups, the present study aims to dive deeper into the significance of each parameter through the compliance of another GLM model. The distribution among participant backgrounds (such as gender or mother language) was also unequal. That said, in the future, the present study seeks stricter participant limits while taking more participant backgrounds into account, such as participants' careers, living environments, or educational degrees.

## References


Edwards, H., & Noller, P. (1993). Perceptions of overaccommodation used by nurses in communication with the elderly. *Journal of Language and Social Psychology, 12,* 207–223.

Filippi, P., Bergassoli, A., Habault, D., & Lefebvre, J. P. (1998). *Acoustics: Basic physics, theory, and methods*. Elsevier.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research, 37*(4), 769–778.

Hollien, H., Gelfer, M. P., & Carlson, T. (1991). Listening preferences for voice types as a function of age. *Journal of Communication Disorders, 24*(2), 157–171.

Ilie, G., & Thompson, W. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23*, 319–330. https://doi.org/10.1525/mp.2006.23.4.319

Janicki, A., & Szczypiorski, K. (2015). Why do older adults prefer some radio stations? Helping to increase speech understanding. *Journal of Communication, 10*(11), 926–931.

Kemper, S., & Harden, T. (1999). Experimentally disentangling what's beneficial about elderspeak from what's not. *Psychology and Aging, 14*, 656–670.

Lessa, A. H., & Costa, M. J. (2013). The impact of speech rate on sentence recognition by elderly individuals. *Brazilian Journal of Otorhinolaryngology, 79*(6), 745–752. https://doi.org/10.5935/1808-8694.20130136

Lowery, M. A. (2013). Elderspeak: Helpful or harmful? A systematic review of speech to elderly adults (Doctoral dissertation, Auburn University).

McGuire, L. C., Morian, A., Codding, R., & Smyer, M. A. (2000). Older adults' memory for medical information: Influence on elderspeak and notetaking. *International Journal of Rehabilitation and Health, 5*, 117–128.

Mohammed, E. A., Naugler, C., & Far, B. H. (2015). Emerging business intelligence framework for a clinical laboratory through big data analytics. In Elsevier eBooks (pp. 577–602). https://doi.org/10.1016/b978-0-12-802508-6.00032-6

Pinto, A. G. L., Crespo, A. N., & Mourão, L. F. (2014). Influence of smoking isolated and associated to multifactorial aspects in vocal acoustic parameters. *Brazilian Journal of Otorhinolaryngology, 80*(1), 60–67.

Re, D. E., O'Connor, J. J., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PLOS ONE, 7*(3), e32719.

Rozsypal, A. J., & Millar, B. F. (1979). Perception of jitter and shimmer in synthetic vowels. *Journal of Phonetics, 7*(4), 343–355. https://doi.org/10.1016/s0095-4470(19)31069-1

Ryan, E. B., Hummert, M. L., & Boich, L. H. (1995). Communication predicaments of ageing: Patronizing behavior towards older adults. *Journal of Language and Social Psychology, 14*, 144–166.

Šebesta, P., Kleisner, K., Tureček, P., Kočnar, T., Akoko, R. M., Třebický, V., & Havlíček, J. (2017). Voices of Africa: Acoustic predictors of human male vocal attractiveness. *Animal Behaviour, 127,* 205–211.

Shen, J., Heller Murray, E., & Kulick, E. R. (2023). The effect of breathy vocal quality on speech intelligibility and listening effort in background noise. *Trends in Hearing, 27*.

Suire, A., Raymond, M., & Barkat-Defradas, M. (2019). Male vocal quality and its relation to females' preferences. *Evolutionary Psychology, 17*(3).

Ting, L. Y. (2021). 聲音特質與文字的情緒向性對於聲音偏好度之影響 [Master's thesis, National Taiwan Normal University]. Airiti Library.

Van Borsel, J., Janssens, J., & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: A perceptual approach. *Journal of Voice, 23*(3), 291–294.