Why Judy Sounds *Older* in English but *Younger* in Mandarin:

The Influence of Acoustic Features on Human Perceptions of Movie Dubbings

KAO, MIN-HSUAN

Kang Chiao International School Xiugang Campus 10th-grade

Keywords: Praat, Acoustics, Movie Dubbings, Perception

Instructor: Su Hung Kuan

September 2023

1 Introduction:

Judy, a young and energetic bunny, is the protagonist of the movie Zootopia. However, its dubbings in English and Mandarin reveal different images of this character. According to Liu et al (1991), humans often employ the properties of sounds to perceive the characteristics of environmental aural sources. In other words, acoustic features, or sound properties of speech, such as pronunciation or word length, can be one possible factor that determines humans' perception.

Previous studies have examined the acoustic properties behind the perceptions of different personalities. For example, Cheng (2013) discovered voices with a higher pitch or a more significant pitch variation in Mandarin are regarded as more energetic, confident, and extroverted in terms of cartoon dubbing. Additionally, according to Park et al (2020), English speakers tend to adopt a slower speech rate to convey a more introverted personality.

Having stated that acoustic features can be a key factor in human perceptions of character traits in a language, little has been explored about the difference in more than one language. That is, would different languages utilize similar acoustic strategies to express similar personal traits? Therefore, varied from the previous research, by selecting dubbings from the same movie character, the present study examines the difference between English and Mandarin speeches and discovers whether acoustic features can be a possible factor in influencing human perceptions. One important merit of using dubbings from the same character's words is that subjects could focus more on the acoustic features of the two dubbings without being affected by the semantics of the sentences since the content of the sentence of the two dubbings is semantically similar and the only difference is the dubbing language. The research questions are as follows:

1) How do Mandarin speakers perceive English and Mandarin dubbing?
2) How did the character ratings by audience participants correlate to acoustic features measured by Praat?

The present study hopes that the findings from this research could help future speakers to shape their personal traits more thoroughly in speeches and understand the connections between acoustic features and personal traits.

2 Literature review:

2.1 Acoustic features

Several determined acoustic features will be introduced and discussed in the following sections.

2.1.1 Fundamental frequency

Fundamental frequency (F0) refers to the lowest frequency of a periodic waveform. According to Keating et al (2010), the F0 in different languages may tend to be different. For instance, Mandarin speakers chose a higher minimum pitch (133 Hz) than English speakers (115 Hz). Furthermore, while both male and female Mandarin speakers chose a higher starting pitch of 171 Hz, English speakers have a lower starting pitch of 165 Hz than Mandarin. Additionally, for age differences, Hollien and Shipp (1972) explore the fundamental frequency of different age groups, observing that 20 to 29 years old obtained a fundamental frequency of 119.5 Hz, 30 to 39 with 112.2 Hz, 40 to 49 with 107.1 Hz, 50 to 59 with 118.4 Hz, 60 to 69 with 112.2 Hz, 70 to 79 with 132.1 Hz and 80 to 89 with 146.3 Hz, displaying a U-shaped relationship.

2.1.2 Speech rate

Referring to Apple et al (1979), speakers with a slower speech rate were judged to be less extroverted and more emotional. Alternatively, another similar experiment presented the personality judgments in terms of acoustic manipulation (Brown et al 1972). The study concludes that with the increase in one's speech rate, judgments regarding personality are often less "benevolent". The study also justified that speech rate is considered to be a better factor in determining personality than pitch variation, demonstrating the significant role speech rate plays in evaluating one's character.

2.1.3 Jitter and shimmer signal

Jitteriness in the linguistic criteria is variations in frequency and amplitudes that result from irregular vocal fold vibration, providing a more coarse and breathy voice (Backstrom, 2022). According to Backstrom, subjects exhibiting higher jitter often obtained a lower frequency and intensity variability whilst reporting a high voice intensity presence.

2.2 Connections between personality and acoustic features

Cheng (2013) investigated the variation in pitch, speech rate, along with voice qualities that included lax/tense and clear/turbid. The study puts forth that a higher pitch or a greater pitch variation in speech refers to characters that are more energetic, confident, and extroverted. Additionally, the speech rate of the character only elicits characteristics of being lazy/energetic and introverted/extroverted, allowing us to extrapolate that

individuals that are energetic may also be more extroverted.

In addition, Liu et al (2010) claimed that large and small response magnitude is measured by comparing the baseline pitch before the stimulus to the highest pitch after the response starts, which is similar to the measurements of pitch variation. Liu stated that pitch variation correlated to age difference. Another similar experiment presented the personality judgments in terms of acoustic manipulation also emphasized the role speech rate plays in evaluating one's character (Brown et al 1972).

3 Methodology:

The present study designed a questionnaire to investigate subjects' perceptions of different dubbings in English and Mandarin. Afterward, Praat is utilized to analyze the acoustic features behind each dubbing.

3.1 Materials

The recordings are excerpted from two animations on Disney Plus: (1) *Coco*, a film about the main character Miguel who travels from the Land of the Livings to the Land of the Deaths to visit his musical idol as he acknowledges the importance of a family (2) *Zootopia*, a movie regarding the journey of the protagonist Judy of overcoming all obstacles and ultimately becoming a competent officer. The selection of the two movies is based on their character variety in all personality, gender, and acoustic aspects. The traits of the character, such as roles that are not human in *Zootopia*, are also one of the investigations we are focusing on in the project. Further, the preprocess of extracting the recording includes the background music, minimizing overlapping voice, and limiting the time of the excerpt to be consistent between 3 to 6 seconds.

Alternatively, with regards to the table below, we selected a total of six characters with three characters from each movie, including Imelda, Hector, and Miguel from *Coco*, along with Judy, Nick, and Chief Bogo for *Zootopia.* Details about the characters and experimental materials can be found at the following link: http://surl.li/jtfvx.

3.2 Questionnaire and Procedures

3.2.1 Structure and Participants

The questionnaire is divided into two main focuses: 1) investigating how different language dubbings affected people's selection of personality traits; 2) examining the

participants' preference for dubbings in different languages. We included a total of 76 participants with an age range between 12-15 years old with 55% of male participants and 45% of female participants. Around 93.4% of participants selected Chinese as the language they used most frequently before the age of seven. The survey includes six groups of contrasting adjectives: lazy/energetic, young/old, cautious/rash, self-doubting/confident, introverted/extroverted, and human/non-human, which will be rated through a four-point scale for each recording by the participants. The selection of the adjectives is mainly based on either previous studies or the characteristics of the movie character chosen. The experiment contains a total of 24 groups of recordings with 4 recordings for each movie character (4 recordings*6 characters=24 recordings). In each group, there are 2 recordings, which include the same context but are spoken in different languages. The sequence of all recordings is scrambled to reduce the learning effect.

3.2.2 Data collection

The questionnaire was conducted by presenting the experiment to a large group of audience in a single session without any interruptions. The sequence began with a brief introduction of the main idea of the project with complete assurance of ethical considerations. Subsequently, the experimental process involved participants listening to an audio file, answering corresponding questions, moving on to the next audio file, answering the respective questions, and ultimately making a preference decision, which the process will be fully explained by the experimenter before being carried out. The entire experiment lasts from 30 to 45 minutes.

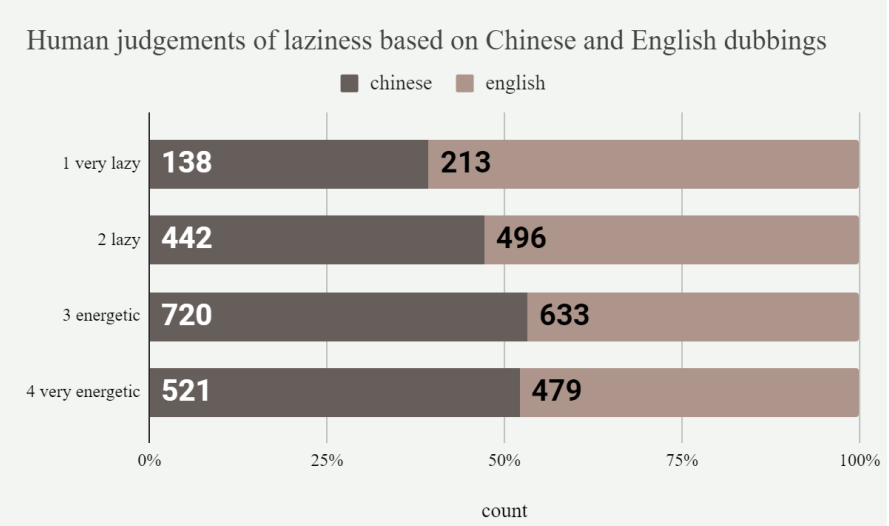
3.3 Measurements of recordings

A total of 24 excerpts were analyzed through the online program Praat. For each recording, we recorded the value of fundamental frequency (F0, F1, F2), intensity, and speech rate. For F0 and intensity, we recorded the maximum, minimum, and range of all data, whilst for F1 and F2, we computed the mean of each value. The speech rate was calculated by dividing the number of words in the sentence by the duration of the recording with the unit of (words/seconds). The duration of each excerpt was restricted to 3-6 seconds.

4 Results and analysis

This section focuses on three main findings: (1) Chinese dubbings are more “energetic” (2) English dubbings are more “old” (3) Chinese dubbings are more “extroverted". The present study focuses on how Mandarin speakers perceive two languages differently. However, the present studies indeed observed similarities between the two languages. For instance, no obvious difference is perceived in the subjects’ judgment of “self-doubting” and “confident”. The observation might serve as support for sound symbolism as a cross-linguistic phenomenon (Johansson et al, 2020). Nevertheless, due to the limitation of the abstract length, the present study would only show the differences between the two languages.

4.1 Chinese dubbings are more “energetic”



[Figure 4.1 Human judgments of laziness based on Chinese and English dubbings]

Based on Figure 4.1, Chinese dubbings are considered to be “more energetic” than those in English. Chinese dubbings received more votes than English dubbings in categories 3 (“energetic”) and 4 (“very energetic”) while English dubbings received more votes than Chinese dubbings in categories 1 (“very lazy”) and 2 (“lazy”).

Group (Language)	Pitch variation	Max Intensity	Speech Rate (w/s)
5a (Chinese)	365.04	84.56	4.42
5b (English)	402.16	85.12	4.50
8b (Chinese)	172.40	84.75	4.54
8a (English)	160.81	72.92	3.99
17a (Chinese)	292.16	84.49	3.23
17b (English)	189.18	72.46	2.82
20b (Chinese)	365.78	82.73	2.34
20a (English)	140.49	70.94	2.18

[Table 4.1 The Praat measurement for Hector]

In order to understand raters' perception of the dubbings, the present study used Praat to analyze the acoustic properties of the dubbings. Further, owing to the fact that different movie characters utilize different dubbing strategies (Cheng, 2003), the present study will provide the Praat measurement for the typical example that refers to the characters in which most subjects consider the Chinese dubbing to be more "energetic". Therefore, the character Hector, who is rated to be more energetic in all four dubbings, becomes the typical example for the following analysis. The results for the Praat measurement are in Table 4.1. The present study has the following observations:

(1) In terms of pitch variation, the present study suggested the higher the pitch variation, the more often the dubbing is marked as "energetic".

(2) The intensity of the recording also plays a role in the audience's perceptions of laziness. To elucidate, audio recordings exhibiting higher intensity averages are often perceived as more energetic, hence implying Chinese dubbings may include higher intensity than English dubbings.

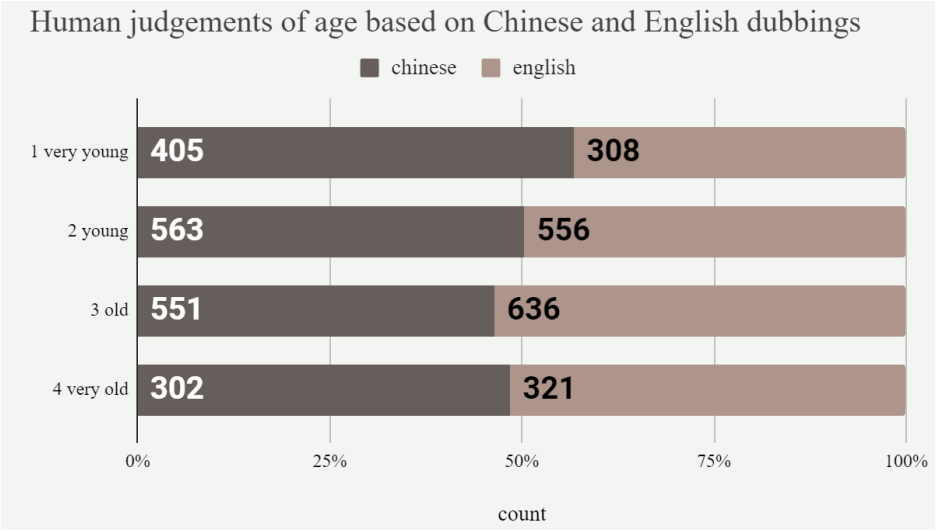
(3) For the speed of speech, while aiming to convey the dubbing in a more energetic tone, the speech rate tends to be faster.

These observations align with findings from previous studies. While previous research (i.e. Cheng (2003)) and the present study both support the influence of pitch variation, intensity, and speech rate on human judgments of the personality of "energetic", the present study does not suggest a clear relationship between changes in intensity range and their impact on laziness. In addition, Hector's recordings do not demonstrate the impact of fundamental frequency on human perceptions of laziness. Nonetheless, the

average of all four audios in Hector reveals a similar pattern to previous research ($F0_{Chinese}$: 248.75Hz; $F0_{English}$: 193.26Hz). Both investigations suggest that higher fundamental frequency (F0) in the audio is associated with more frequent perceptions of energy.

In Table 4.1, recordings 5a and 5b seem to be a counter-example in which the pitch variation, intensity, and speech rate go contrary to our expected trend. Therefore, with a higher rate of all pitch variation, maxim intensity, and speech rate in group 5 English recording, English is perceived as more "energetic" as English receives 19 votes and Chinese receives 17 votes for energetic.

4.2 English dubbings are more "old"



[Figure 4.2 Human judgments of age based on Chinese and English dubbings]

Based on Figure 4.2, English dubbings are rated as "older" than those in Chinese. While English dubbings received more votes than Chinese dubbings in categories 3 ("old") and 4 ("very old"), Chinese dubbings received more votes than English dubbings in categories 1 ("very young") and 2 ("young").

Group (Language)	Max F0 (Hz)	Pitch Variation	Max Intensity	Intensity Range	Jitterness
7a (Chinese)	218.62	303.52	84.21	39.15	2.28%
7b (English)	100.99	422.52	82.01	29.48	3.39%
12b (Chinese)	80.76	323.15	83.55	26.52	3.41%
12a (English)	74.49	430.50	86.25	34.07	2.80%
18b (Chinese)	76.13	333.95	82.61	39.38	2.19%
18a (English)	87.72	272.34	86.31	39.23	3.39%
21a (Chinese)	77.35	327.11	81.90	4.56	3.06%
21b (English)	75.02	237.45	85.56	42.27	3.17%

[Table 4.2 The Praat measurement for Judy]

The present study focuses on Judy's acoustic features and their correlation with human preference in age. The results for the Praat measurement are in Table 4.2. The findings are mentioned below:

(1) For fundamental frequency (F0), while aiming to convey the dubbing in an older tone, the F0 data tends to be lower.

(2) The present study suggests that greater pitch variation may refer to an older age.

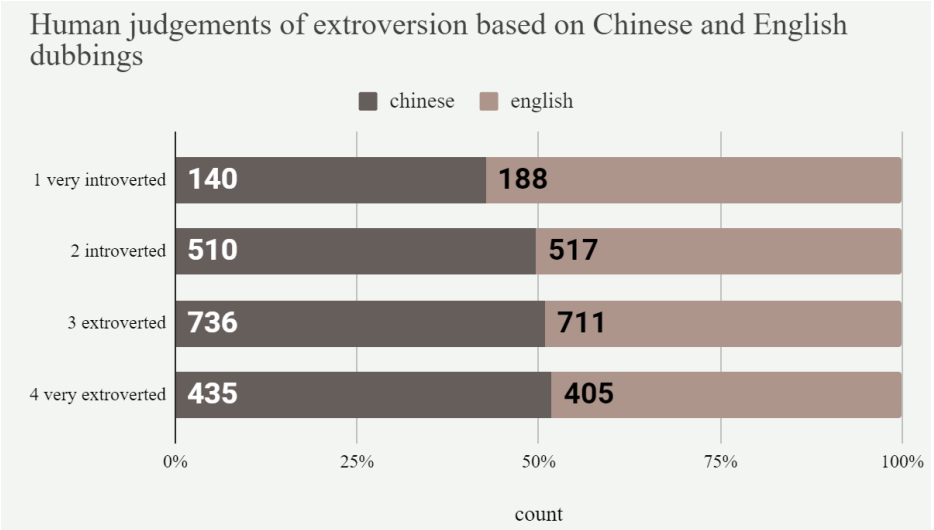
(3) The intensity of the recording also plays a role in the audience's perceptions of age. To exemplify, dubbings with lower maximum intensity values are more often recognized as "old", demonstrating English dubbings to have relatively lower intensity values.

(4) With regards to intensity range, audio recordings exhibiting a greater range of intensity are often perceived as "older", thus indicating English dubbings to have a greater intensity range than Chinese dubbings.

(5) For the jitterness of the dubbing, in order to express an "older" tone, the speakers tend to have a higher rate of jitterness.

By comparing the observations above with the findings in earlier studies, an opposite trend was discovered. The observation regarding pitch variation corresponds to the previous research (i.e. Liu et al (2010)) that indicates that a greater pitch variation may refer to an older age. Further, the present study also investigates that jitterness has also influenced human judgments in age.

4.3 Chinese dubbings are more "extroverted"



[Figure 4.3 Human judgments of extroversion based on Chinese and English dubbings]

Based on Figure 4.3, Chinese dubbings are rated as "more extroverted" than those in English. While Chinese dubbings received more votes than English dubbings in categories 3 ("extroverted") and 4 ("very extroverted"), English dubbings received more votes than Chinese dubbings in categories 1 ("very introverted") and 2 ("introverted").

Group (Language)	Pitch variation	Max Intensity	Speech Rate (w/s)	Jitterness
9a (Chinese)	374.72	84.99	3.37	1.99%
9b (English)	277.64	70.32	2.94	2.57%
6b (Chinese)	363.20	85.89	4.54	2.82%
6a (English)	437.17	81.92	3.84	3.36%
13a (Chinese)	321.46	84.77	4.55	1.64%
13b (English)	436.25	79.38	4.40	1.97%
22b (Chinese)	447.31	86.06	2.74	2.96%
22a (English)	312.44	78.40	2.29	3.24%

[Table 4.3 The Praat measurement for Miguel]

The present study of human judgments of extroversion provided the Praat measurement for the typical example of Miguel. The results for the Praat measurement are mentioned in Table 4.3, while the present study has the following observations:

(1) The pitch variation of the dubbing plays a role in the audience's judgments of extroversion. As the present study discovered, dubbings with higher pitch variation often received more votes for "extroverted". This unveils that with a higher pitch

variation, Chinese dubbings are marked as more “extroverted” than English dubbings.

(2) With regards to the intensity average, the higher intensity average the audio recordings are measured, the more frequently the audio is marked as “extroverted”.

(3) While aiming to convey the dubbing in a more extroverted tone, the speech rate tends to be faster whilst the jitterness tends to be lower.

The above investigations are associated with the findings from previous studies. While the previous research (i.e. Park et al (2020)) and the present study both support the influence of pitch variation, intensity, and speech rate on human judgments of the personality of “extroverted”, the present study discovers another factor. As the present study highlights, the lower the jitterness of the recording, the more often the recording is recognized as “extroverted”. According to Backstrom (2020), a lower jitter yields a higher frequency and intensity. Therefore, subjects marked as “extroverted” have higher intensity averages and lower jitterness, which suggests the correlation between jitterness and intensity that both affect human perceptions of “extroversion”.

5. Conclusion

The present study investigates the difference in human perceptions of English and Mandarin dubbing, along with its correlation with their acoustic features. The results are as follows: (1) Mandarin dubbings are often rated as “more energetic”, influenced by pitch variation, intensity, and speech rate (2) English dubbings are perceived as “older”, affected by F0, pitch variation, intensity, energy variation, and jitterness (3) Mandarin dubbings are marked as “more extroverted”, resulting from pitch variation, intensity, speech rate, and jitterness. The present study not only supports the observations from previous research but also identifies the role of jitterness in human perception of “extroversion” and “age”.

However, there are some limitations in the experiment. The age groups are targeted under the range of 12-15 without considering the perceptions of the elders or younger ones. Individuals with an older age have more experience in life and thereby hold different opinions compared to youngsters. Furthermore, there are no obvious differences in human perception of “self-doubting” and “confidence”, hence allowing us to extrapolate other factors not discussed in the study influencing the result. These probable factors should be taken into consideration for future studies.

References

Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. Journal of Personality and Social Psychology, 37(5), 715–727. https://doi.org/10.1037/0022-3514.37.5.715

Bäckström, T. (2022, January 27). *Introduction to speech processing. Jitter and Shimmer.* Aalto University Wiki. https://speechprocessingbook.aalto.fi/index.html

Brown, B. L., Strong, W. J., & Rencher, A. C. (1972). *Manipulation of vocal qualities by speech synthesis: A new way to study person perception.* Proceedings of the Annual Convention of the American Psychological Association, 7(Pt. 1), 197–198.

Cheng, C. (2003). *The Relation between Characters and Dubbing Voice in Cartoon: Examples from the Cartoon Televised in Taiwan.* National Digital Libray of Theses and Dissertations in Taiwan. https://hdl.handle.net/11296/dee8hk

Hollien, H., & Shipp, T. (1972, March 1). *Speaking fundamental frequency and chronologic age in males.* ashawire. https://pubs.asha.org/doi/10.1044/jshr.1501.155

Erben Johansson, Niklas & Anikin, Andrey & Carling, Gerd & Holmer, Arthur. (2020). *The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features.* Linguistic Typology. 24. 10.1515/lingty-2020-2034.

Keating, P., & Kuo, G. (2010, September 20). WPP, no. 108: *Comparison of speaking fundamental frequency in English and Mandarin*. eScholarship, University of California. https://escholarship.org/uc/item/1gh6x943

Liu, H., Russo, N. M., & Larson, C. R. (2010). Age-related differences in vocal responses to pitch feedback perturbations: a preliminary study. *The Journal of the Acoustical Society of America*, *127*(2), 1042–1046. https://doi.org/10.1121/1.3273880

Mendoza, E., Valencia, N., Muñoz J., & Trujillo H. (1996). *Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS)*, Journal of Voice, Volume 10, Issue 1, 1996, Pages 59-66, ISSN 0892-1997, https://doi.org/10.1016/S0892-1997(96)80019-1.

Park, J., Lee, S., Brotherton, K., Um, D., & Park, J. (2020, March 23). *Identification of speech characteristics to distinguish human personality of introversive and extroversive male groups*. International journal of environmental research and public health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7143196/

Zhou, W. (n.d.). *A study on change of the aesthetics of timbre of Chinese pop music*. AISALLC. https://webofproceedings.org/proceedings_series/ART2L/AISALLC%202019/AISALLC19022.pdf