Does size matter? An preliminary investigation on the effects of physical size on pitch level in pet-directed speech

Yu-Fai Li, Peggy Mok

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong
yfli@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

Abstract

This study investigated the use of pitch in pet-directed speech (PDS) when the size of the pet differs. Two Cantonese-speaking owners were recruited to talk with three dogs with different sizes. Results demonstrated that pitch level and dog size were inversely related, i.e., a higher pitch was used with a smaller dog, and vice versa. Other factors may also influence the pitch level of PDS, like the attitude towards the animal. This study provided further details on how speech register is adjusted to the characteristics of human and non-human recipients.

Index Terms: speech register, pet-directed speech, Cantonese

1. Introduction

Pet-Directed Speech (PDS) is a speech register used when a speaker interacts with a pet, which is characterized by a higher pitch and affect when compared to the speech directed to an adult [1]. Sometimes it is also known as "doggerel" when the speech is directed to dogs [2], but it can also be directed to other animals like cats [1] and parrots [3].

As PDS is perceptually similar to infant-directed speech (IDS) which is the register used to interact with infants, PDS is often described in the three aspects utilized by IDS, namely, attentional, affective, and didactic [4]. First of all, the raised pitch of PDS is considered as the attentional component, which is measured by the fundamental frequency (F0) and its range [5]. The second aspect of PDS is the affective component, which is usually measured by ratings of low-pass filtered speech so that judgment of raters is based on the intonation and rhythm of the speech but not segmental and semantic information [6]. The last aspect of PDS is the didactic component, which concerns the quality of the articulated vowels in the speech and is measured by the area of a vowel triangle formed by joining the first and second formants (F1, F2) values of [a], [i], and [u] [7].

It has been shown in previous studies that speakers are able to fine-tune their speech to accommodate the characteristics of the recipients, both humans and pets. For example, while PDS and IDS are similar in terms of their high pitch (i.e., the attentional component) and affect (i.e., the affective component), for the didactic component vowels are often hyperarticulated in IDS but not in PDS to dogs and cats, suggesting that speakers are aware of the potential linguistic ability of infants and the lack of such ability in dogs and cats [1]. Similarly, it was found that speech directed to a parrot has more hyperarticulated vowels, a didactic component, when compared to the speech directed to a dog, which can be explained by the fact that parrots show some degree of linguistic ability while dogs show a lack of such ability, and speakers are well aware of this fact [3].

In this study, it is proposed that the physical size of the animal would influence the pitch height of the PDS, so that speech directed to a bigger animal would have a lower pitch and that directed to a smaller animal should have a higher pitch. According to the Frequency Code proposed by Ohala [8], pitch height is related to the size of the animal because bigger animals tend to have a more massive vibrating membrane (vocal cords in mammals and syrinx in birds) so their voices have a lower pitch height, and vice versa for smaller animals as they have a less massive vibrating membrane. Therefore, it is possible that when a speaker interacts with a small dog, he or she would speak with a higher pitch level so that he or she would be perceived as smaller and appeared less aggressive to the small dog, as a result a good human-animal interaction can be ensured. In contrast, a speaker would speak with a lower pitch level when his or her speech is directed to a big dog, so that he or she would be perceived as bigger and appeared more assertive to the big dog, as a result the big dog would be more obedient to the speaker and a harmonic human-animal interaction is facilitated.

The present study aimed at investigating the above notion that the use of pitch in PDS, whether it is high or low, is related to the physical size of the animal. Specifically, it was hypothesized that a speaker would use a higher pitch when talking with a small animal, and a lower pitch when talking with a big animal. By investigating this previously unstudied aspect of PDS, the present study hopes to shed light on how speakers adjust their speech to match with the characteristic of the recipients.

2. Methods

2.1. Participants

Two Cantonese-speaking adults of a family, one female and one male, were recruited to be the human participants of this study. They have over ten years of experience in keeping dogs, and they are currently keeping over 20 dogs with different physical sizes. We chose to observe the naturalistic interaction between dog owners and their dogs, which is analogical to the general practice of observing the interaction between caregivers and their infants in IDS researches. Because of this methodological decision, only two participants were recruited because of the difficulty to find participants who keep multiple dogs with different sizes at the same time.

Three dogs of different size and breeds were selected as the animal participants. They were chosen among the dogs that are currently kept by the human participants, because of their docility and general good health condition. All of them have been neutered. In this study, the notion "size" was defined by

the overall body volume (body length × body width × body height, LWH) of the dogs. Please refer to Table 1 for the information of the three dogs in this study.

Table 1. *Details of the three dogs in the present study.*

Name	Breed	Age	Weight (kg)	LWH (cm)	Overall volume (cm^3)
Piggy	Wire Fox Terrier	10	8	50×37×12	22200
Duke	Beagle	8	11	60×30×20	36000
Fanny	Mongrel	3	18	70×45×22	69300

2.2. Procedures

To enhance naturalness, all interactions between the human participants and the three participating dogs were in the house of the human participants. Each human participant interacted with each dog separately in a room. Each interaction was about five minutes, so that the whole data collection process was about 30 minutes (including time of taking a dog to leave the room, and introducing another dog to the room).

All interactions were recorded with a Zoom H1 handy recorder, attached with a Rode SmartLav lavalier microphone through a TRRS to TRS Adaptor. The recorder is equipped with a low cut filter which was turned on during the whole course of recording so as to minimize noise. Recordings were in WAV format, with 16 bit and 96 kHz sample rate. Input level was tested before actual recording, and it was subsequently set to 50% as the recording environment was quiet.

2.3. Analysis

Coding of the recordings was done using PRAAT (Version 5.0.21). Boundaries that correspond to the onset and offset of each word were annotated. For the purpose of this study, each word was manually identified with reference to the pitch contour generated by the built-in function of PRAAT, with the default settings of Hertz range (75 to 500 Hz) and pitch analysis method (autocorrelation) applied. Labels reflecting the lexical tone carried by the word were used in the annotation. Utterances directed to humans were not the focus of this study so they were not annotated. In addition, fused syllables were not annotated as they carried a fusion lexical tone. Also, words that have a boundary tone (i.e., the rising contour at the end of a question) were not annotated. Unclear utterances were also not annotated.

A PRAAT script was ran to determine the pitch of each annotated word. The script was designed to measure each word at five equidistant time points, respectively located at the beginning of the word (referred as p1 in subsequent section), the 25th percentile of the total duration of the word (p2), the middle of the word (p3), the 75th percentile of the total duration of the word (p4), and the end of the word (p5). An average value of F0 was taken across these five time points for words that carried the lexical tones 1, 3 and 6, since they are three level tones. On the other hand, words that carried the lexical tones 2, 4 and 5 were compared directly at these five time points, since they are contour tones.

3. Results

3.1. Level tones

For the female participant, she produced 90, 56, and 115 usable Tone 1 (a high-level tone) tokens to the three participating dogs Piggy, Duke, and Fanny respectively. A one-way between-subject Analysis of Variance (ANOVA) revealed that the average F0 of Tone 1 tokens directed to the dogs was significantly different, F (2, 258) = 32.32, p = .00, ω^2 = .19. Post-hoc comparisons with the Bonferroni test shown that the average F0 of Tone 1 tokens directed to Piggy (M = 352.03, SD = 59.35) was significantly higher than that directed to Duke (M = 311.95, SD = 57.61), which was significantly higher than that directed to Fanny (M = 289.42, SD = 51.14).

For the mid-level Tone 3, the female participant produced 49, 38, and 98 usable tokens directed to Piggy, Duke, and Fanny respectively. A one-way between-subject ANOVA shown that the average F0 of Tone 3 tokens directed to different dogs was again significantly different, F (2, 182) = 36.51, p = .00, ω^2 = .28. Bonferroni test demonstrated that the average F0 of Tone 3 tokens directed to Piggy (M = 291.49, SD = 42.85) was significantly higher than that directed to Duke (M = 263.92, SD = 32.29), which in turn was significantly higher than that directed to Fanny (M = 233.25, SD = 40.49).

For the low-level Tone 6, the female participant produced 20, 4, and 42 usable tokens to Piggy, Duke, and Fanny respectively. A one-way between-subject ANOVA indicated that the average F0 of Tone 6 tokens directed to different dogs was also significantly different, F (2, 63) = 14.68, p = .00, ω^2 = .29. Bonferroni test shown that the average F0 of Tone 6 tokens directed to Piggy (M = 271.63, SD = 42.51) was significantly higher than that directed to Fanny (M = 219.56, SD = 32.18), but not Duke (M = 233.76, SD = 26.17). The average F0 of Tone 6 tokens directed to Duke was not significantly different from that directed to Fanny, possibly due to the lack of data (only 4 tokens for Duke).

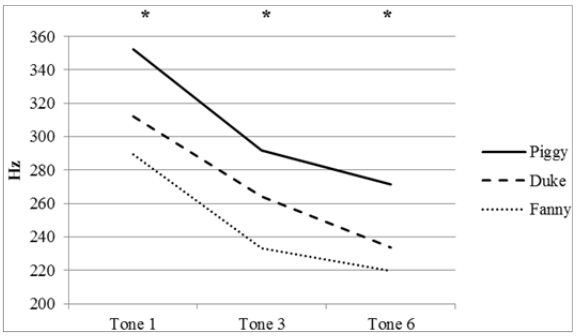

Figure 1: *The average F0 of Tone 1, 3, and 6 tokens directed to the three dogs by the female human participant. An asterisk on the top indicated a significant difference between F0 values of different dogs for a particular tone.*

For the male participant, he produced a total of 305 usable tokens that carried the high-level Tone 1, in which 99 tokens, 130 tokens, and 76 tokens were directed to Piggy, Duke, and Fanny respectively. As the homogeneity of variance assumption was not supported by the Levene's test (p < .001), the Welch correction was applied. A one-way between-subject ANOVA revealed that the average F0 of Tone 1 tokens

directed to different dogs was significantly different, *Welch's* F (2, 183.75) = 9.27, p = .00, *est.* ω^2 = .05. Post-hoc comparisons were done using the Games-Howell procedure as homogeneity of variance was not assumed. It was found that the overall average F0 of Tone 1 tokens directed to Piggy (M = 172.32, SD = 30.43) was significantly higher than that directed to Duke (M = 162.33, SD = 23.77) and Fanny (M = 155.59, SD = 20.83). The average F0 of Tone 1 tokens directed to Duke was not significantly different from that directed to Fanny.

For the mid-level Tone 3, the male participant produced 47, 77, and 43 usable tokens which were directed to Piggy, Duke, and Fanny respectively, accounted for a total of 167 tokens. A one-way between-subject ANOVA revealed that the overall average F0 of Tone 3 tokens directed to Piggy (M = 140.61, SD = 18.70), Duke (M = 135.28, SD = 18.96), and Fanny (M = 134.62, SD = 53.95) was not significantly different, *Welch's* F (2, 81.39) = 1.22, p = .30.

A similar result was also obtained for the male participant's production of the low-level Tone 6. In total he produced 151 usable tokens with Tone 6, in which 38 tokens, 66 tokens, and 47 tokens were directed to Piggy, Duke, and Fanny respectively. A one-way between-subject ANOVA indicated that the overall average F0 of Tone 6 tokens directed to Piggy (M = 126.12, SD = 17.67), Duke (M = 119.96, SD = 16.93), and Fanny (M = 122.94, SD = 39.91) was not significantly different, F (2, 148) = 0.67, p = .52.

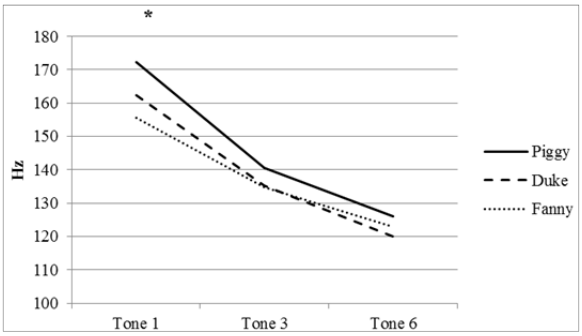

Figure 2: *The average F0 of Tone 1, 3, and 6 tokens directed to the three dogs by the male human participant.*

3.2. Contour tones

For the high-rising Tone 2, the female participant produced 56, 34, and 76 usable tokens directed to Piggy, Duke, and Fanny respectively. Five one-way between-subject ANOVAs were conducted in relation to this tone as well as all subsequent contour tones to compare the F0 of the words directed to different dogs at the five equidistant time points. Results revealed that there was a significant difference among different dogs in terms of the F0 of the words directed to them at four of the five time points, namely p2, p3, p4 and p5 (p2: *Welch's* F (2, 87.37) = 3.73, p = .03, *est.* ω^2 = .03; p3: *Welch's* F (2, 87.64) = 5.29, p = .05, *est.* ω^2 = .29; p4: *Welch's* F (2, 85.48) = 10.49, p = .00, *est.* ω^2 = .10; and p5: *Welch's* F (2, 83.77) = 10.71, p = .00, *est.* ω^2 = .10). Games-Howell test shown that at p2 the F0 of the words directed to Piggy was significantly higher than that directed to Fanny, but not Duke. At p3, p4, and p5, the F0 of the words directed to Piggy was significantly higher than that directed to both Duke and Fanny. Please refer to the upper panel of Figure 3 for the contours of the three contour tones produced by the female participant.

For the low-falling Tone 4, the female participant produced 18, 22, and 46 usable tokens to Piggy, Duke, and Fanny respectively. It was found that that the F0 of the words directed to the dogs was significantly different at all the five time points, p1, F (2, 83) = 8.39, p = .00, ω^2 = .15, p2, F (2, 83) = 8.33, p = .00, ω^2 = .15, p3, F (2, 83) = 7.97, p = .00, ω^2 = .14, p4, F (2, 83) = 6.07, p = .00, ω^2 = .11, and p5, F (2, 83) = 4.60, p = .01, ω^2 = .08. Bonferroni test demonstrated that the F0 of Tone 4 tokens directed to Piggy and Duke was significantly higher than that directed to Fanny at p1, p2, and p3. At p4 and p5, the F0 of Tone 4 tokens directed to Piggy was significantly higher than that directed to Fanny, but that directed to Duke was not significantly different from Piggy and Fanny.

For the low-rising Tone 5, the female participant produced 9, 12, and 81 usable tokens directed to Piggy, Duke, and Fanny respectively. There was a significant difference among different dogs in terms of the F0 of the words directed to them at all the five time points, namely, p1, F (2, 99) = 8.83, p = .00, ω^2 = .13, p2, F (2, 99) = 13.15, p = .00, ω^2 = .19, p3, F (2, 99) = 13.81, p = .00, ω^2 = .20, p4, F (2, 99) = 18.09, p = .00, ω^2 = .25, and p5, F (2, 99) = 17.07, p = .01, ω^2 = .24. Bonferroni test shown that the F0 of Tone 5 tokens directed to Piggy was significantly higher than that directed to Duke and Fanny at all the five time points. On the contrary, the F0 of Tone 5 tokens directed to Duke at all five time points was not significantly different from that directed to Fanny.

For the male participant, he produced 59, 49, and 47 usable tokens of high-rising Tone 2 to Piggy, Duke, and Fanny respectively. There was a significant difference among different dogs in terms of the F0 of the words directed to them at all the five time points, namely, p1, F (2, 152) = 14.87, p = .00, ω^2 = .15, p2, F (2, 152) = 10.08, p = .00, ω^2 = .10, p3, F (2, 152) = 8.39, p = .00, ω^2 = .09, p4, *Welch's* F (2, 101.17) = 11.79, p = .00, *est.* ω^2 = .12, and p5, *Welch's* F (2, 101.33) = 15.88, p = .00, *est.* ω^2 = .16. Bonferroni test (at p1, p2, and p3) as well as Games- Howell test (at p4 and p5) demonstrated that the F0 of Tone 2 tokens directed to Piggy was significantly higher than that directed to Duke and Fanny at all the five time points. On the contrary, the F0 of Tone 2 tokens directed to Duke at all five time points was not significantly different from that directed to Fanny. Please refer to the lower panel of Figure 3 for the contours of the three contour tones produced by the male participant.

For the low-falling Tone 4, the male participant produced 22, 41, and 22 usable tokens directed to Piggy, Duke, and Fanny respectively. It was shown that the F0 of the words directed to the dogs was significantly different only at p1, F (2, 82) = 4.12, p = .02, ω^2 = .07. Bonferroni test shown that the F0 of Tone 4 tokens directed to Piggy was significantly higher than that directed to Fanny, but not Duke at p1. At the other four time points, no significant difference was found in terms of the F0 of Tone 4 tokens directed to different dogs, specifically, p2, F (2, 82) = 1.22, p = .30, p3, F (2, 82) = 0.73, p = .49, p4, F (2, 82) = 0.97, p = .39, and p5, F (2, 82) = 1.20, p = .31.

Finally, for the low-rising Tone 5, the male participant produced 32, 38, and 28 usable tokens directed to Piggy, Duke, and Fanny respectively. It was revealed that the F0 of the words directed to the dogs was not significantly different at all the five time points, specifically, p1, F (2, 95) = 2.42, p = .09, p2, F (2, 95) = 1.78, p = .18, p3, F (2, 95) = 1.58, p = .21, p4, F (2, 95) = 1.65, p = .20, and p5, F (2, 95) = 2.15, p = .12.

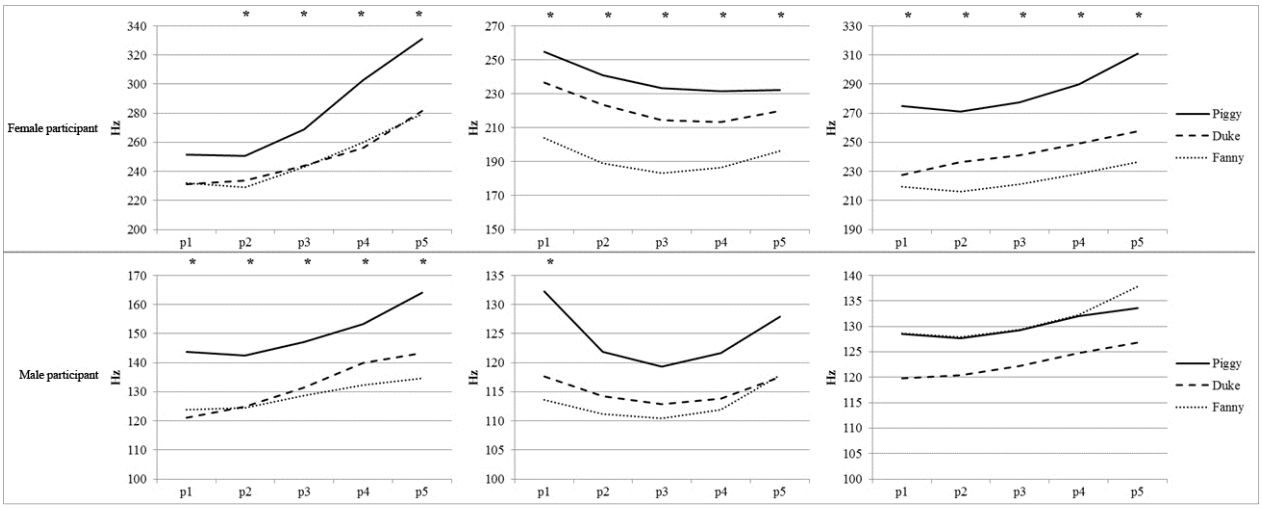

Figure 3: *F0 contours of, from left to right, Tone 2, Tone 4, and Tone 5, produced by, from top to bottom, the female participant and male participant. An asterisk on a time point represents a significant difference between F0 values of different dogs at that time point.*

4. Discussion

In general, the hypothesis of the present study that the pitch of speech directed to an animal is related to the physical size of that animal is largely supported. Specifically, it was hypothesized that the pitch in PDS is inversely related to the physical size of the animal, such that the pitch would be higher when the speech is directed to a smaller animal, and lower when the speech is directed to a bigger animal. As revealed by the data obtained from the female participant, speech directed to Piggy, which was the smallest dog in this study, had the highest pitch when compared to speech directed to dogs with ascending physical size, namely, Duke and Fanny, and vice versa. This pattern was especially clear in her production of words that carry particular lexical tones, like Tone 1 and Tone 3. Admittedly, one would argue that other factors may contribute to this pattern, like the cuteness of the dogs, or the physiological sex of them. However, as the female participant reported that the dogs in this study were equally cute to her, and all of them had been neutered, the present pattern on the relationship between physical size of the dogs and pitch height of speech should be well justified.

However, one point that needs clarification is the inconsistent pattern of pitch use between the female and male participant in the study, a point that probably can be explained by the attitude of different speakers. While the female participant in this study largely followed the hypothesized pattern of pitch use that pitch height and dog size are inversely related, this pattern could only be partially observed from the male participant. In particular, while the pitch level used to talk with the smallest dog Piggy and the biggest dog Fanny was almost always the highest and the lowest respectively, the difference between them did not reach significance in most of the cases. While the small pitch range used by the male participant could be a reason on the lack of significant difference, another possible cause would be his negative attitude towards dogs. During the recording session, it was observed that the male participant was reluctant to interact with the dogs, and further comments from him suggested that he is not really a dog-lover. Therefore in his speech, the lack of significant difference among the pitch level directed to different dogs could be attributed to his lack of interest in dogs, a point that is in line with the claim that a higher pitch is associated with positive emotions [9]. This also implies that the speaker's attitude towards the animal is a possible variable that influences the pitch level of PDS, and future investigation on the topic should pay attention to it.

In addition, the present study only looked at PDS in a simple setting, which imposed a limit on how the findings can be generalized to other situations of speech use. For example, in some instances of human-dog interaction, a dog owner would give orders to his or her dog, or discipline the dog when it is disobedient, and the pattern of pitch use would be different in such situations, and thus it would be interesting to see in these situations whether a difference in pitch level could still be observed. Therefore, future investigation on PDS can also try to test the findings of the present study in different types of human-dog interactions, so that a clearer understanding on the pitch use of PDS can be obtained.

To conclude, the present preliminary study has suggested a probable relationship between physical size of the animals and the pitch used in speech directed to them. While the need to further describe and explain this relationship is warranted with more participants, future study on the same topic should also include the attitude of the owner, as well as different types of human-animal interactions as possible factors influencing the pitch level of PDS. As a final note, as PDS resembles IDS in many similar ways [1] [3], future study can also try to see whether the relationship between the age of infants and pitch height of speech as found in IDS [6] [10] [11] has anything to do with the possible influence of physical size (of infants), which can further elucidate the adjustment of speech register to the characteristics of communicative partners.

5. References

[1] D. Burnham, C. Kitamura, and U. Vollmer-Conna, "What's new, pussycat? On talking to babies and animals," *Science Magazine*, vol. 296, p. 5572, May 2002.

[2] K. Hirsh-Pasek and R. Treiman, R, "Doggerel: Motherese in a new context," *J. Child Language*, vol. 9, pp. 229–237, 1982.

[3] N. Xu, D. Burnham, C. Kitamura, and U. Vollmer-Conna, "Vowel hyperarticulation in Parrot-, Dog- and Infant-Directed

Speech," *Anthrozoos: A Multidisciplinary Journal of the Interactions of People and Animals*, vol. 26, pp. 373–380, 2013.
[4] D. Burnham, S. Joeffry, and L. Rice, "Computer- and Human-directed speech before and after correction," in *Proc. 13th Australasian Int. Conf. Speech Science and Technology*, Melbourne, Dec 2010, pp. 13–17.
[5] A. Fernald and P. Kuhl, "Acoustic determinants of infant preference for motherese speech," *Infant Behavior and Development*, vol. 10, pp. 279–293, 1987.
[6] C. Kitamura and D. Burnham, "Pitch and communicative intent in mother's speech: Adjustment for age and sex in the first year," *Infancy*, vol. 4, pp. 85–110, 2003.
[7] P. K. Kuhl, J. E. Andruski, I. A. Chistovich, L. A. Chistovich, E. V. Kozhevnikova, V. L. Ryskina, … F. Lacerda, "Cross-language analysis of phonetic units in language addressed to infants," *Science Magazine*, vol. 277, pp. 684–686, Aug 1997.
[8] J. J. Ohala, "Cross-language use of pitch: an ethological view," *Phonetica*, vol. 40, pp. 1–18, 1983.
[9] I. R. Murray and J. L. Arnott, "Toward to simulation of emotion in synthetic speech: A review of the literature on human vocal emotion", *JASA*, vol. 93, pp. 1097–1108, 1993.
[10] H.-M. Liu, F.-M. Tsao, and P. K. Kuhl, "Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study," *J. Child Language*, vol. 36, pp. 909–922, 2009.
[11] N. Xu Rattanasone, D. Burnham, and R. G. Reilly, "Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age," *J. Phonetics*, vol. 41, pp. 332–343, 2013.