HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 1 (2022)

# Hearing Dogs and Seeing Barks: Multimodal Sensory Perception of Dogs

Hannah Yu[1] and Dr. Trevor Chen[1#]

[1]Quarry Lane School
[#]Advisor

## ABSTRACT

The field of multisensory integration in human speech perception has been well explored throughout a multitude of studies. Perceptual phenomena and models have been reported, including most notably, the McGurk effect and the Morton-Massaro law of information integration. Existing research of multimodal integration in human perception, include studies that involve perception of drinks and food. Beyond this, scientists have also studied how non-human species, such as dolphins and bats, integrate different sources of information. The purpose of the current study was to test whether multimodal sensory integration is used by humans when perceiving the size of dogs. Three separate audio-visual sources of barking dogs – sized small, medium, and large – were used. A 3 × 3 expanded factorial design was applied to test all possible pairings of auditory and visual combinations from the three sources. The stimuli were included on a questionnaire and then digitally distributed to participants, the majority of whom were high-school students aged 13-17. After viewing each stimulus, participants were prompted to answer what sized dog barked. The results of the current study were consistent with the McGurk effect. When the auditory-visual stimuli were consistent, participants were able to accurately identify what kind of dog barked. However, when the auditory-visual stimuli were inconsistent, the intermediate responses were most frequent. Thus, the results of this study are not consistent with the motor theory of perception. This study yields practical applications in the field of interspecies perception and human-canine communication.

## Introduction

It has been well established that humans integrate auditory and visual sources of information for speech perception (Massaro, 1998). For example, in an experiment using a visual of a face articulating a consonant that was dubbed with a voice articulating a different consonant, MacDonald and McGurk (1978) found that participants perceived another consonant. When the audio speech signal was played independent of the visuals, participants were able to accurately recognize the consonant (Tiippana, 2014). Thus, studies supported the theory that speech perception is not unimodal, and that visuals do affect how humans perceive speech.

Moreover, the Morton-Massaro law of information integration (Movellan & McClelland, 2001), also known as the fuzzy logic model of perception (Massaro, 1998), has been developed as a mathematical model and repeatedly tested. The law outlined three stages in speech perception: evaluation, integration, and decision (Massaro, 1987). Results from this study supported the idea that auditory-visual integration occurs in a variety of speech-perception tasks. These findings were also applicable for different language speakers (Chen & Massaro, 2004; Chen & Massaro, 2008).

Furthermore, speech perception is not the only instance of sensory integration when humans attempt to impose meaning onto their experiences. Morrot, Brochet, and Dubourdieu (2001) explored the connection between the color of wine and how participants perceived the taste of the wine. Participants were asked to describe the aroma of two glasses of wine – one white, and the other red. After a week, the experiment was repeated with the same group of participants, except this time both glasses contained white wine, with one glass of wine dyed red with tasteless food coloring (Spencer, 2010). The descriptors used by participants for the aroma of the dyed white wine were consistent with the descriptors used for the red wine in the first experiment (Morrot et al., 2001). A similar phenomenon explored the integration of sight, smell, and taste when perceiving the sensory qualities of wine. Wang and Spence (2019) studied the influence of wine color on the perceived aroma and smell by having wine novices and experts taste three types of wine: white wine, rosé wine, and white wine that was dyed a rosé color. The results from this study also supported the conclusions from Morrot et al. (2001) – that humans integrate multiple sources of information when perceiving the smell and taste of wine.

In addition to drinks, this principle of multimodal sensory integration has also been found in tests that involved food. For example, Van Stokkom, Blok, Van Kooten, de Graaf, and Stieger (2018) tested how the senses of smell, taste, and texture were related to the perception of vegetables that are commonly consumed in the Netherlands (including broccoli, cauliflower, bell pepper, carrot, cucumber, lettuce, onion, and tomato). It was found that taste and olfactory information increased the identification of vegetables as compared to taste-only and smell-only conditions (Van Stokkom, Blok, Van Kooten, de Graaf, & Stieger, 2018).

Other research studies have also shown that multimodal sensory integration is not unique to humans. For example, information from sound (i.e., echolocation) and vision are integrated for perception by dolphins (Kuczaj, Solangi, Hoffland, & Romagnoli, 2008) and by bats (Kugler, Luksch, Peremans, Vanderelst, Wiegrebe, & Firzlaff, 2019). Given the ubiquity of multisensory integration from mammals in well-practiced tasks, the aforementioned research findings inspired the current study on the interspecies perception between humans and dogs.

Dogs are among the most common pets for humans, so presumably many humans are familiar with the sounds and the faces of dogs. Exploring the auditory-visual perception of dogs is important because it can provide a further test for theories of auditory-visual perception. Therefore, the current experiment tests whether the visual faces and/or the auditory sounds of barking can influence the perception of the size of the dog.

## Methods

### Stimuli

The current study applies a 3 x 3 expanded factorial design because it allows the experiment to test all possible pairings of auditory and visual combinations, similar to those in McGurk-effect conditions. Three audio-visual videos of three different dogs barking were selected from YouTube.

**Figure 1.** Small Dog (NyGirlVic, 2017).

**Figure 2.** Medium Dog (Susan Wedell, 2020).

**Figure 3.** Large Dog (World Class K-9, 2020).

The corresponding videos and sounds were chosen based on the size of the dog and the pitch of the bark. The videos showcased three visually distinct dogs of different sizes — small (Fig. 1), medium (Fig. 2), and large (Fig. 3). The videos were also selected so that the sound of each dog's barking was distinguishable from one another, with the bark of the small dog having the highest pitch, the bark of the large dog having the lowest pitch, and the pitch of the medium dog's bark slightly lower than the pitch of the small dog's bark. Using a video-editing software (iMovie), each video was trimmed so that it only displayed the dog barking once.

A total of 15 stimuli were created, and Table 1 shows the experimental design. Three were visual-only (V), three were audio-only (A), and nine were audio-visual stimuli (AV). To create the three V stimuli, the audio was removed from each of the three videos. To create the three A stimuli, the extracted audio from each video was attached to a black visual. To create the nine AV stimuli, each visual-only video was paired and dubbed with the three bark audios.

**Table 1**. The 3 × 3 Expanded Factorial Experimental Design.

| | **Audio (Bark Sound)** | | | | |
|---|---|---|---|---|---|
| **Visual (Video of Dog Barking)** | | **Small Dog** | **Medium Dog** | **Large Dog** | **No Audio** |
| | **Small Dog** | | | | |
| | **Medium Dog** | | | | |
| | **Large Dog** | | | | |
| | **No Visual** | | | | |

## Questionnaire Design

A digital questionnaire was developed for this experiment. The instructions for participants were displayed in a text-based format on the first page of the questionnaire. The instructions explained that some of the stimuli would be audio-visual, while others would be visual-only or audio-only. Before moving onto the next page of the questionnaire, participants were required to check a box to acknowledge that they had carefully read the instructions. The first page of the questionnaire also included a diagram categorizing nine dog silhouettes into the sizes: Small, Medium, and Large (Figure 4).

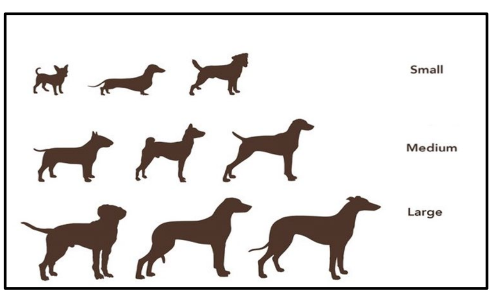


**Figure 4.** Size Chart (Kona Cave, n.d.).

To prevent the order of the stimuli from becoming a confounding variable, the order was randomized for each participant. Each stimulus appeared on a separate page and was followed by the question, "What kind of dog barked?". Participants were required to select one of the following answers: *Small*, *Medium*, or *Large*. Participants were not allowed to move between pages of the questionnaire to prevent them from changing previous answers, lest they figure out the design of the study.

At the end of the questionnaire, three additional questions were included to collect information about participants' demographics and their familiarity with dogs. For demographics, participants indicated their age range and gender. Then, participants ranked their own familiarity with dogs on a scale from 1-5, with 1 indicating "No Experience" and 5 indicating "Very Experienced." Overall, this method is designed to address the research question of how visual and auditory sources of information can affect the perception of dog barks.

### Participants

The majority of participants were high-school students in Advanced Placement courses of 13-17 years of age. These participants accounted for 75.5% of the study group. The questionnaire was also digitally distributed to young adults and adults, with the age range spanning from 21-59 years old.

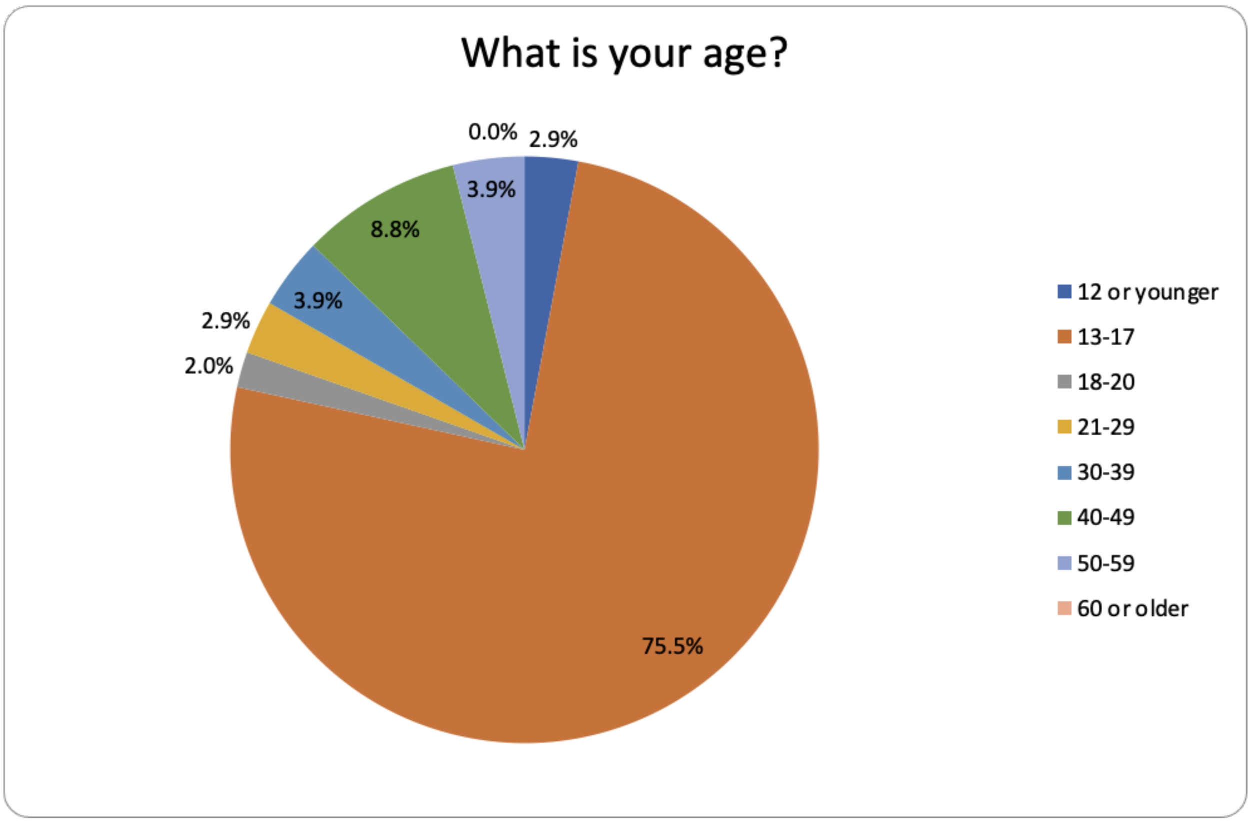


**Figure 5.** Distribution of Participants' Ages.

## Results

Table 2 presents the overall results. For the Auditory-only (A) conditions, A-small stimuli produced mostly small responses, A-medium stimuli were often confused with small responses, and A-large stimuli produced mostly large responses. For the Visual-only (V) conditions, V-small stimuli produced mostly small responses, but V-medium and V-large stimuli were confused with each other.

**Table 2**. Participant Responses of 3 × 3 Expanded Factorial Design Stimuli (sm = small; med = medium; lg = large).

| Visual Stimulus | | Auditory Stimulus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Small | | | Medium | | | Large | | | None | | |
| Small | sm | 88.2% | 90 | sm | 60.8% | 62 | sm | 31.4% | 32 | sm | 78.4% | 80 | |
| | med | 7.8% | 8 | med | 35.3% | 36 | med | 39.2% | 40 | med | 17.6% | 18 | |
| | lg | 3.9% | 4 | lg | 3.9% | 4 | lg | 29.4% | 30 | lg | 3.9% | 4 | |
| Medium | sm | 40.2% | 41 | sm | 21.6% | 22 | sm | 8.8% | 9 | sm | 12.7% | 13 | |
| | med | 40.2% | 41 | med | 55.9% | 57 | med | 32.4% | 33 | med | 50.0% | 51 | |
| | lg | 19.6% | 20 | lg | 22.5% | 23 | lg | 58.8% | 60 | lg | 37.3% | 38 | |
| Large | sm | 50.0% | 51 | sm | 19.6% | 20 | sm | 9.8% | 10 | sm | 15.7% | 16 | |
| | med | 39.2% | 40 | med | 58.8% | 60 | med | 38.2% | 39 | med | 52.9% | 54 | |
| | lg | 10.8% | 11 | lg | 21.6% | 22 | lg | 52.0% | 53 | lg | 31.4% | 32 | |
| None | sm | 87.3% | 89 | sm | 62.7% | 64 | sm | 12.7% | 13 | | | | |
| | med | 10.8% | 11 | med | 26.5% | 27 | med | 38.2% | 39 | | | | |
| | lg | 2.0% | 2 | lg | 10.8% | 11 | lg | 49.0% | 50 | | | | |

For the Auditory-Visual (AV) conditions, when AV stimuli were consistent with each other, the consistent responses were always the most frequent. However, when the AV stimuli were inconsistent, there was a different pattern of results. For example, in the A-large + V-small condition, the medium responses were the most frequent, analogous to a McGurk-like effect for the perception of dogs. For the combined A-small + V-medium condition and the A-small + V-large condition, medium and large responses were as frequent as small responses, showing a McGurk-like influence of the visual information. Figure 6 shows one condition of this McGurk-like effect for the perception of dogs. When the V-small stimulus was paired with a A-large stimulus, the responses were intermediate and medium – compromises between the large and small response.

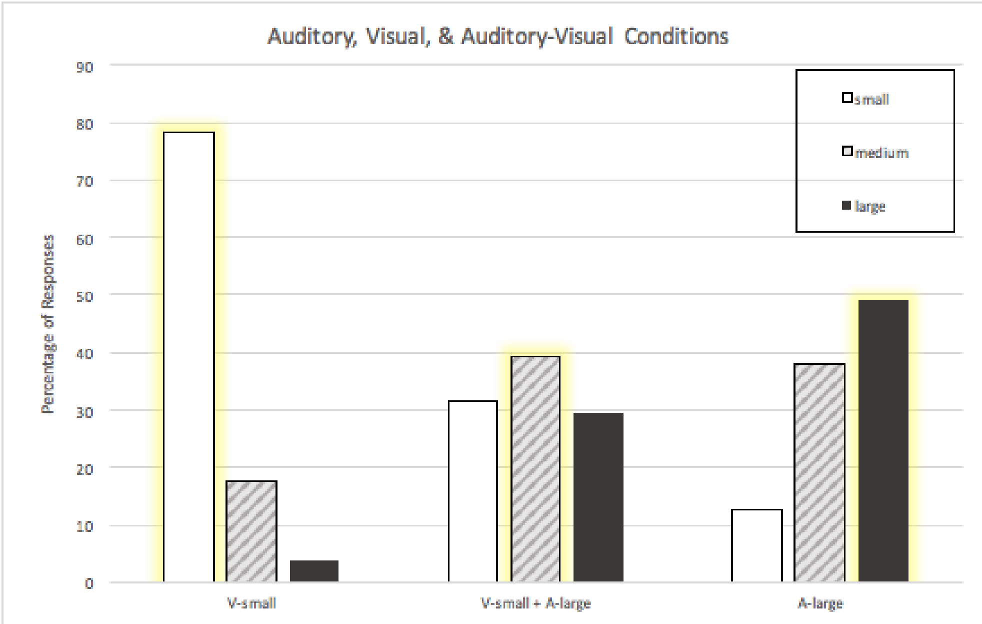


**Figure 6**. One example of the McGurk-like effect for the auditory-visual perception of dogs.

These AV small-large inconsistent conditions were especially interesting, so correlations were carried out to assess whether the number of medium/intermediate (McGurk) responses were related to participants' familiarity with dogs. Figure 7 shows the probability of medium responses for the A-large + V-small condition (grouped by the level of familiarity with dogs), and Figure 8 shows the probability of medium responses A-small + V-large conditions (grouped by the level of familiarity with dogs. In both cases, medium/intermediate responses were the most frequent when participants were somewhat familiar (i.e., familiarity ratings of 3 and 4 out of 5) with dogs. For these participants, the A-large + V-small condition produced the most medium responses, $X^2$ (2, 37) = 5.41, $p < 0.05$. Also, the A-small + V-large condition produced significantly more medium responses as well, $X^2$ (2, 37) = 10.76, $p < 0.05$.

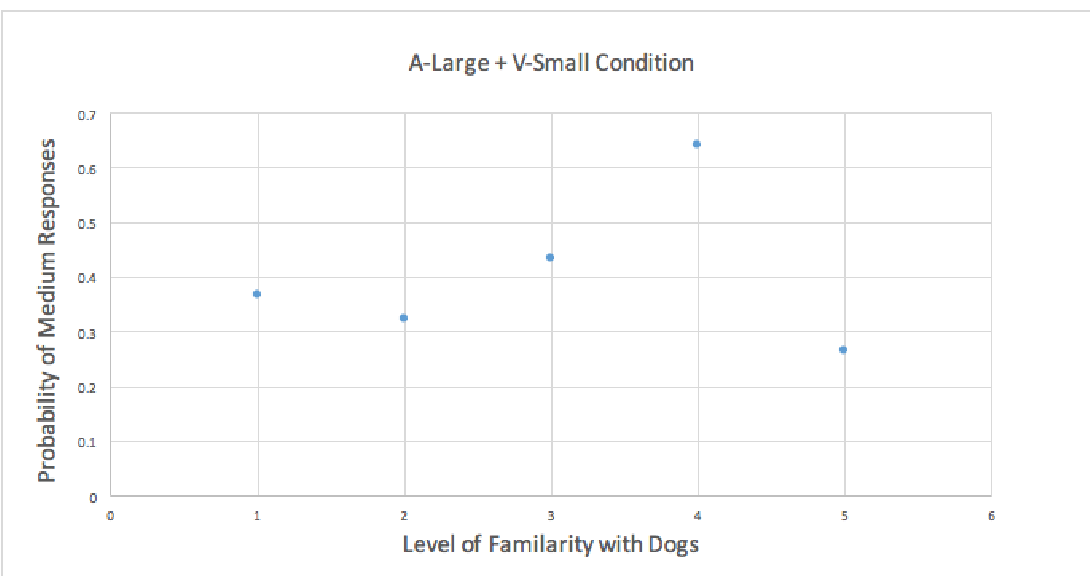


**Figure 7**. The probability of medium responses for the A-large + V-small condition, grouped by the level of familiarity with dogs.

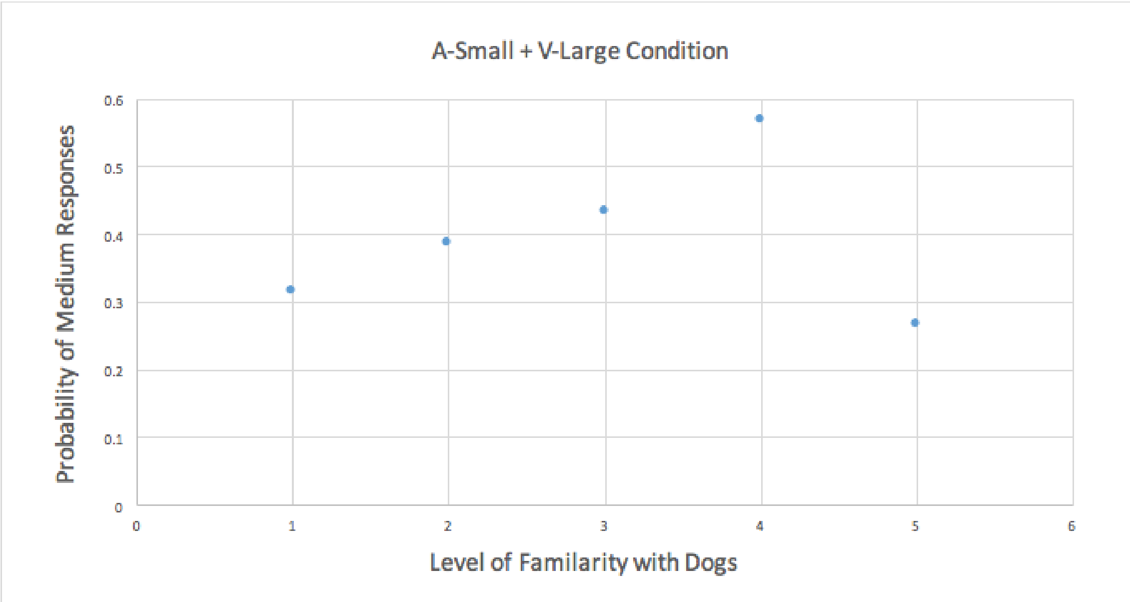


**Figure 8.** The probability of medium responses A-small + V-large conditions, grouped by the level of familiarity with dogs.

# Discussion

The current study found results analogous to a McGurk-like effect for the perception of dogs. For the combined A-small + V-medium condition and the A-small + V-large condition, medium and large responses were as frequent as small responses. This seems like an interesting pattern found in speech perception, specifically the McGurk effect, in which the combination of auditory /ba/ and visual /ga/ can lead to the perception of /da/. Clearly, in the current experiment, both the visual information and the auditory information affected the perception of dog sizes. However, the effects of visual information were most prominent for those who were somewhat familiar with dogs, illustrating the role experience analogous to that of speech perception.

It is possible that those participants who were extremely familiar with dogs knew more about which sounds were likely to come from which dogs, so they might have relied more on auditory information. On the other hand, it is possible that those unfamiliar with dogs did not have enough experiences to associate certain dog sizes with certain visual information, so they might have relied less on their sense of sight. Nevertheless, although there were different degrees of visual influence, all of the participants' responses were still influenced by the integration of visual information.

## Theoretical Implications

The current results are significant in terms of theoretical implications. Even when humans perceive dog barking, the face of the dog still influences perception. In regards to Occam's razor and the principle of parsimony, the current findings are consistent with the Morton-Massaro Law but inconsistent with the motor theory (Massaro & Chen, 2008). After all, patterns of responses analogous to the McGurk-effect were produced in the AV-inconsistent conditions, even though there seems to be no obvious reason to assume that the average human knows much about the articulators of dogs. Therefore, the results of this study support the idea that speech is not special, and the McGurk effect should not be used as evidence for the motor theory of perception (Massaro & Chen, 2008).

### Limitations

Given that the questionnaire only provided 15 stimuli, it is possible that some participants figured out the design of the study, leading them to base their decisions solely on the visual or auditory information. However, it is possible that these participants still could have been integrating both sources of information, even if they did not think they were consciously doing so. The randomized nature of the questions may have also accidentally revealed the design of the study to some participants. For example, one individual mentioned that two of the stimuli that she experienced back-to-back in the experiment were likely the same sound. Although this individual was extremely familiar with dogs, it would be better for future studies to apply pseudo-randomization – i.e., randomization but making sure that no two stimuli with the same sound appear one after the other.

Another drawback of the questionnaire design was that the instructions were text-based. The digital format of the questionnaire and the text-based instructions for the participants increased the probability that some participants may have neglected to read the instructions at the start of the questionnaire. The instructions clearly stated that some of the stimuli videos were audio only, some were visual only, and some of them contained both visuals and audio. If participants did not carefully comprehend the instructions, as they moved throughout the questionnaire they may have been confused if they watched a video that only contained one source of information (audio or visual). This confusion may have affected how these participants responded to the unimodal videos. If participants believed that these videos were faulty or "broken," this may have led them to select a random answer.

Finally, it would be more informative if the current study allowed each participant to provide multiple responses (repeated measures) for all of the auditory and visual conditions. This adjustment would not only make the results more reliable and increase statistical power, but it would also enable additional tests to further assess the response patterns of the participants.

## Conclusion

The findings of this study can be applied in a practical way to contribute to a better understanding of human-canine communication. The study also has implications for a wide range of people – dog owners, dog trainers, and laypeople who are not familiar with dogs. An aspect of this study that can be explored in future research is examining whether the wording of the questions on the questionnaire affects how participants answer. Other possible future research directions include studying the perception of dogs' body language as well as dogs' emotional states, with the potential goal of creating an interspecies "vocabulary" or "language system" that can help humans understand other organisms. In general, findings on interspecies perception and communication can inform people working with non-human animals.

In closing, the area of multimodal interspecies perception holds great promise, especially for humans and canines, who have lived cooperatively for a long time. Across thousands of years of history, the paths of dogs and humans have become increasingly intertwined (Katz, 2019). Dogs have gradually assumed a greater presence in our societies. From serving as war dogs in World War II (Pearson, 2013), to assisting humans as modern search-and-rescue dogs (Jones, Dashfield, Downend, & Otto, 2004), to acting as loyal companions, dogs hold invaluable roles in everyday life. Dogs have observed, sniffed at, and listened to humans for at least tens of thousands of years; it is time that the favor is returned so that their voices can be heard – and be seen as well.

## References

Chen, T. H., & Massaro, D. W. (2004). Mandarin speech perception by ear and eye follows a universal principle. *Perception & Psychophysics, 66*(5), 820-836. https://doi.org/10.3758/BF03194976.

Chen, T. H., & Massaro, D. W. (2008). Seeing pitch: Visual information for lexical tones of Mandarin-Chinese. *The Journal of the Acoustical Society of America 123*(4), 2356-2366. https://doi.org/10.1121/1.2839004.

Jones, K. E., Dashfield, K., Downend, A. B., & Otto, C. M. (2004). Search-and-rescue dogs: an overview for veterinarians. *Journal of the American Veterinary Medical Association*, 225(6), 854-860. https://doi.org/10.2460/javma.2004.225.854.

Katz, B. (2019, June 17). *When did we come together?* Learning from Dogs. https://learningfromdogs.com/2019/06/17/when-did-we-come-together/.

Kona Cave. (n.d.). Breed Size Chart. In *Kona Cave*. https://cdn.shopify.com/s/files/1/1069/9082/files/Breed_Size_Chart_KC_Travel_Beds_48d81154-2d04-49f2-8054-8be5bd687540_1024x1024.jpg?v=1543928989.

Kuczaj, S., Solangi, M., Hoffland, T., & Romagnoli, M. (2008). Recognition and Discrimination of Human Actions Across the Senses of Echolocation and Vision in the Bottlenose Dolphin: Evidence for Dolphin Cross-modal Integration of Dynamic Information. *International Journal of Comparative Psychology*, 84-95. https://escholarship.org/uc/item/8805892k.

Kugler, K., Luksch, H., Peremans, H., Vanderelst, D., Wiegrebe, L., & Firzlaff, U. (2019). Optic and echo-acoustic flow interact in bats. *Journal of Experimental Biology*, 222(6), https://doi.org/10.1242/jeb.195404.

Macdonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253–257. https://doi.org/10.3758/bf03206096.

Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press. https://doi.org/10.1006/jpho.2000.0108.

Massaro, D. W. (1987). *Speech Perception By Ear and Eye: A Paradigm for Psychological Inquiry* (1st ed.). Psychology Press. https://doi.org/10.4324/9781315808253.

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review, 15*, 453–457. https://doi.org/10.3758/PBR.15.2.453.

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746–748. https://doi.org/10.1038/264746a0.

Morrot, G., Brochet, F., & Dubourdieu, D. (2001). The Color of Odors. *Brain and Language*, *79*(2), 309–320. https://doi.org/10.1006/brln.2001.2493.

Movellan, J. R., & McClelland, J. L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review, 108*(1), 113–148. https://doi.org/10.1037/0033-295X.108.1.113.

NyGirlVic. (2017, February 27). *Cutest face howling 6-months old border terrier puppy*. YouTube. https://www.youtube.com/watch?v=eRmBg7u82Dg.

Pearson, C. (2013). Dogs, history, and agency. History and Theory, 52(4), 128-145. https://doi.org/10.1111/hith.10683.

Spence, C. (2010, November). *The multisensory perception of flavour | The Psychologist*. https://thepsychologist.bps.org.uk/volume-23/edition-9/multisensory-perception-flavour.

Susan Wedell. (2020, April 21). *Red Heeler Cattle Dog crazy bark !* [Video]. YouTube. https://www.youtube.com/watch?v=1pz_EDeThvA.

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00725.

Van Stokkom, V. L., Blok, A. E., Van Kooten, O., de Graaf, C., & Stieger, M. (2018). The role of smell, taste, flavour and texture cues in the identification of vegetables. *Appetite*, 121, 69-76. https://doi.org/10.1016/j.appet.2017.10.039.

World Class K-9. (2020, January 23). *Black Russian Terrier Protection*. YouTube. https://www.youtube.com/watch?v=DgtAs0YFuzU.