IS AN AUDITORY EVENT MORE TAKETE?

Federico FONTANA (federico.fontana@uniud.it) (0000-0002-1692-2603)[1],
Hanna JÄRVELÄINEN (hanna.jarvelainen@zhdk.ch) (0000-0001-6255-8657)[2], and **Maurizio FAVARO**[1]

[1]*Department of Mathematics, Computer Science and Physics (DMIF)*, **University of Udine**, Udine, Italy
[2]*Institute for Computer Music and Sound Technology (ICST)*, **Zurich University of the Arts**, Zürich, Switzerland

ABSTRACT

Recent experiments have demonstrated that the words Takete and Maluma, as well as Kiki and Bouba, once heard stimulate a cross-modal response in humans that goes beyond visual associations, and in particular affects the trajectory of human motion patterns. Inspired by such experiments, in a binary (Takete/Maluma) response test we presented to sixteen individuals a random sequence of either sonic or silent videos reproducing a smooth and a notched ball rolling down along a rounded or zig-zagged path. Bayesian estimation revealed a credible effect of the zig-zagged path in participants choosing Takete, and an equally strong effect of the notched ball. On the other hand, the silent videos had a negative effect on subjects' probability of choosing Takete. This means that in absence of auditory feedback, subjects tend to choose Maluma compared to similar situations with sound. Though exploratory, such a result suggests that the auditory modality may have significantly biased the decision toward Takete when our participants were exposed to the audio-visual event. If supported by more extensive tests, this experiment would emphasize the importance of sound in the cognition of audio-visual events eliciting sense of sharpness in humans.

1. INTRODUCTION

In 1929, Wolfgang Köhler asked a group of Spanish speakers to make an association between the words Takete or Maluma and the images of two shapes, one jagged and the other rounded, like those in Figure 1.

His results showed a significant preference of the speakers for associating Takete with the jagged, and Maluma with the rounded shape. The experiment has been repeated by several psychologists using different pairs of words, in particular Kiki and Bouba, as well as involving speakers from different languages and levels of literacy. Apart from some specific exceptions reported for a population of Papua New Guinea, these experiments have all shown a general tendency of speakers, including young children aged 2.5 years old [1], to map rounded shapes on words containing the vowels "o" and "u", and, conversely, jagged shapes on words containing "e" and "i".

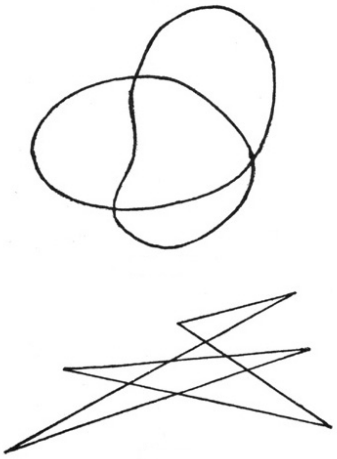

Figure 1. Images similar to those used by Köhler in his experiment.

Taken together, these results provide evidence of a powerful cross-modal effect, linking visual shapes to sounds of words. The presence of this effect in pre-literate children suggests the existence of active connections among contiguous cortical areas, making possible for humans to link characteristic geometrical shape contours to similar geometries assumed by the speaker's lips. According to Ramachandran and Hubbard [2] such connections exist before language, hence they represent a general invariant speeding up and constraining its development.

This research embraced other sensory modalities in more recent decades, investigating associations that are not dominated by vision. Spence and colleagues investigated effects of taste [3, 4]: by asking subjects to associate food and liquids to Kiki/Bouba, they concluded that counter-intuitive branding and packaging may be detrimental to the success of a food product. Similar effects were found for odors [5]. A stronger input to our work, however, comes from experiments that involved motion patterns. Independently of each other, in 2016 Shinohara *et al.* and Koppensteiner *et al.* presented an experiment linking gestures to Takete/Maluma by using animated human figures [6, 7]. Earlier in 2013, Fontana had experimented on the same link by guiding the dominant hand of blindfolded partici-

pants along rounded or jagged trajectories by means of a robotic arm, hence excluding vision completely from the tests [8]. Together, these experiments shed further light on the human ability to associate words to sensations involving motion. Further considerations about this ability and its relationships with previously memorized mental imageries were pointed out by Fryer *et al.* while testing haptic-word associations made by blind individuals [9], and then rediscussed by Graven & Desebrock [10].

Sensation of motion becomes unavoidable if an experiment is designed involving auditory stimuli. Familiar sounds in fact are almost inevitably linked to dynamic events, in which motion is inherently implied. However, the association between words and auditory stimuli is a fragile concept by definition. As words encode sounds, an experiment of this kind should first provide evidence that the association between an auditory stimulus and the sound of a word is not merely onomatopoeic. Probably due to this issue, that puts the own concept of *association* under discussion, experiments linking auditory feedback to words are apparently absent in the literature. However, two ideas convinced us to proceed along this uneven path:

- if the sound of our words of interest is a consequence of onomatopoeia [2], then auditory stimuli should be chosen among familiar sounds that do not imply the words Takete/Maluma or Kiki/Bouba via an evident onomatopoeic link, as e.g. a tik-tok or mumbling sound would suggest;

- if auditory feedback is able to define a genuine, that is, not onomatopoeic association with such words, then the effect can be controlled by removing sound from a multi-sensory stimulus in which this feedback is superimposed as part of a multi-modal event presentation.

Moved by such ideas, we designed an experiment in which participants had to classify a ball rolling down as Takete or Maluma. Two audio-visual components were present in each stimulus: the ball surface and the path trajectory. The surface was either smooth or notched; the trajectory was either rounded or zig-zagged. In what follows, the smooth/round conditions are marked with M (Maluma), and the notched/zig-zagged conditions are marked with T (Takete) according to their respective hypothesized association. Holding such two visible differences, the corresponding rolling sounds of the two balls and the collision sounds they did against the side walls while traversing the respective paths were different as well.

2. METHOD

With all laboratories at the university being inaccessible to students and guests due to the covid pandemic, the experiment took place in a quiet room at one of the Authors' home.

Figure 2. Balls (above) and paths (below) used in the experiment.

2.1 Participants

Sixteen participants (8 female and 8 male, ages M=40.6, sd=17.6 years), all reporting normal sight and hearing volunteered for the experiment. Two of them reported previous knowlwedge of the Köhler experiment.

2.2 Setup and stimuli

Two balls were made of white play dough covered with vinyl glue (Figure 2, above), both having an external diameter of about 6 cm and a weight of about 250 g. In parallel, two paths were prepared on a plywood base sized $1\times0.5\times0.15$ m (Figure 2, below), again using play dough covered with vinyl glue for the side walls delimiting the paths. Once such paths were refined so as to provide an approximately identical time to reach the bottom, the side walls were secured to the base with permanent glue and the setup was painted. A contrast between dark still and bright moving objects was created, similar to the scenario that Shinohara *et al.* had presented to their participants [6].

The two balls were video- and audio-recorded while they rolled down along both paths, once being left free to roll by one Author who wore a dark cloth. Four short sonic videos hence were recorded, three times each. Each video, then, was duplicated by removing the soundtrack. Twenty-four stimuli, twelve sonic and twelve silent videos, were finally

made available for the tests. [1]

2.3 Procedure

While sitting in front of a PC equipped also with speakers, each participant was asked to attend some videos of what was told to be a simple passtime game popular in Polinesia. Two different versions were told to exist about this game, Takete or Maluma as they were called by locals, and participants had to label each video accordingly when it was finished, by verbally reporting their choice to the experimenter; alternatively they could see it again, by pressing the space bar of the PC keyboard. Once a decision was made, they attended the next video.

The videos were included in a randomly balanced sequence of trials, for a total of 2 balls {T, M} × 2 paths {T, M} × 2 modality {V, AV} × 3 repetitions = 24 trials. Each session lasted about 10 minutes. At the end of it, each participant left comments to the experimenter in particular including information about his or her previous knowledge of this experiment.

3. RESULTS

Results are presented in Figure 3. Inspecting the raw data, it seems that incongruent combinations of ball and path (M-T or T-M) lead to relatively even distributions of Takete and Maluma responses – however, in favor of Takete in audio-video conditions (AV) and in favor of Maluma in the video only conditions (V). In congruent ball-path combinations M-M and T-T, responses are biased toward Maluma and Takete respectively, as expected. Yet again, Takete responses are generally favored in the AV conditions and Maluma in the V conditions, so much so that with T ball and T path, responses based on video only (V) approach random.

Statistical analysis was carried out by logistic regression as explained below. The model coefficients were estimated by Bayesian methods using the R program and the *brms* package [11–13].

The Takete/Maluma response, a binary-outcome dependent variable, was mapped to values $k = 0$ (Maluma response) and $k = 1$ (Takete response). Such an outcome follows the Bernoulli distribution, taking value 1 with unknown probability p and value 0 with probability $1 - p$:

$$f(k;p) = \begin{cases} p & \text{if } k = 1 \\ 1-p & \text{if } k = 0 \end{cases} \quad (1)$$

The unknown probability p of a Takete outcome was predicted by a logistic regression model given by

$$\log \frac{p}{1-p} = \beta_0 + \sum_{i=1}^{m} \beta_i \cdot x_i, \quad (2)$$

where $m = 3$ is the number of predictors, x_i are the predictors (path, ball, and modality; the effect of repetition was not modeled), β_0 is the intercept, and β_i are the regression coefficients estimated by the model.

[1] The videos are available at `https://doi.org/10.5281/zenodo.4770168`.

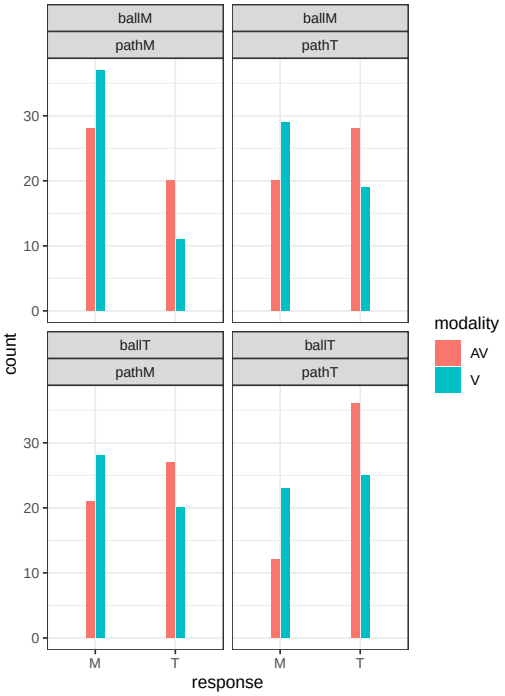

Figure 3. Results. y-axis = response counts (M: Maluma; T: Takete) for both modalities (AV: audio-video; V: video) in each factor combination (ballM: smooth ball; ballT: notched ball; pathM: rounded path; pathT: zig-zagged path).

Figure 4 presents the parameter estimates and their 95% Credible Intervals from the posterior distribution, produced by Markov chain Monte Carlo (MCMC) draws. These logit-transformed values [2] cannot be interpreted in terms of probabilities; however, a 95% CI either entirely above or below zero indicates a credible non-zero positive or negative effect on p, respectively. Hence, Takete path and Takete ball both have an equally strong positive effect. In contrast, video without audio produces credibly more often a Maluma response than video and audio combined. The conditional effects, transformed back to probabilities, are presented in Figure 5.

4. DISCUSSION

This explorative experiment demonstrated that the event of a ball rolling on a surface can be perceived as Takete or Maluma depending on both smoothness of the ball and shape of the trajectory. Our statistical model was additive; modeling the ball-path interaction would require a larger dataset. In a larger experiment, measurement of decision times should also be informative, given that decisions tend to take longer under increasing uncertainty [14, 15]. This could help in investigating, whether either the ball or the

[2] The logit function maps values from $p \in [0, 1]$ to $x \in [-\infty, \infty]$ according to $x = \log(\frac{p}{1-p})$; the inverse mapping is given by the logistic function $p = \frac{1}{1+e^{-x}}$

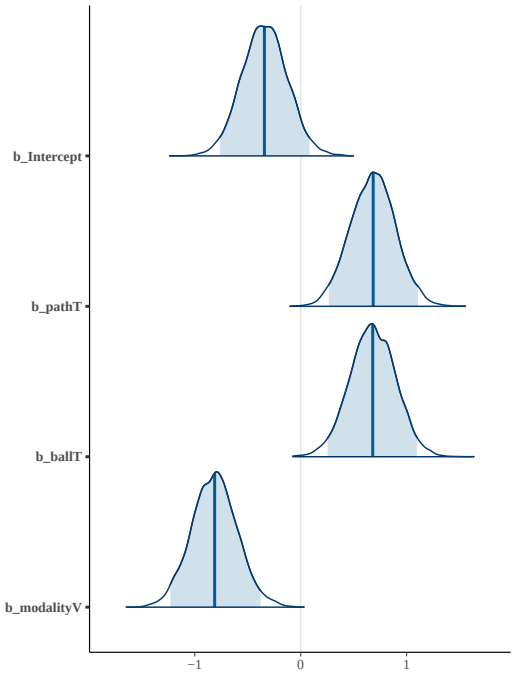

Figure 4. Parameter estimates from the posterior distribution of the Bayesian model.

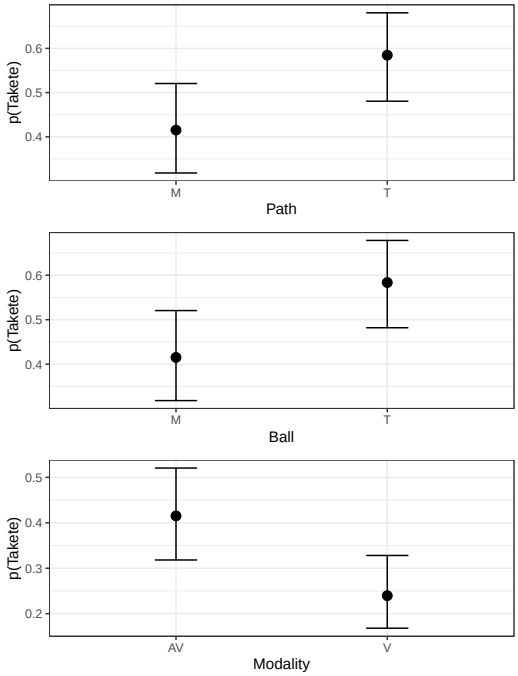

Figure 5. Conditional effects of path, ball, and modality. y-axis = estimated probability of a Takete response; errorbars = 95% Credible Intervals.

trajectory is a dominant feature in the rolling event. In the present model, their effects were approximately equal.

Interestingly, a Takete response was more probable in presence of sound. As the audio and video signals were always congruent, we should expect responses at least in the congruent ball-path conditions (T-T and M-M) to be overwhelmingly in favor of Takete and Maluma, respectively. In both cases however, we see the bias towards Takete when sound is present and towards Maluma in the silent videos. Although our statistical model does not allow very refined conclusions, the raw data suggests that sound is crucial for making non-random decisions in the T-T condition (bottom-right panel in Figure 3). It is of course possible that these specific trajectories or balls happened to produce acoustic cues that were perceived as Takete and visual cues that were perceived as Maluma; using a pseudo-random variety of both might reduce the bias.

Humans (as well as great apes, to some extent) show visual preference for curved objects [16, 17]. In this experiment, most trials contained a curved path or a smooth ball, or both. This might explain part of the Maluma bias in the silent videos, if participants' decisions were guided by higher attention to the pleasant curved components. Auditory information, in contrast, has shown potentially higher alerting power than visual information [18]. Some studies have also reported higher attention to the auditory over the visual channel in high-arousal conditions, although contradictory evidence also exists [19, 20]. Altogether, audiovisual associations to Takete and Maluma are not yet explored in detail.

Regarding the auditory channel, associations of musical excerpts to Takete or Maluma were experimentally found [21], although the related analysis did not explain which factors may have determined the associations.We plan a further experiment, adding an auditory only condition and related signal analysis, to identify cues underlying Maluma and Takete responses. We hope to recruit more participants such that the A/V/AV modalities could be split between subjects. Literature into acoustic cues driving sound-shape symbolism mentions links between angularity, pitch, and other spectral aspects [22]. Material characteristics, such as hardness, might also drive the responses; high importance of auditory cues in identification of materials from bouncing events has been demonstrated [23].

As our data do not yet include the auditory only condition, it is possible that the Takete bias in the AV condition was caused by cross-modal enhancement, similar to the effect observed by Stein et al. [24]. They reported increased visual brightness in presence of an auditory noise burst, although later research has offered other than perceptual explanations for the effect [25].

Assuming that the audiovisual Takete effect was indeed caused by auditory influence, we turn to the question of how the auditory channel may have achieved such dominance; there is ample evidence of general visual dominance, for example in terms of the Colavita effect [26, 27]. In our experiment, the strong auditory influence could be explained, firstly, by higher attention to the auditory channel when sound appears. Attention is known to modulate the visual dominance effect [28]. Secondly, our congruent

stimuli likely increased the importance of auditory cues in feature integration based on the whole-object bias – the spreading of attention to other modalities containing coherent information (see [29]). Auditory dominance has also been demonstrated in situations involving temporal processing [30], or when the auditory channel is more reliable or contains more information, such as music.

5. CONCLUSIONS

Our results showed a credible Takete bias in the audio-visual versus visual condition. We cannot, however, provide a general conclusive answer about the potential of sound to bias a sensation. As we have discussed, the auditory feedback coming from the rolling balls may have biased our participants toward Takete due to specific "associative cues" in those sounds operating above their obvious interpretation, in terms of the physical events they reported about. If existing, such cues are yet to be understood. On the other hand, the suggestions posed by these results motivate the design of further experiments that could contribute to clarifying the role of sound in multi-sensory associations.

6. REFERENCES

[1] D. Maurer, T. Pathman, and C. Mondloch, "The shape of boubas: sound-shape correspondences in toddlers and adults," *Developmental Science*, vol. 9, no. 3, pp. 316–322, 2006.

[2] V. Ramachandran and E. Hubbard, "Synaesthesia: A window into perception, thought and language," *J. of Counsciousness Studies*, vol. 8, no. 12, pp. 3–34, 2001.

[3] A.-S. Crisinel, S. Jones, and C. Spence, "'The sweet taste of Maluma': Crossmodal associations between tastes and words," *Chemical Perception*, vol. 5, pp. 266–273, Aug. 2012.

[4] A. Gallace, E. Boschin, and C. Spence, "On the taste of "Bouba" and "Kiki": An exploration of word-food associations in neurologically normal participants," *Cognitive Neuroscience*, vol. 2, no. 1, pp. 34–46, 2011.

[5] H.-S. Seo, A. Arshamian, K. Schemmer, I. Scheer, T. Sander, G. Ritter, and T. Hummel, "Cross-modal integration between odors and abstract symbols," *Neuroscience Letters*, vol. 478, no. 3, pp. 175–178, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304394010005744

[6] K. Shinohara, N. Yamauchi, S. Kawahara, and H. Tanaka, "Takete and maluma in action: A cross-modal relationship between gestures and sounds," *PLOS ONE*, vol. 11, no. 9, pp. 1–17, 09 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0163525

[7] M. Koppensteiner, P. Stephan, and J. P. M. Jäschke, "Shaking takete and flowing maluma. non-sense words are associated with motion patterns," *PLOS ONE*, vol. 11, no. 3, pp. 1–13, 03 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0150610

[8] F. Fontana, "Association of haptic trajectories to takete and maluma," in *Haptic and Audio Interaction Design*, I. Oakley and S. Brewster, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 60–68.

[9] L. Fryer, J. Freeman, and L. Pring, "Touching words is not enough: How visual experience influences haptic–auditory associations in the "Bouba–Kiki" effect," *Cognition*, vol. 132, no. 2, pp. 164–173, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027714000559

[10] T. Graven and C. Desebrock, "Bouba or kiki with and without vision: Shape-audio regularities and mental images," *Acta Psychologica*, vol. 188, pp. 200–212, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001691817305176

[11] J. K. Kruschke, *Doing Bayesian data analysis - A tutorial with R, JAGS, and Stan*, 2nd ed. Academic Press, 2014.

[12] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[13] P.-C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *J. Stat. Softw.*, vol. 80, no. 1, 2017. [Online]. Available: http://www.jstatsoft.org/v80/i01/

[14] H. Piéron, "Recherches sur les lois de variation des temps de laten- ce sensorielle en fonction des intensités excitatrices," *L'Année Psychol.*, vol. 20, pp. 17–96, 1914.

[15] ——, *The Sensations*. New Haven, CT: Yale University Press, 1950.

[16] E. Munar, G. Gómez-Puerto, J. Call, and M. Nadal, "Common Visual Preference for Curved Contours in Humans and Great Apes," *PLoS One*, vol. 10, no. 11, p. e0141106, nov 2015. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0141106

[17] M. Bertamini, L. Palumbo, T. N. Gheorghes, and M. Galatsidas, "Do observers like curvature or do they dislike angularity?" *Br. J. Psychol.*, vol. 107, no. 1, pp. 154–178, 2016.

[18] G. A. D. Souza, L. A. Torres, V. S. Dani, D. S. Villa, A. T. Larico, A. MacIel, and L. Nedel, "Evaluation of visual, auditory and vibro-tactile alerts in supervised interfaces," *Proc. - 2018 20th Symp. Virtual Augment. Reality, SVR 2018*, vol. 2, no. April 2020, pp. 163–169, 2018.

[19] K. L. Shapiro, B. Egerman, and R. M. Klein, "Effects of arousal on human visual dominance," *Percept. Psychophys.*, vol. 35, no. 6, pp. 547–552, nov 1984.

[Online]. Available: http://link.springer.com/10.3758/BF03205951

[20] S. Van Damme, G. Crombez, and C. Spence, "Is visual dominance modulated by the threat value of visual and auditory stimuli?" *Exp. Brain Res.*, vol. 193, no. 2, pp. 197–204, feb 2009. [Online]. Available: http://link.springer.com/10.1007/s00221-008-1608-1

[21] M. Murari, A. Rodà, S. Canazza, G. D. Poli, and O. D. Pos, "Is Vivaldi smooth and takete? Non-verbal sensory scales for describing music qualities," *J. of New Music Research*, vol. 44, no. 4, pp. 359–372, 2015. [Online]. Available: https://doi.org/10.1080/09298215.2015.1101475

[22] K. Knoeferle, J. Li, E. Maggioni, and C. Spence, "What drives sound symbolism ? Different acoustic cues underlie sound-size and sound-shape mappings," *Sci. Rep.*, no. December 2016, pp. 1–11, 2017. [Online]. Available: http://dx.doi.org/10.1038/s41598-017-05965-y

[23] Y. De Pra, F. Fontana, H. Järveläinen, S. Papetti, and M. Simonato, "Does it ping or pong? Auditory and tactile classification of materials by bouncing events," *ACM Transactions on Applied Perception (TAP)*, vol. 17, no. 2, pp. 1–17, 2020.

[24] B. E. Stein, N. London, L. K. Wilkinson, and D. D. Price, "Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis." *J. Cogn. Neurosci.*, vol. 8, no. 6, pp. 497–506, nov 1996. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23961981

[25] C. Spence and M. K. Ngo, "Does attention or multisensory integration explain the cross-modal facilitation of masked visual target identification?" in *The New Handbook of Multisensory Processing*, B. E. Stein, Ed. MIT Press, 2012, ch. 18.

[26] F. B. Colavita, "Human sensory dominance," *Percept. Psychophys.*, vol. 16, no. 2, pp. 409–412, mar 1974. [Online]. Available: http://link.springer.com/10.3758/BF03203962

[27] C. Spence, C. Parise, and Y.-C. Chen, *The Colavita Visual Dominance Effect*. CRC Press/Taylor & Francis, 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22593876

[28] S. Sinnett, C. Spence, and S. Soto-Faraco, "Visual dominance and attention: The Colavita effect revisited," *Percept. Psychophys.*, vol. 69, no. 5, pp. 673–686, 2007.

[29] I. Fiebelkorn, J. Foxe, and S. Molholm, "Attention and multisensory feature integration," in *The New Handbook of Multisensory Processing*, B. E. Stein, Ed. MIT Press, 2012, ch. 21.

[30] B. H. Repp and A. Penel, "Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences." *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 28, no. 5, pp. 1085–1099, 2002. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.28.5.1085