Multisensory Research

The Impact of Singing on Visual and Multisensory Speech Perception in Children on the Autism Spectrum

Jacob I. Feldman[1,2,*], Alexander Tu[3,4], Julie G. Conrad[3,5], Wayne Kuang[3,6], Pooja Santapuram[3,7] and Tiffany G. Woynaroski[1,2,8,9]

[1]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA

[2]Frist Center for Autism and Innovation, Vanderbilt University, Nashville, TN 37212, USA

[3]Neuroscience Undergraduate Program, Vanderbilt University, Nashville, TN 37212, USA

[4]Present address: Department of Otolaryngology and Communication Sciences, Medical College of Wisconsin, Milwaukee, WI 53226, USA

[5]Present address: Department of Pediatrics, University of Illinois, Chicago, IL 60612, USA

[6]Present address: Department of Pediatrics, Los Angeles County and University of Southern California (LAC+USC) Medical Center, University of Southern California, Los Angeles, CA 90033, USA

[7]Present address: Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY 10032, USA

[8]Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, TN 37203, USA

[9]Vanderbilt Brain Institute, Vanderbilt University, Nashville, TN 37240, USA

*Corresponding author; e-mail: j.i.feldman@vumc.org

ORCID iDs: Feldman: 0000-0002-5723-5834; Tu: 0000-0003-2387-1688; Kuang: 0000-0003-3155-4121; Santapuram: 0000-0002-8284-9547; Woynaroski: 0000-0001-6513-1181

Supplementary Material

S1. Supplemental Methods

S1.1. Stimulus Visual and Auditory Properties

To evaluate differences in the auditory properties of the ‘ba’ and ‘ga’ syllables in each presentation format, we extracted the auditory signal and analyzed it in Praat (Boersma, 2002). First, we found the peak intensity for each of the three repetitions of the syllable and the time at which that peak intensity occurred. Then, we found the onsets and offsets for the auditory component of each repetition and subtracted the onset from the offset to measure the syllable duration.

To evaluate differences in the stimuli’s visual properties, we utilized Adobe Photoshop CC 2017 to measure the speaker’s peak mouth opening in pixels. This was accomplished in two steps: first, the frame from the time at which the peak intensity occurred, obtained via Praat, was extracted. Then, open mouth was extracted using the automated extraction tool (see Supplementary Fig. S1 for an example). and the area of the mouth opening, in squared pixels, was calculated.

The data for the intensity, duration, and mouth opening are presented in Supplementary Table S1. Data were then analyzed in a 2 (Syllable; ba vs. ga) × 2 (Presentation; spoken vs. sung) mixed-model ANOVA that treated repetition as an independent case. Syllable was analyzed as a between-case factor and presentation format was analyzed as a within-case factor. Only the main effects of presentation were presented in the manuscript, but see Supplementary Table S2 for full analysis of variance (ANOVA) results.

S1.2. Procedures

For an example time series for the audiovisual stimuli, see Supplementary Fig. S2. For visualizations of trials in each modality, see Supplementary Fig. S3.

S2. Supplemental Results

S2.1. Responses to McGurk Stimuli

To determine whether groups differed in their response patterns to McGurk stimuli, we conducted a 2 (Presentation) x 3 (Response; auditory [‘ba’] vs. visual [‘ga’] vs. fusion [‘da’ and ‘tha’]) × 2 (Group) mixed-model ANOVA (see Supplementary Table S2 for responses to McGurk stimuli by presentation and group). The interactions between presentation and response, $F_{2,76} = 0.44$, $p = 0.65$; response and group, $F_{2,76} = 2.36$, $p = 0.10$; and presentation and response and group, $F_{2,76} = 0.83$, $p = 0.44$, were all non-significant. There was a significant main effect of response. Across presentations and groups, participants were significantly more likely to (a) report the auditory stimulus ‘ba’ than they were the visual stimulus ‘ga’ ($p < 0.001$) or fusion (i.e., ‘da’ or ‘tha’; $p < 0.001$) and (b) report a fusion than they were the visual stimulus ($p < 0.001$).

Reference

Boersma, P. (2002). Praat, a system for doing phonetics by computer, *Glot Int.* **5**, 341–345.

Table S1.

Peak intensity, duration, and peak mouth opening of stimuli by syllable and presentation format.

Syllable and Repetition	Peak Intensity		Duration		Peak Mouth Opening	
	Sung	Spoken	Sung	Spoken	Sung	Spoken
Ba						
1	82.29	82.45	524	564	2953	2693
2	84.09	82.21	575	629	2780	2580
3	85.84	83.78	515	537	2618	2587
Ga						
1	78.73	84.28	455	597	2872	2159
2	83.87	83.96	485	575	2680	1978
3	84.68	83.57	484	579	2545	1853

Peak intensity was measured as the amplitude of the waveform using Praat (Boersma, 2002). Duration was measured in ms using Praat. Peak mouth opening was measured in squared pixels using Adobe Photoshop CC 2017 (Adobe Inc, San Jose, CA, USA).

Table S2.

Complete ANOVA results for visual and auditory stimulus properties.

Stimulus property	ME of syllable *F* *p*	ME of presentation *F* *p*	Interaction effect *F* *p*
Peak intensity	0.06	0.01	1.63
	0.825	0.913	0.270
Duration	1.53	60.14	13.61
	0.285	0.001	0.021
Peak mouth opening	9.93	157.92	61.25
	0.001	<0.001	0.001

ME, Main effect. All *p* values are based on df = (1,4). Peak intensity was measured as the amplitude of the waveform using Praat (Boersma, 2002). Duration was measured in ms using Praat. Peak mouth opening was measured in squared pixels using Adobe Photoshop CC 2017.

Table S3

Proportion of closed choice responses to McGurk stimuli by group and presentation.

	Spoken				Sung			
	‘ba’	‘ga’	‘da’	‘tha’	‘ba’	‘ga’	‘da’	‘tha’
Group	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)	*M* (SD)
Autism	0.69	0.02	0.08	0.21	0.70	0.03	0.10	0.17
(*n* = 20)	(0.39)	(0.03)	(0.16)	(0.32)	(0.35)	(0.04)	(0.16)	(0.25)
Non-autism	0.50	0.08	0.07	0.35	0.52	0.02	0.24	0.21
(*n* = 20)	(0.36)	(0.21)	(0.13)	(0.32)	(0.39)	(0.07)	(0.31)	(0.23)

‘Ba’ responses reflect perceptions that are consistent with the auditory signal, ‘ga’ responses reflect perceptions that are consistent with the visual signal, and both ‘da’ and ‘tha’ responses reflect fused perceptions.

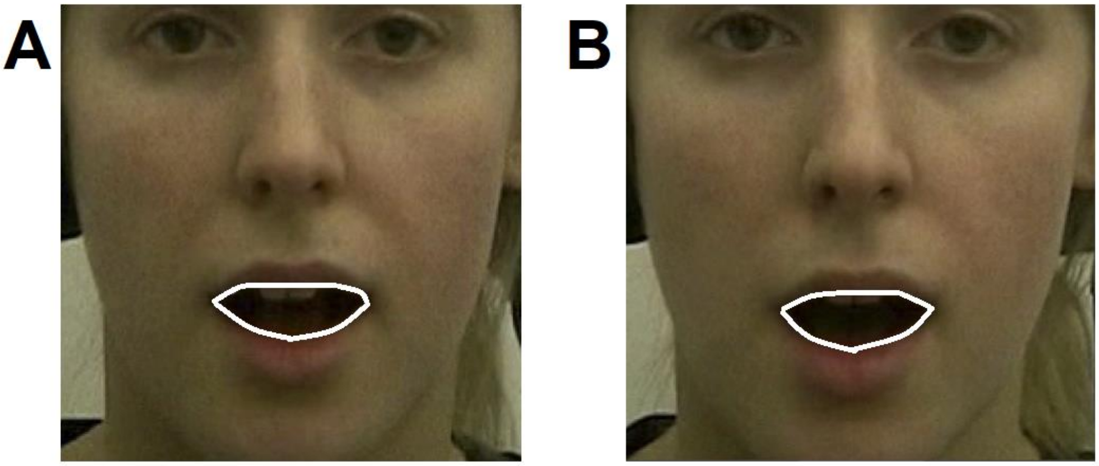

Figure S1. Visualization of measurement of peak mouth openness for ‘ba’ stimuli in the (A) spoken and (B) sung modalities.

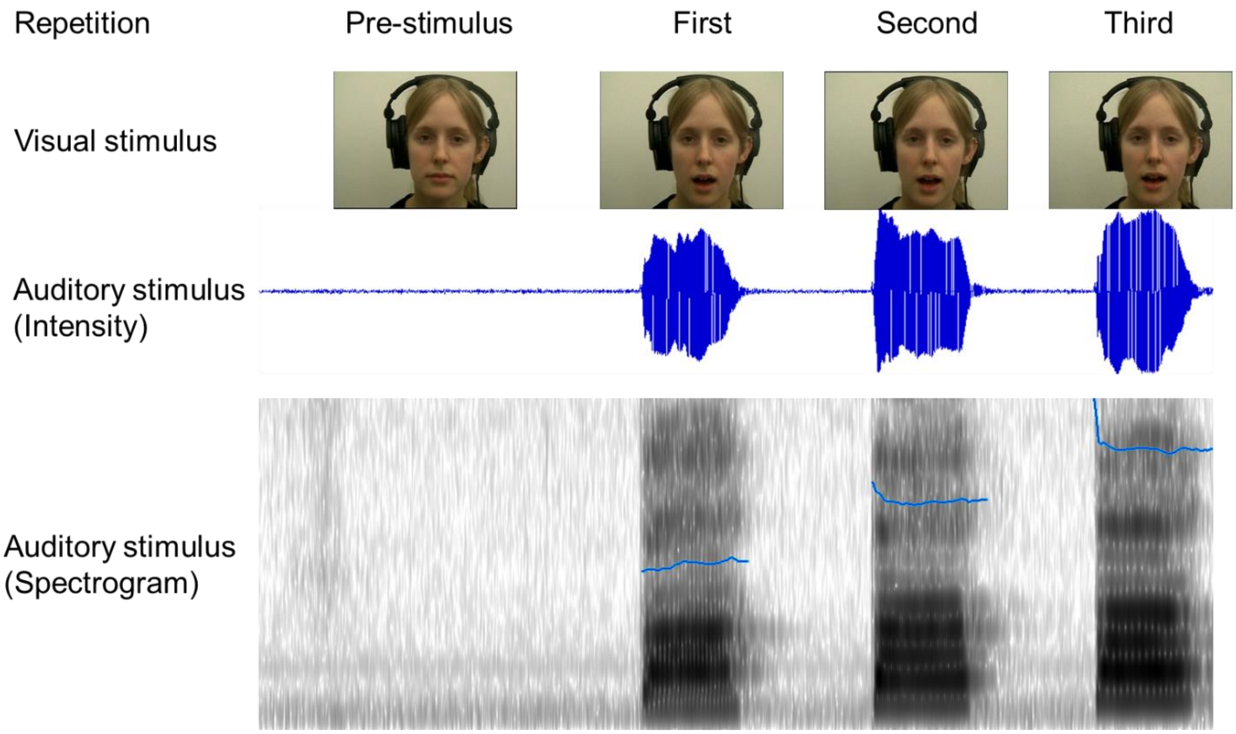
Repetition
Pre-stimulus
First
Second
Third
Visual stimulus
Auditory stimulus (Intensity)
Auditory stimulus (Spectrogram)

Figure S2. Time series of selected stimulus (incongruent audiovisual in sung modality). Presented stimulus is the three repetitions of the incongruent audiovisual (i.e., McGurk) stimulus in the sung modality. Prior to each repetition, each stimulus begins with silence (auditory-only condition and audiovisual conditions) and/or the speaker’s face with neutral expression and affect (visual-only condition and audiovisual conditions). Then, the syllable (in this example, auditory ‘ba’ and visual ‘ga’) is presented three times.

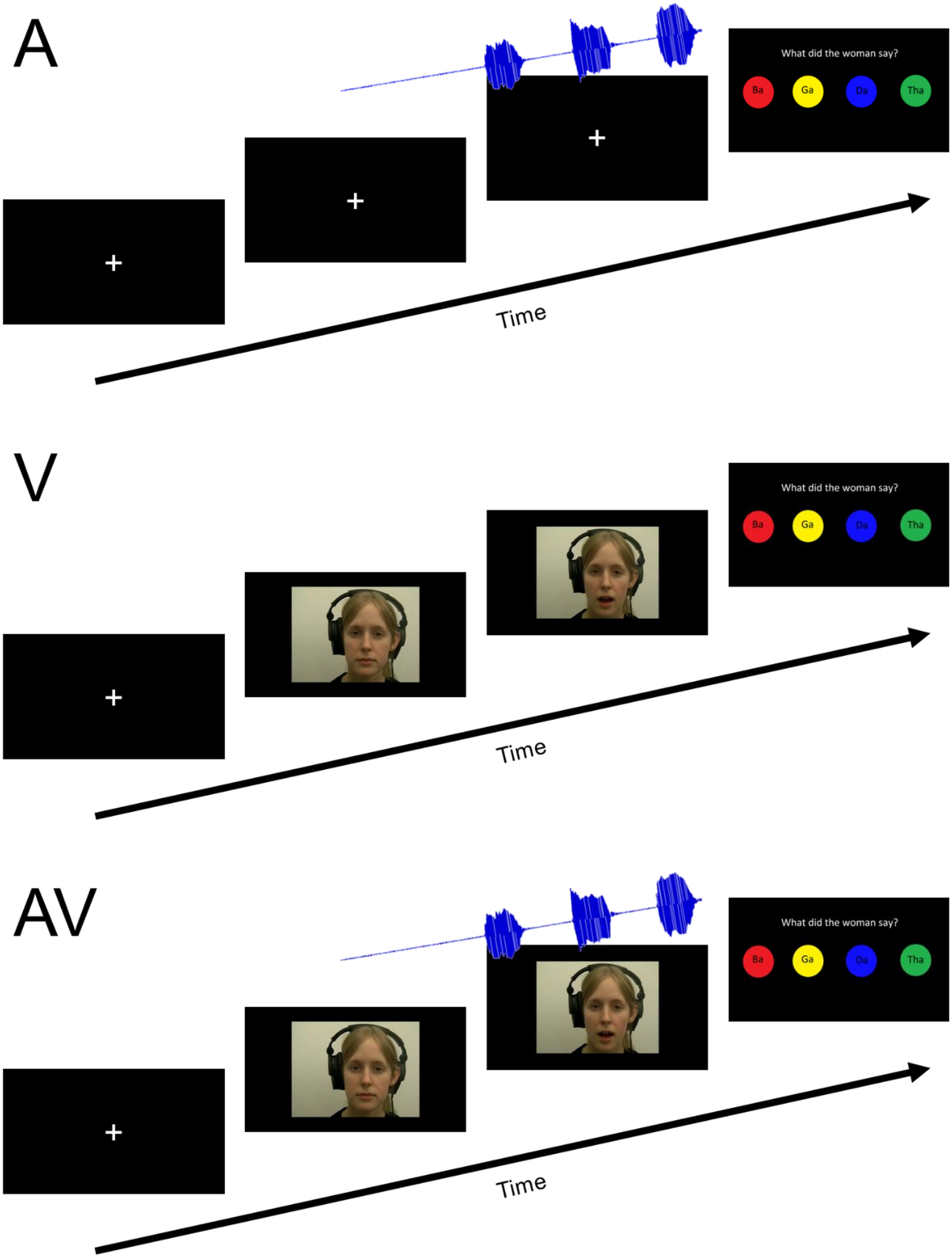

Figure S3. Visualizations of trials in the auditory-only, visual-only, and audiovisual modalities.