Generality of the Memory Advantage for Vocal Melodies

Michael W. Weiss, E. Glenn Schellenberg, &
Sandra E. Trehub
University of Toronto, Toronto, Canada

Children and adults, with or without music training, exhibit better memory for vocal melodies (without lyrics) than for instrumental melodies (Weiss, Schellenberg, Trehub, & Dawber, 2015; Weiss, Trehub, & Schellenberg, 2012; Weiss, Trehub, Schellenberg, & Habashi, 2016; Weiss, Vanzella, Schellenberg, & Trehub, 2015). In the present study, we compared adults' memory for vocal and instrumental melodies, as before, but with two additional singers, one female (same pitch level as the original female) and one male (7 semitones lower). In an exposure phase, 90 participants ($M = 4.1$ years training, $SD = 3.9$) rated their liking of 24 melodies—6 each in voice, piano, banjo, and marimba. After a short break, they heard the same melodies plus 24 timbre-matched foils (6 per timbre) and rated their recognition of each melody. Recognition was better for vocal melodies than for melodies in every other timbre, replicating previous findings. Importantly, the memory advantage was comparable across voices, despite the fact that liking ratings for vocal melodies differed by singer. Our results provide support for the notion that the vocal advantage in memory for melodies is independent of the idiosyncrasies of specific singers or of vocal attractiveness, arising instead from enhanced processing of a biologically significant timbre.

Received: February 8, 2016, accepted May 10, 2016.

Key words: melody recognition, voice perception, memory, music, timbre

Adults remember vocal melodies without lyrics (sung to *la*) better than instrumental melodies (Weiss, Trehub, & Schellenberg, 2012). This vocal melody advantage is demonstrable by 7 years of age (Weiss, Schellenberg, Trehub, & Dawber, 2015) and is also evident in Brazilian musicians and nonmusicians (Weiss, Vanzella, Schellenberg, & Trehub, 2015). In fact, Brazilian pianists remember vocal melodies better than piano melodies.

Replications across age, culture, and levels of expertise attest to the reliability of the findings, which are presumed to arise from the biological and functional significance of the voice. Greater pupil dilation to vocal than to piano melodies (Weiss, Trehub, Schellenberg, & Habashi, 2016), reflecting heightened arousal, is consistent with that presumption. Nevertheless, questions remain about the *generality* of the vocal melody advantage because all studies addressing this issue featured the same vocalist—an amateur, female singer in the alto range. In those studies, moreover, listeners liked the vocal melodies systematically less than the instrumental melodies despite remembering them better, raising the possibility that the vocal performances were unique in some respects.

In principle, female vocalizations could be more memorable than male vocalizations because of women's links to emotionality or nurture. In fact, women's affective vocalizations are perceived as more emotionally intense than those of men (Belin, Fillion-Bilodeau, & Gosselin, 2008). Moreover, their voices are more rapidly categorized as human or machine-like than are men's voices, even when the stimuli overlap in pitch level (Lévêque, Giovanni, & Schön, 2012). In such cases, the speed of categorization is affected by the gender of the speaker but not the listener. Gender identification for male and female stimuli is faster for women's voices but only for male listeners (Junger et al., 2013).

Male and female voices elicit different patterns of neural activation (Junger et al., 2013; Lattner, Meyer, & Friederici, 2005; Sokhi, Hunter, Wilkinson, & Woodruff, 2005) and voice recognition. For example, male students are less accurate at recognizing the voices of female classmates than male classmates, but female students' recognition of classmates' voices is independent of gender (Skuk & Schweinberger, 2013). These findings reveal no common theme other than the possibility of influences of stimulus and observer gender on discrimination and memory.

In the present study, we sought to confirm the generality of the vocal melody advantage, specifically its independence from person-specific features or from female voices in general. On the assumption that the biological and functional salience of the human voice promotes enhanced processing, we predicted better memory for vocal melodies than for instrumental melodies regardless

Music Perception, volume 34, issue 3, pp. 313–318, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2017.34.3.313

of voice identity or gender. We compared memory for vocal and instrumental melodies, as in previous research, but with additional vocalists. All participants heard the same instrumental timbres but were assigned randomly to one of three vocal conditions featuring (1) the female vocalist from previous studies of the vocal melody advantage (i.e., original vocalist), (2) a new female vocalist who sang at the same pitch level, or (3) a male vocalist who sang 7 semitones lower. If the vocal melody advantage stems from factors associated with biological salience, as hypothesized, then the advantage should be comparable for the three voices.

Method

PARTICIPANTS

The final sample of 90 undergraduates (24 male, 66 female, $M_{\text{age}} = 18.6$ years, $SD = 2.5$ years) was recruited without regard to music training and had moderate levels of classroom and formal lessons ($M = 4.1$ years, $SD = 3.9$, range 0-14 years). Twenty-nine participants (6 men) heard the original female voice, 32 (9 men) heard the new female voice, and 29 (9 men) heard the male voice. A chi-square test of independence confirmed that the male-to-female ratio was similar across conditions, $p > .60$. Seven additional participants were excluded from the final sample because of self-reported hearing problems ($n = 1$), equipment failure ($n = 1$), overall memory performance more than 2 *SD* below the sample mean (reflecting inattention to the task, $n = 2$), or misidentification of the gender of the vocalist ($n = 1$ for original female voice, $n = 2$ for male voice). All participants received partial course credit.

STIMULI AND APPARATUS

The stimuli were 48 excerpts of unfamiliar British and Irish folk melodies in vocal and instrumental (piano, banjo, marimba) renditions, 35 from previous research (Weiss, Vanzella, et al., 2015), and an additional 13 in the same style. Melodies were selected so that all notes remained at or higher than the low D on the banjo after transposition to the pitch level of male performances. The melodies varied in tempo (70 to 130 beats per minute), mode (major, minor), time signature (3/4, 4/4, 6/8), number of notes (20 to 52), and duration (12.7-21.0 s).

Each melody was sung without lyrics (i.e., to *la la*) and with neutral affect by three amateur vocalists: (1) the original female vocalist (alto) from Weiss and colleagues (Weiss, Schellenberg, et al., 2015; Weiss et al., 2012, 2016; Weiss, Vanzella, et al., 2015), (2) a new female vocalist at the same pitch level, and (3) a male vocalist 7 semitones lower than the female renditions. Vocalists sang to a MIDI "backing track" that was later discarded. Their renditions were pitch corrected in Melodyne (Celemony) by centering each note to the correct pitch and removing pitch drift. Corrected performances sounded natural, retaining the usual pitch fluctuations within notes. MIDI note information (pitch, onset, offset, velocity) generated from the vocal recordings of the original female vocalist (Melodyne) was quantized (Logic) and used to trigger digital sample-based instruments (piano, banjo, marimba). Separate instrumental renditions were created at the female and male pitch levels. MIDI instruments were used for convenience with respect to transposition because in previous research, no memory differences were observed between melodies with real and MIDI instruments (Weiss et al., 2012). All stimuli were amplitude (RMS) normalized with Sample Manager (Audiofile Engineering) and saved as high-quality monophonic audio files (16 bit, 44.1 kHz, .wav encoding).

PsyScript (version 2.3; Slavin, 2007) was used to present stimuli and collect responses on an iMac computer. Participants, who were tested individually in a sound-attenuating booth (Industrial Acoustics), heard the stimuli at a comfortable volume through high-quality headphones (Sony MDR-NC6).

PROCEDURE

Melodies were randomly assigned to exposure level (old, new) and timbre (voice, piano, banjo, marimba), separately for each participant. Participants were assigned randomly to one of the three voice conditions—original female, new female, or male—and heard that vocalist and the corresponding instrumental performances at the same pitch level. During the exposure phase, they heard half of the melodies ($n = 24$), distributed across four timbres (6 per timbre). Participants rated their liking of each on a 5-point scale ranging from 1 (*dislike*) to 5 (*like*). After exposure, participants completed a background questionnaire for 5-10 min. In the subsequent test phase, they heard the same 24 melodies (from exposure) intermixed with the remaining 24 melodies (6 per timbre). For each melody, they judged whether it had been presented before on a 7-point scale ranging from 1 (*definitely new*) to 7 (*definitely old*). Following the memory test, participants were asked if the voice they heard was male or female.

Results

LIKING

For each participant, four liking scores were calculated by averaging six original liking ratings separately for each timbre. Liking scores were analyzed with a two-way mixed-design ANOVA with vocalist as a between-subjects factor (original female, new female, male) and timbre as a repeated measure (voice, piano, banjo, marimba), with Greenhouse-Geisser correction of *p* values for violation of sphericity. There was no main effect of vocalist, $p > .20$, but there was a significant main effect of timbre, $F(3, 261) = 22.23$, $p < .001$, partial $\eta^2 = .20$, and a marginally significant interaction between timbre and vocalist, $F(6, 261) = 2.16$, $p = .06$, partial $\eta^2 = .05$, which motivated separate analyses of each vocalist.

One-way repeated-measures ANOVAs confirmed that liking differed across timbres for each of the three groups: original female, $F(3, 84) = 8.84$, $p < .001$, partial $\eta^2 = .24$; new female, $F(3, 93) = 10.69$, $p < .001$, partial $\eta^2 = .26$; male, $F(3, 84) = 7.14$, $p = .001$, partial $\eta^2 = .20$. Follow-up pairwise comparisons within each group (Bonferroni corrected) showed that all groups rated the banjo lower than the piano or marimba, $ps < .02$, and the banjo fared worse than the new female vocalist, $p = .001$. Participants in the original female group rated the voice lower than the piano and marimba, $p < .03$, but there were no other significant pairwise differences among timbres, $ps > .10$. In short, the differences among instruments were consistent from group to group (i.e., banjo liked the least), but liking for the voice relative to instruments varied by group. An additional one-way ANOVA confirmed that liking differed across voices, $F(2, 87) = 4.67$, $p = .01$, $\eta^2 = .10$, but pairwise comparisons revealed that the only significant difference was between the two female voices, $p = .010$. The original female voice was liked less than the new female voice (see Figure 1).

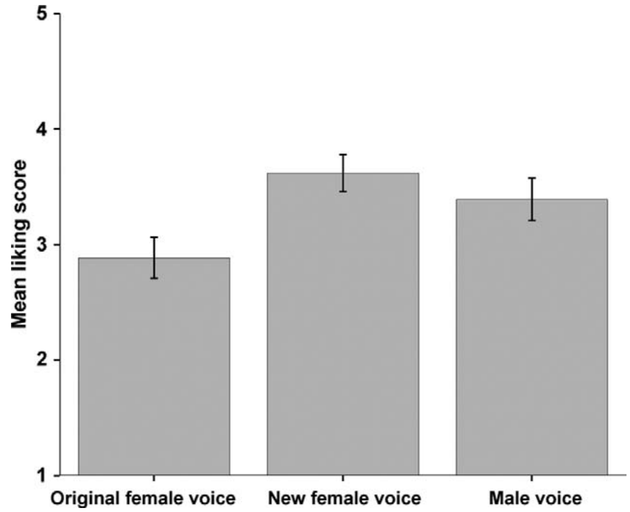

FIGURE 1. Average liking rating (1 = *dislike* to 5 = *like*) for vocal melodies heard during the exposure phase (i.e., between groups). Liking for the original female voice was lower than the new female voice. Error bars represent standard error of the mean.

RECOGNITION

Eight scores, each averaged over six original ratings, were calculated for each participant. A 3-way mixed-design ANOVA with vocalist (original female, new female, male) as the between-subjects factor and exposure level (old, new) and timbre (voice, piano, banjo, marimba) as repeated measures revealed no main effect of vocalist, $p > .60$, and no interactions with vocalist, $ps > .10$. There was a robust main effect of exposure level, $F(1, 87) = 594.17$, $p < .001$, partial $\eta^2 = .87$, confirming that the old melodies were remembered well. There was also a main effect of timbre, $F(3, 261) = 16.64$, $p < .001$, partial $\eta^2 = .16$, which was qualified by a significant interaction between exposure level and timbre, $F(3, 261) = 27.07$, $p < .001$, partial $\eta^2 = .24$. In short, memory performance (i.e., the *difference* between old and new ratings) differed as a function of timbre but not as a function of the specific voice. Descriptive statistics are provided in Figure 2.

Follow-up analyses revealed that difference scores (i.e., average old rating minus average new rating) for the voice ($M = 2.99$, $SD = 1.25$) were significantly larger than for all instrumental timbres (piano: $M = 1.77$, $SD = 1.05$; banjo: $M = 1.99$, $SD = 1.28$; marimba: $M = 1.95$, $SD = 1.24$), $ps < .001$, and there were no differences among instrumental timbres, $ps > .60$. As shown in Figure 2, this finding was evident and similar in magnitude for all three voices. The effect size (Cohen's *d*) of the vocal memory advantage (i.e., voice vs. average of three instruments) was 1.07 for the original female voice, 0.78 for the new female voice, and 0.97 for the male voice. In other words, vocal melodies elicited superior memory performance, as in previous research (Weiss, Schellenberg, et al., 2015; Weiss et al., 2012, 2016; Weiss, Vanzella, et al., 2015), and the vocal advantage extended across three different voices.

Gender, which was not balanced during recruitment (see *Participants*), had no influence on the results when considered as an additional between-subjects factor in the initial mixed-design ANOVA. Moreover, regrouping participants based on gender-matching with the vocalist ($n = 55$ matched, $n = 35$ mismatch) did not influence the results. When music training (median split) was added as an additional between-subjects variable in the original ANOVA, untrained participants (≤ 3 years, $n = 51$, $M = 3.98$, $SD = 1.44$) provided slightly higher recognition ratings for *all* melodies (old and new) compared

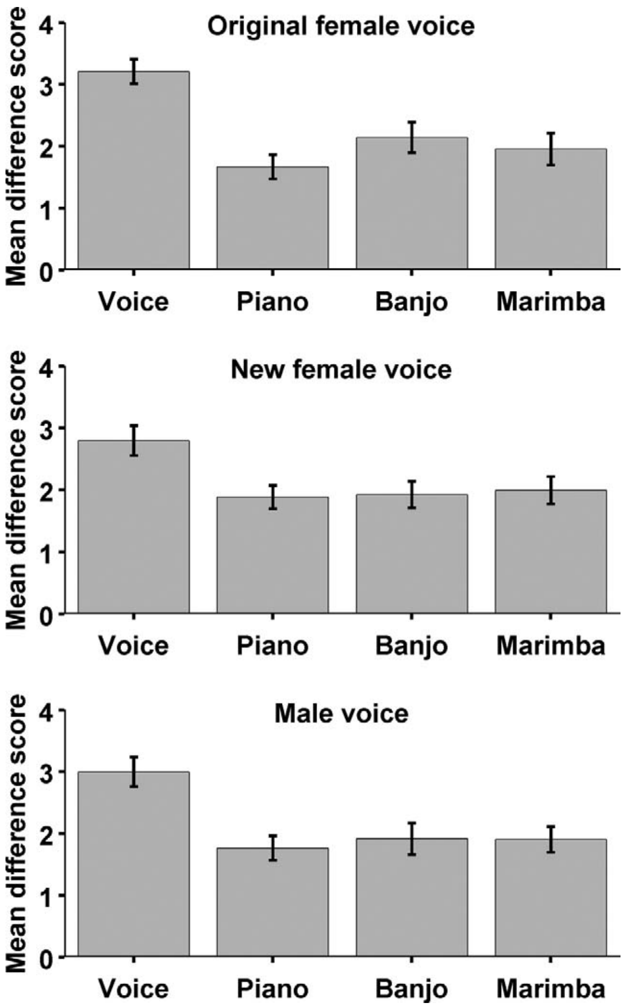

FIGURE 2. Average difference scores as a function of timbre in the recognition test, with separate plots for participants who heard different vocalists. Difference scores were calculated by subtracting average ratings (1 = *definitely new* to 7 = *definitely old*) for new melodies from average ratings for old melodies (i.e., larger difference scores indicate better memory; maximum score = 6, chance = 0). Error bars represent standard error of the mean.

to trained participants (> 3 years, $n = 39$, $M = 3.72$, $SD = 1.48$), $F(1, 84) = 6.57$, $p = .01$, partial $\eta^2 = .07$, but music training had no influence on the other results and did not interact with any other variable, *p*s > .10. Finally, all findings were replicated using modified *d*-prime scores, which were not used as the primary measure because of perfect performance in 179 of the 720 cells in the present design.

LIKING AND RECOGNITION

To confirm that liking did not contribute to the vocal memory advantage, we examined the correlation between the average liking rating at exposure and the average recognition rating for the same melodies at test, separately for each timbre. A significant positive correlation was observed for the piano timbre, $r(88) = .31$, $p = .003$, but not for the vocal, banjo, or marimba timbres, *p*s > .10. Repeating the correlation for the vocal timbre separately for the different groups of participants (original female, new female, male), revealed no significant correlations, *p*s > .20. It is unclear why liking and recognition for the piano were correlated in the present study but not in previous research (Weiss et al., 2012). It is clear, however, that liking for vocal melodies does not contribute to the memory advantage.

We also used an item analysis to examine correlations between liking and memory for individual melodies. For each melody, we calculated the average liking rating at exposure and the average recognition rating at test. Because liking ratings were collected only for melodies heard at exposure (i.e., old melodies), 45 of the 90 participants, on average, provided liking *and* recognition ratings for a given melody (range = 34-54). For all timbres combined, there was a strong positive correlation between liking of individual melodies and subsequent recognition confidence, $r(46) = .72$, $p < .001$, confirming the greater memorability of likeable melodies, as in Weiss, Vanzella, et al. (2015), and for music more generally (Stalinski & Schellenberg, 2013). Limiting the item analysis to melodies in a specific timbre necessarily reduced the number of ratings included in the rating averages ($M = 11.25$ ratings, range = 3-20). Nevertheless, significant positive correlations were observed for the voice, $r(46) = .29$, $p = .04$, the piano, $r(46) = .38$, $p < .01$, and the banjo, $r(46) = .52$, $p < .001$, with a marginally significant correlation for the marimba, $r(46) = .26$, $p = .07$. In other words, melodies that elicited higher liking ratings—regardless of timbre—were recognized with greater confidence.

Discussion

The present study explored the generality of the memory advantage for vocal melodies relative to instrumental melodies (Weiss, Schellenberg, et al., 2015; Weiss et al., 2012, 2016; Weiss, Vanzella, et al., 2015), specifically its independence from one particular vocalist and from female vocalists in general. Participants first rated how much they liked each melody, which was presented in vocal, piano, banjo, or marimba timbre. Their ratings revealed lower liking for the original female vocalist than for the other female vocalist and no differences between male and female vocalists. In the subsequent memory test, vocal melodies were recognized more confidently than instrumental melodies, regardless of vocalist and pitch level. Moreover, liking of melodies, gender of listener, and music training did not influence the memory advantage for vocal melodies.

We found no evidence that gender of listener or stimulus affected recognition, as it has in other auditory identification tasks (e.g., Skuk & Schweinberger, 2013). The difference in outcome may be attributable to our design, which did not emphasize structural features of the vocalist and presented only a single vocalist to each listener. Presenting several voices as a repeated measure may reveal subtle memory differences across vocalists. We predict, however, that any such differences would be trivial compared to the memory advantage for sung over instrumental melodies.

Our decision to compare a single voice to three instruments, and to include three percussive instruments, was driven by the intention to replicate the previously observed memory advantage (Weiss et al., 2012) while manipulating only one element of the design. Our design ensured an appropriate number of trials per cell, which could not be achieved by a fully within-participants design. We acknowledge, however, that item distinctiveness could affect memory (Talmi & McGarry, 2012). Future research could vary the composition of the list (i.e., number of vocalists, ratio of vocal to instrumental timbres, number of melodies per timbre) to explore the limits of the vocal memory advantage.

The present findings, in conjunction with previous evidence from listeners of different age, cultural background, and music training (Weiss, Schellenberg, et al., 2015; Weiss et al., 2012, 2016; Weiss, Vanzella, et al., 2015), confirm the memory advantage for vocal melodies as a robust and generalizable phenomenon. The acoustic features or constellation of features that underlie the processing advantage for vocal melodies remain to be determined. One important difference between vocal and instrumental sounds is considerable pitch variability in vocalizations of all kinds, which contrasts with the limited pitch variability of the piano, banjo, and marimba timbres in the present study and in previous studies of memory for melodies (Weiss, Schellenberg, et al., 2015; Weiss et al., 2012, 2016; Weiss, Vanzella, et al., 2015). Perhaps violin renditions of melodies, which incorporate voice-like pitch variation, or renditions created with artificial timbres and comparable pitch variation would yield a comparable boost in memory. Because hybrid or morphed timbres have been useful in the study of prosody (Kawahara, 2015), they may also be useful in specifying the features that are critical for enhanced processing of vocal melodies. It is possible, however, that the observed processing and memory advantages for vocal stimuli necessitate the full suite of features in natural vocal timbres.

Item analyses revealed correlations between liking and memory for melodies that were independent of the vocal memory advantage. The present performances of melodies, vocal or instrumental, were intentionally neutral rather than pleasant, which probably made them less pleasant than typical renditions. More attractive performances, achieved by averaging several renditions, could generate an even greater vocal melody advantage. Averaging or morphing multiple productions of a single syllable increases its attractiveness independent of speaker or listener gender, an effect attributed to by-products of averaging such as smoother voice texture (i.e., reduction of aperiodic noise) and greater similarity in pitch and timbre to prototypical voices (Bruckert et al., 2010).

Finally, the vocal processing advantages observed in the present study and in our previous studies raise questions about the continued use of synthesized or natural instrumental music to explore various aspects of music cognition. The use of vocal music in a variety of tasks is likely to reveal further processing advantages and more nuanced perspectives on the ability of human listeners to perceive and remember music.

Author Note

This research was funded by the Natural Sciences and Engineering Research Council of Canada.

We thank Sana Irshad and Agatha Rodrigues for assistance in testing.

Correspondence concerning this article should be addressed to E. Glenn Schellenberg, Department of Psychology, Deerfield Hall, 4th Floor, Room 4098, 3359 Mississauga Road, Mississauga, ON, Canada, L5L 1C6. E-mail: g.schellenberg@utoronto.ca

References

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*, 531-539. DOI: 10.3758/BRM.40.2.531

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., et al. (2010). Vocal attractiveness increases by averaging. *Current Biology, 20*, 116-120. DOI: 10.1016/j.cub.2009.11.034

Junger, J., Pauly, K., Bröhr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., et al. (2013). Sex matters: Neural correlates of voice gender perception. *NeuroImage*, *79*, 275-287. DOI: 10.1016/j.neuroimage.2013.04.105

Kawahara, H. (2015). Temporally variable multi attribute morphing of arbitrarily many voices for exploratory research of speech prosody. In K. Hirose & J. Tao (Eds.), *Speech prosody in speech synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis* (pp. 109-120). Heidelberg, Germany: Springer. DOI: 10.1007/978-3-662-45258-5_8

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, *24*, 11-20. DOI: 10.1002/hbm.20065

Lévêque, Y., Giovanni, A., & Schön, D. (2012). Effects of humanness and gender in voice processing. *Logopedics Phoniatrics Vocology*, *37*, 137-143. DOI: 10.3109/14015439.2012.687763

Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research*, *296*, 131-140. DOI: 10.1016/j.heares.2012.11.004

Slavin, S. (2007). *Psyscript* (Version 2.3) [Computer software]. Lancaster, UK: Lancaster University. Retrieved from http://www.lancaster.ac.uk/psychology/research/research-software/psyscript2/

Sokhi, D. S., Hunter, M. D., Wilkinson, I. D., & Woodruff, P. W. R. (2005). Male and female voices activate distinct regions in the male brain. *NeuroImage*, *27*, 572-578. DOI: 10.1016/j.neuroimage.2005.04.023

Stalinski, S. M., & Schellenberg, E. G. (2013). Listeners remember music they like. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 700-716. DOI: 10.1037/a0029671

Talmi, D., & McGarry, L. M. (2012). Accounting for immediate emotional memory enhancement. *Journal of Memory and Language*, *66*, 93-108. DOI: 10.1016/j.jml.2011.07.009

Weiss, M. W., Schellenberg, E. G., Trehub, S. E., & Dawber, E. J. (2015). Enhanced processing of vocal melodies in childhood. *Developmental Psychology, 51*, 370-377. DOI: 10.1037/a0038784

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, *23*, 1074-1078. DOI: 10.1177/0956797612442552

Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils dilate for vocal or familiar music. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1061-1065. http://doi.org/10.1037/xhp0000226. Advance online publication. DOI: 10.1037/xhp0000226

Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *Quarterly Journal of Experimental Psychology*, *68*, 866-877. DOI:10.1080/17470218.2015.1020818