Article

IJRCS

International Journal of Research in Choral Singing

The Scientific Research Journal of the American Choral Directors Association

International Journal of Research in Choral Singing
(2022) Vol. 10 108-130

The Impact of Manual Mimicry Gestures on the Learning of Sung German Phonemes

Caron Daley,[1] Heather L. Rusiewicz,[2] James Schreiber,[3] and Julianna Grabowski[4]

Abstract

The integration of embodied pedagogies has a long history in music education, and especially in choral singing (Benson, 2011; Ehmann, 1968; Jaques-Dalcroze, 1921). Manual mimicry gestures are hand movements that mirror the spatiotemporal attributes of speech sounds with an analogous gesture (Rusiewicz & Rivera, 2017). The present study investigated the effects of manual mimicry gestures on the articulatory accuracy, vocal technique, and expressive artistry of non-native German speech sounds in singing. Twenty-four college-aged voice majors were assigned to three groups, each with a different instructional mode (i.e., no gesture, viewed gestures, viewed and produced gestures). Four German sounds were tested in isolation and in the context of a musical phrase (fricatives ichlaut /ç/ and achlaut /x/, and mixed vowels /y/ and /Y/). Expert listeners rated the participants' singing at three time points (baseline, immediate post-instruction, and 48-hours post-instruction) using visual analog scales. Results revealed improved articulatory accuracy and vocal technique for all sounds in all training conditions as perceived by the raters. Individuals who produced gestures during training were not rated significantly higher than the groups trained without gestures or by viewing gestures only. Implications for vocal pedagogy and related professions are discussed, as well as future directions for research.

[1] Mary Pappert School of Music, Duquesne University
[2] John G. Rangos Sr. School of Health Sciences, Duquesne University
[3] School of Nursing, Duquesne University
[4] Mary Pappert School of Music (alumna), Duquesne University

Corresponding author:
Caron Daley, Duquesne University, Mary Pappert School of Music, 600 Forbes Ave, Pittsburgh, PA 15282.
Email: daleyc1@duq.edu

The Impact of Manual Mimicry Gestures on the Learning of Sung German Phonemes

Singing requires a range of embodied competencies, such as the efficient coordination of parts of the body, the integration of ideas with actions, and the pronunciation and expression of text. Soloists and choral singers sing in diverse languages, and as such, vocal instruction must address the articulation of individual speech sounds and sound combinations not found in the vocalist's first language (Mahaney, 2006). The International Phonetic Alphabet has been a common tool to help singers pronounce non-native sounds but does not substantially relate to the physical actions of singing, including the spatiotemporal configurations of the vocal apparatus. Other common instructional methods have included listening to audio and video pronunciation recordings, rote teaching, or learning from culture bearers and native speakers (Chris, 2019; Sieck, 2013).

This investigation was motivated by two separate but related lines of scholarly work in the disciplines of choral pedagogy and speech-language pathology. In the choral setting, physical gesturing and body movement have been a primary mode of communication for the conductor and an effective choral pedagogical tool (Benson, 2011; Hibbard, 1994; Wis, 1999). Speech-language pathologists (SLPs) have integrated gestural cues into many structured therapeutic approaches for speech sound disorders and accent modification (DeThorne et al., 2009; Hammer, 2006; Square et al., 2014). Choral conductors instruct singers how to pronounce non-native phonemes, a process that Daley & Rusiewicz (2021) have suggested is akin to learning speech sound targets in a speech-language pathology setting.

Movement and Gesture in Vocal Learning

The roots of kinesthetic teaching and learning in music can be traced to Émile Jaques-Dalcroze (1865–1950), who theorized that the human body was the primary instrument of musical instruction. Dalcroze Eurhythmics is designed to cultivate mind-body connection through the integrated development of hearing, moving, and feeling in music learning. It has served as a type of double education, strengthening perception of movement and spatial awareness while also generating a vocabulary of movement for use in playing and singing repertoire (Caldwell, 1995; Dutoit, 1971).

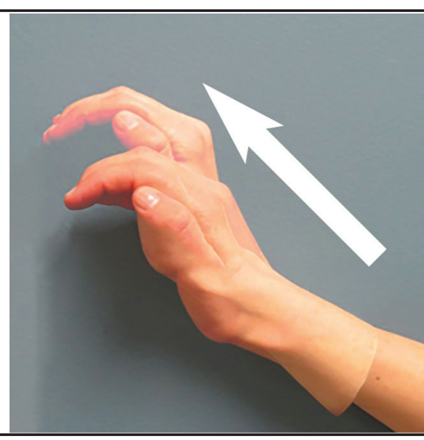
Applied to choral music learning, several authors have noted the benefit of body movement activities on musicianship, including the development of inner pulse and sense of rhythm, precision in intonation, vocal balance within the ensemble, and group interpretation of a musical score (Crosby, 2008; Daley, 2013; Liao & Davidson, 2016; Shenenberger, 2008). The use of movement has also facilitated vocal technical development by assisting in the coordination of various structures of the vocal mechanism for optimal sound production (Galván, 2008; Nafisi, 2013). Movement has likewise been applied to expressive goals in performance to support phrasing, text declamation and tone of voice, and stylistic decision-making (Caldwell, 1995; Galván, 2008; Pierce, 2007). Oney (2017) asserted that Dalcroze Eurhythmics developed singers' confidence and lowered performance anxiety, while several studies have indicated the benefit of movement instruction in choirs on student engagement, motivation, and enjoyment of learning (Manganello, 2011; McCoy, 1986; Wis, 1993).

A guiding principle of Dalcroze Eurhythmics is that sound and gesture exist in a reciprocal relationship, such that any sound has an analogous gesture and any gesture has an analogous sound (Jaques-Dalcroze, 1921). This relationship is often a feature of choral teaching and performance gesture, but it also occurs spontaneously in the rehearsal setting. Mimicry arises between the conductor's bodily movement (including conducting gestures) and the singer's bodily movements, suggesting that conductors' gestures affect singer's production of sound (Manternach, 2016). Daugherty and Brunkan (2013) discovered that choral singers responded nonverbally to conductor's lip shaping while singing, mimicry that resulted in a change in vocal tone and acoustical spectra. Grady (2014) indicated that conductors' gestural vocabularies had a direct impact on choral singers' use of spectral energy while singing, and that similar types of conducting gestures elicited similar sound results in choirs.

Manual Mimicry in Speech-Language Pathology

Physical gestures have been frequently employed during intervention for speech sound objectives in the discipline of speech-language pathology. These movements included tapping out syllables, pointing to a vocal articulator, using a gesture that is specific to a phoneme but not similar in shape or movement to the phoneme (e.g., handshapes used with Cued Speech; Krause et al., 2011), using the hands to manipulate the articulators (e.g., clinician closing the lips to facilitate lip closure) as an oral placement technique (Bahr & Rosenfeld-Johnson, 2010), and myriad other functions. Most hand gestures used in therapeutic contexts are arbitrary and although they may spark a visual mnemonic, they do not capitalize on underlying interactions of the speech and hand motor systems. Manual mimicry gestures (i.e., hand movements that mirror the movements and characteristics of speech sound and voice targets), by contrast, are entrained (i.e. coordinated) with the underlying motor processes of the speech mechanism (Rusiewicz & Rivera, 2017) (see Figure 1). In additional to behavioral

Figure 1
Starting and ending position for the manual mimicry gesture for /ç/.



(e.g., Smith et al., 1986), and kinematic evidence (e.g., Gentilucci et al., 2001), the entrainment of movements of the vocal apparatus and hands has been hypothesized to arise from shared neuroanatomical substrates including areas such as the lateral perisylvian cortex, supplementary motor cortex, premotor cortex, cerebellum, and Broca's area (e.g., Heiser et al., 2003; Rizzolatti & Arbib, 1998).

In speech-language pathology, manual mimicry gestures can represent three aspects of the targeted sound. These parameters include:

> (1) the spatiotemporal configuration required for the accurate production of a segmental target (e.g., using the hand to mirror the articulators in space, time, and tension for /r/); (2) the perceptual quality of a suprasegmental target (e.g., using the hand to represent a desired intonation contour), and (3) the physiological and/or perceptual characteristics of a voice target (e.g., moving the hands open and forward to cue key movements and sound quality for forward resonance). (Rusiewicz, 2020, pp. 9–10)

Manual mimicry gestures can also represent other prosodic features of speech, such as pulsing or tapping the hand to mark prosodic prominence or moving the hand in time to mirror the desired rate of speech.

Previous investigations supported the effect of manual mimicry gestures on the accuracy of /r/ produced by a college-aged woman with persisting childhood apraxia of speech and distortions of the /r/ phoneme in five vowel + /r/ combinations (e.g., /ar/). Using only manual mimicry gestures and minimal verbal instructions as cues, the participant increased accuracy of /r/ from an average 11% at baseline to 90% after the second phase of treatment in this ABAB withdrawal design according to judgments made by 28 naïve listeners (Rusiewicz & Rivera, 2017). Additionally, case study data of two children with CAS indicated that manual mimicry gestures facilitated more accurate speech production (Koshut et al., 2016).

There is also evidence of the impact of hand movements for pronunciation training for non-native sounds in Japanese (e.g., Hirata et al., 2014). Most recently, Xi et al. (2020) found that viewing hand movements that mirrored the phonetic characteristics of Mandarin consonants (e.g., burst of air for aspirated plosives) resulted in higher pronunciation accuracy for these sounds produced by fifty Catalan speakers as rated by five native Mandarin speakers. Importantly, increased accuracy was related to the appropriateness of the gesture to the desired phonetic characteristics. The effects of participants producing gestures while learning non-native speech skills was not studied by Xi et al.

Purpose

The purpose of this study was to examine the impact of simple hand gestures that mirror the movements of the speech articulators in terms of spatial configuration, timing, and tension (i.e., manual mimicry gestures) on the immediate learning and retention of four German phonemes sung by college-aged voice majors. More specifically, the aim was to study the effect of training with and without the integration of manual mimicry gestures on the following parameters of vocal performance: (a) articulatory accuracy, (b) vocal technique, and (c) expressive artistry. These vocal variables were measured across three time points (baseline, immediately following training, and 48hr +/- 8 hours post training) to assess immediate learning and retention of knowledge. The following research questions guided this investigation:

1. What effect does training (No Gesture, View Gesture, or View+Do Gesture) have on articulatory accuracy and vocal technique of sung German consonants and vowels immediately following training and 48-hours after training?

2. What is the impact of training for German consonants and vowels on the transfer to sung phrases immediately following training compared to 48-hours after training according to ratings of articulatory accuracy, vocal technique, and expressive artistry?

Method

Study Design

A 3 (learning condition) x 3 (time of assessment) factorial design was used for the study. The participants were assigned to one of three learning condition groups: (a) no gestures during instruction or *No Gesture,* (b) gestures viewed during instruction or *View Gesture,* and (c) gestures viewed and mimicked during instruction or *View+Do Gesture.* This design was chosen to control for confounding variables and practice effects. The times of assessment were (a) pre-testing (baseline), (b) immediate post-instruction, and (c) 48-hour post-instruction (+/- 8 hours) to assess immediate learning and retention of knowledge. Participants were pseudo-randomized to one of three groups based on their class standing (freshman, sophomore, junior, senior, or graduate) to ensure equal numbers of students in each year of undergraduate and graduate study in each group. Four speech sound targets were tested for articulatory accuracy and vocal technique, the voiceless palatal and velar fricatives spelled with “ch” (/ç/ and /x/) and the mixed vowels spelled “ü” (/y/ and /ʏ/). The gestures used for the View Gesture and View+Do Gesture condition groups can be seen in Figure 2. Expert listeners perceptually rated the sung phrases for (a) articulatory accuracy, (b) technical accuracy, and (c) expressive artistry using a 100 mm visual analog scale by three expert raters. Transfer of learning was also explored using a phrase sung by each training group immediately and 48 hours after training.

Institutional Review Board approval for human subject research was obtained in August 2018 and participants were recruited through live invitation and posted flyers. Twenty-four college-aged students (18 women and 6 men) between the ages of 18 and 25 who were majoring in voice completed the study. Students were excluded from participation if they had concerns about their vocal health or if they spoke German fluently or had previously studied German for two or more semesters. All procedures were video-recorded and audio-recorded.

Following informed consent procedures, the participants completed baseline pretesting. Participants were seated in an armless chair approximately 18 inches away from a mounted iPad. The instructions, stimuli, training, and assessments were presented via PowerPoint on an Apple iPad Air with a 12.9 inch display. Participants listened to the training via Sennheiser RS120 headphones. An AKG C520L condenser microphone and Presonus AudioBox USB Preamplifier recorded their sung productions. The microphone was kept at a consistent two inch distance to the left of participants' mouths. Participants were instructed to sing five repetitions of each of the target consonants and vowels following a sung model by the instructor for these initial baseline data (on F3/F4). Training began immediately following these baseline procedures.

Training and Testing

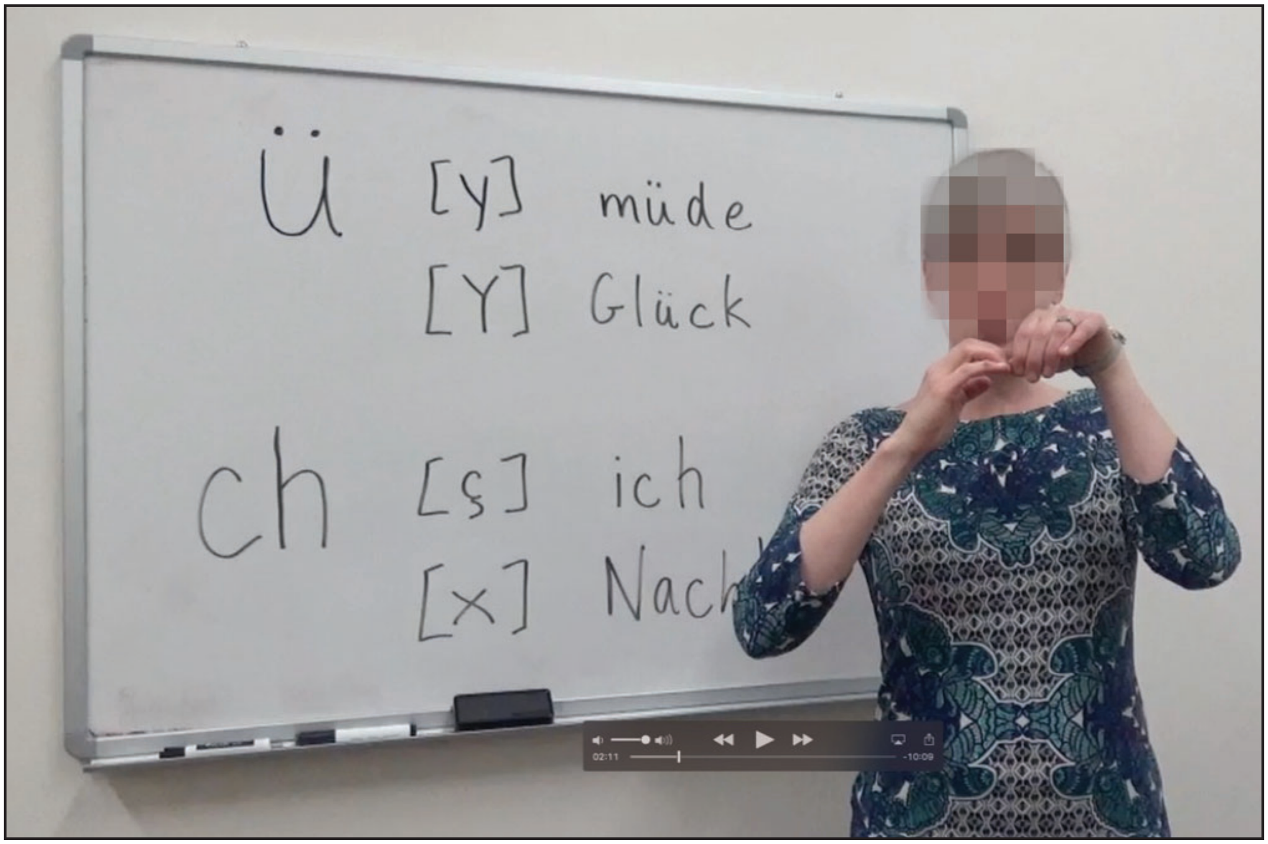
Study participants received instruction via video recordings of an instructor presented on the iPad (i.e., No Gesture, View Gesture, View+Do Gesture) in a counterbalanced manner across the two training units (i.e., palatal and velar fricatives, and mixed vowels; see above). The principal investigator was the instructor in all videotaped segments. The instructor was videotaped with a view of torso, arms, and face. A script was followed for all conditions to equate language structures, general content, and speech intonation characteristics of the sessions. Each training session lasted approximately 10–12 minutes, regardless of condition. Participants received verbal instruction and modeling of the target phonemes in all training conditions, with the number of models consistent in each of the three training units. For each sound, participants were instructed to sing the sound five times on a single sung pitch (F3 or F4 depending on voice type) and told that their fifth production would be be used for rating and analysis.

For only the View Gesture and View+Do Gesture groups, the instructor used manual mimicry hand gestures that approximated the spatiotemporal positioning and movement of the vocal articulators while performing the target phonemes (see Figure 2 on the next two pages). For the palatal and velar fricatives, a one-handed gesture indicated the movement of the air toward the hard palate (/ç/) or the velum (/x/). For the mixed vowel targets, one hand represented tongue advancement and tension, and one hand represented the rounding of the lips (a high and tense tongue combined with tightly-rounded lips for /y/ and and slightly lower and more lax tongue and slightly less-rounded lips for /ʏ/). These gestures were developed by the authors. For the View+Do Gesture groups, participants were instructed to produce the hand gestures simultaneously with their sung productions.

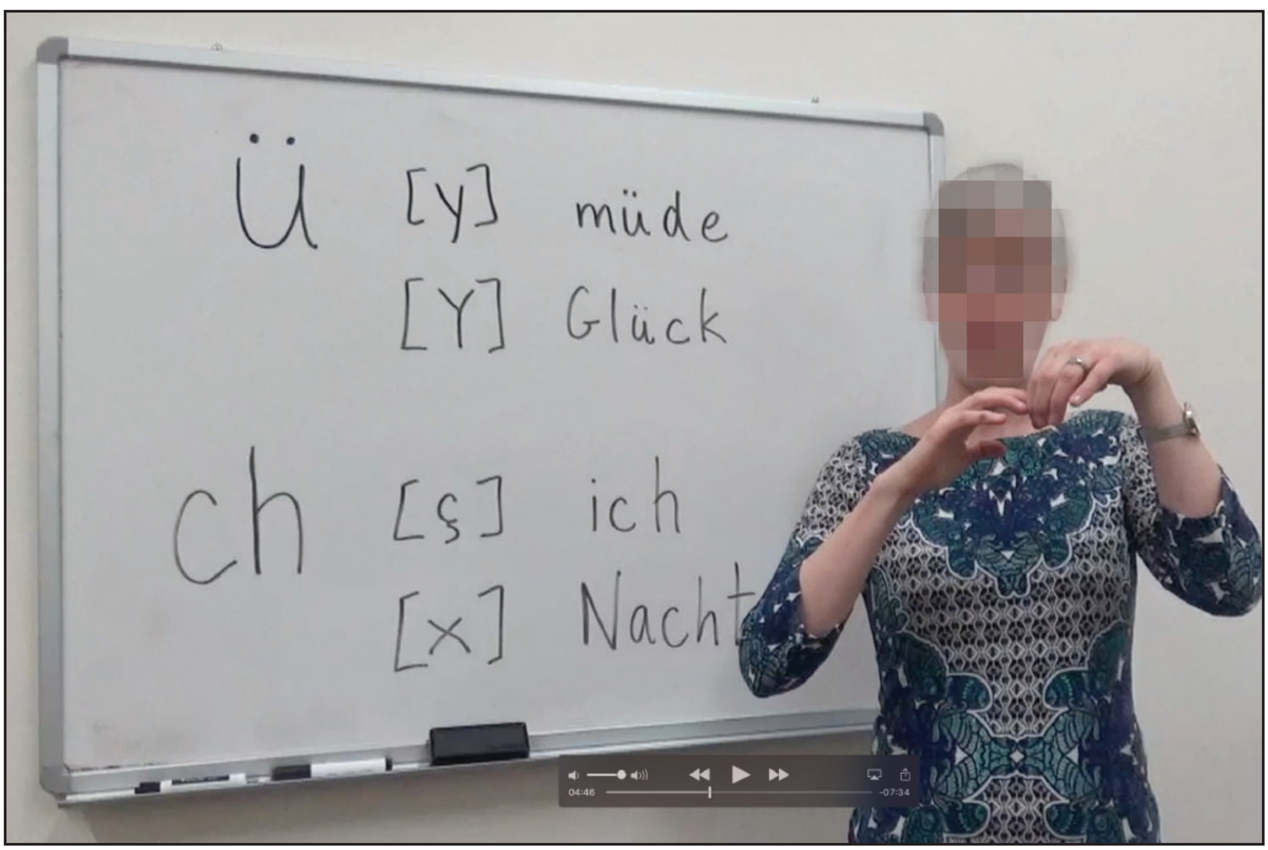
Immediately following instruction, the same assessment procedures were conducted as in the pre-test/baseline procedures. In addition, the participants were instructed to implant the target sounds into a model phrase, "Ich bin so müde, aber es ist ein Glück, dass die

Figure 2, Parts a, b, c, & d

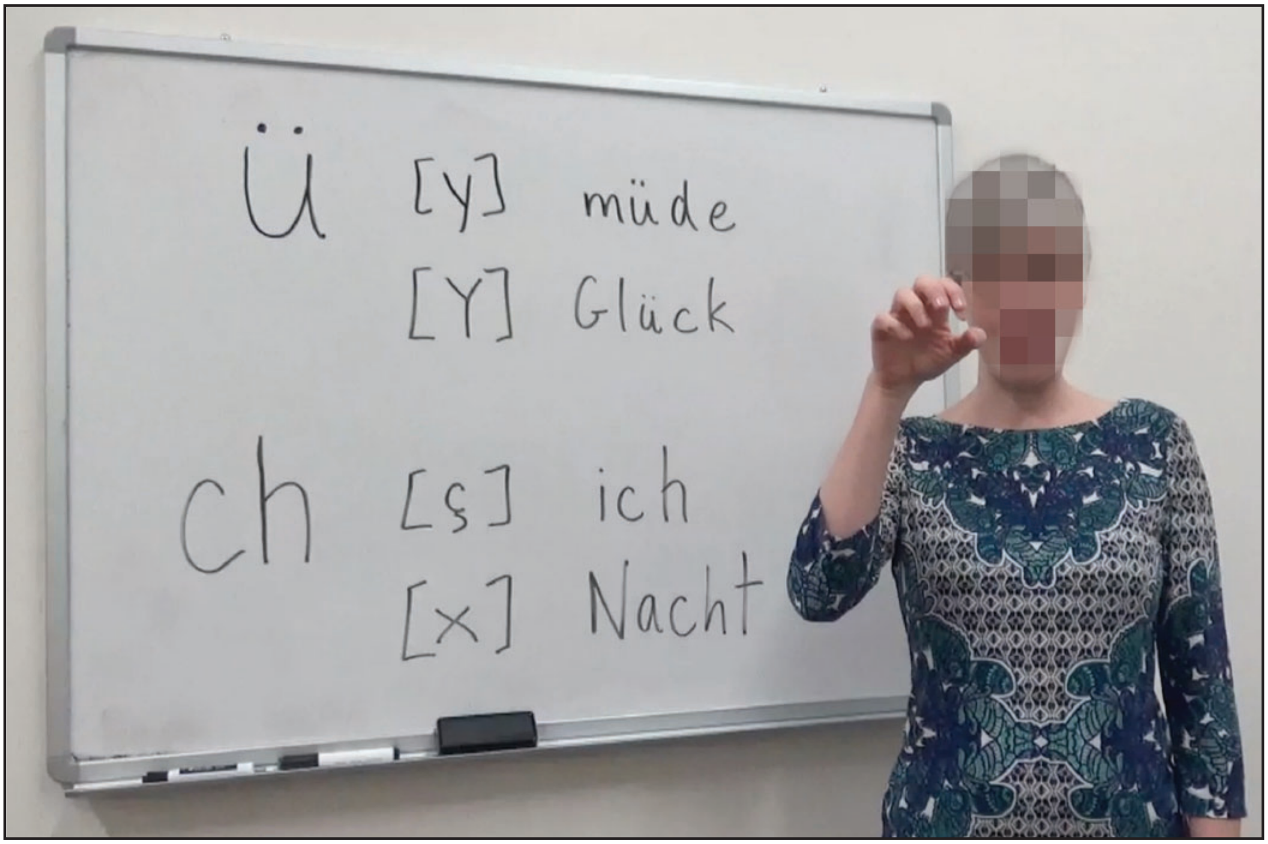
Screenshot of instructional video demonstrating a manual mimicry gesture for /y/.



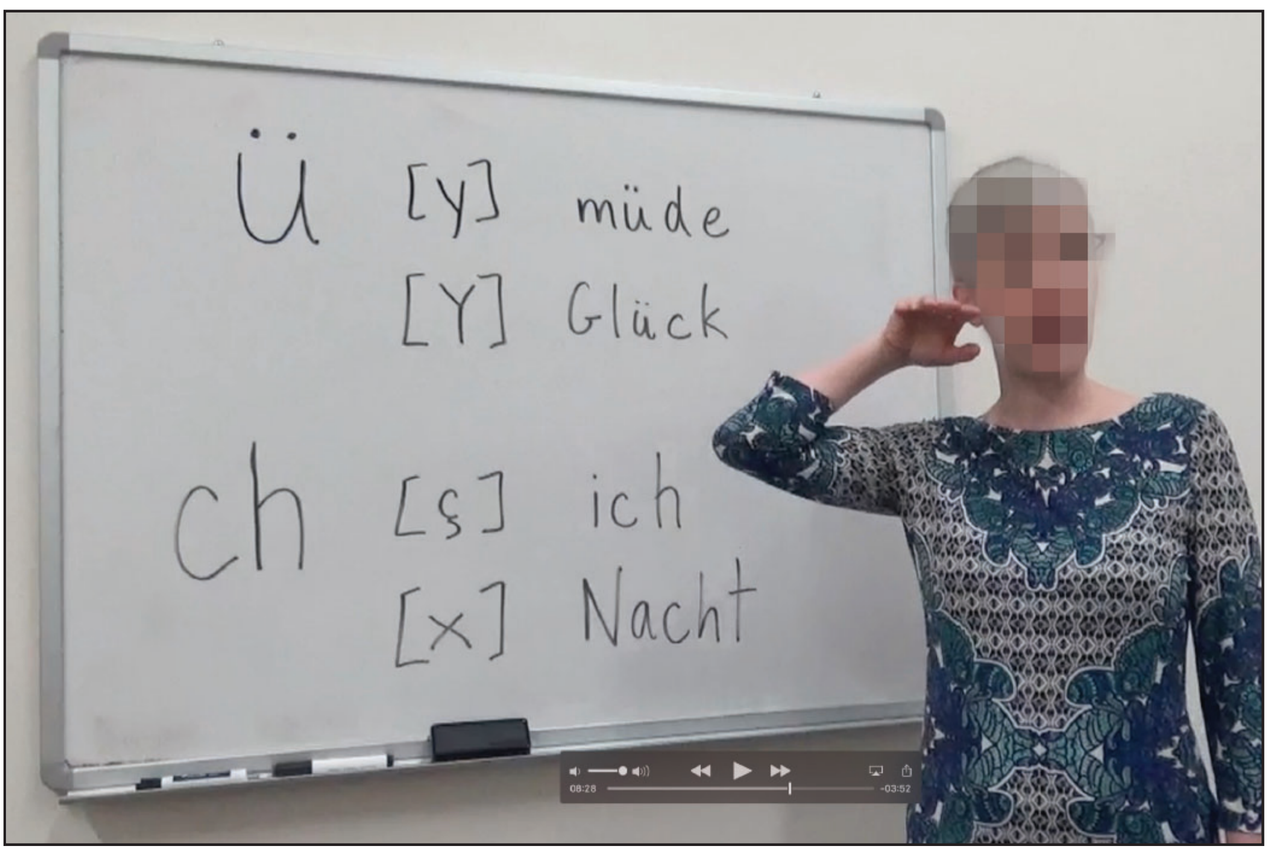
Screenshot of instructional video demonstrating a manual mimicry gesture for /Y/.



Screenshot of instructional video demonstrating a manual mimicry gesture for /ç/.



Screenshot of instructional video demonstrating a manual mimicry gesture for /x/.

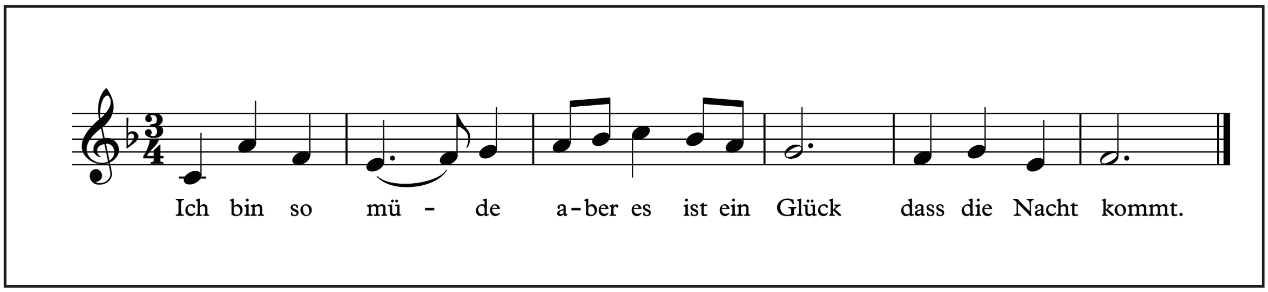


Nacht kommt" (Figure 3). The words and music for this phrase were composed by the first author. This phrase was modeled in a speaking voice by the instructor. The participants then viewed musical notation and heard the model phrase played three times on a piano. They were instructed to sing the phrase five times and told that their fifth production would be used for rating and analysis.

The participants returned for a follow-up session (48 hours +/- 8 hours) to assess consolidation and retention of learning at 48 hours post-instruction (e.g., Cook et al., 2013). This second and final session lasted approximately 20 minutes. Participants recorded each of the four sounds and the phrase in the same manner as in immediate post-instruction.

Figure 3

Sample musical phrase with sounds embedded



Data Reduction and Analysis

A total of 1,584 sung productions were recorded across the 24 participants. The final production of each trial was extracted from the audio signal using Audacity software and saved as an individual .wav file, for a total of 336 targets (288 individual phonemes and 48 phrases). Task fidelity was assessed for all trials for all 24 participants by a trained student research assistant who was blinded to training conditions and the purpose of the study. This research assistant viewed each video and coded if the participant gestured. As expected, 100% of the participants in the View+Do gesture condition produced gestures for all productions during training, while 100% of the participants in the No Gesture and View Gesture training conditions did not produce gestures during the training.

Collegiate vocal instructors with doctoral degrees served as expert raters for the data. None of the voice instructors were familiar with study participants. The participants' productions were presented via PowerPoint in the following order: [ʏ], [x], [y], [ç], and the musical phrase. Each sung production was embedded as an isolated .wav file into a single slide. The presentation order of the trials was randomized across the time of assessment. Ratings were inputted into QualtricsXM software using a continuous 100 mm graphical/visual analog scale. Raters slid the scale to mark their perception of skill for each dependent measure for each production. These ratings were converted to quantitative data between 0 and 100 and then averaged across the three raters. Data for the individual sounds were analyzed using a 3 (Group: No Gesture, View Gesture, View+Do Gesture) x 2 (Sound: Consonants, Vowels) x 3 (Time: Pre-Training, Immediate Post-Training, 48-Hours Post-Training) repeated measures analysis of variance (ANOVA; Schreiber & Asner-Self, 2011). Phrase data were analyzed using 3 (Group: No Gesture, View Gesture, View+Do Gesture) x 2 (Time: Immediate Post-Training, 48-Hours Post-Training) repeated measures ANOVA.

Results

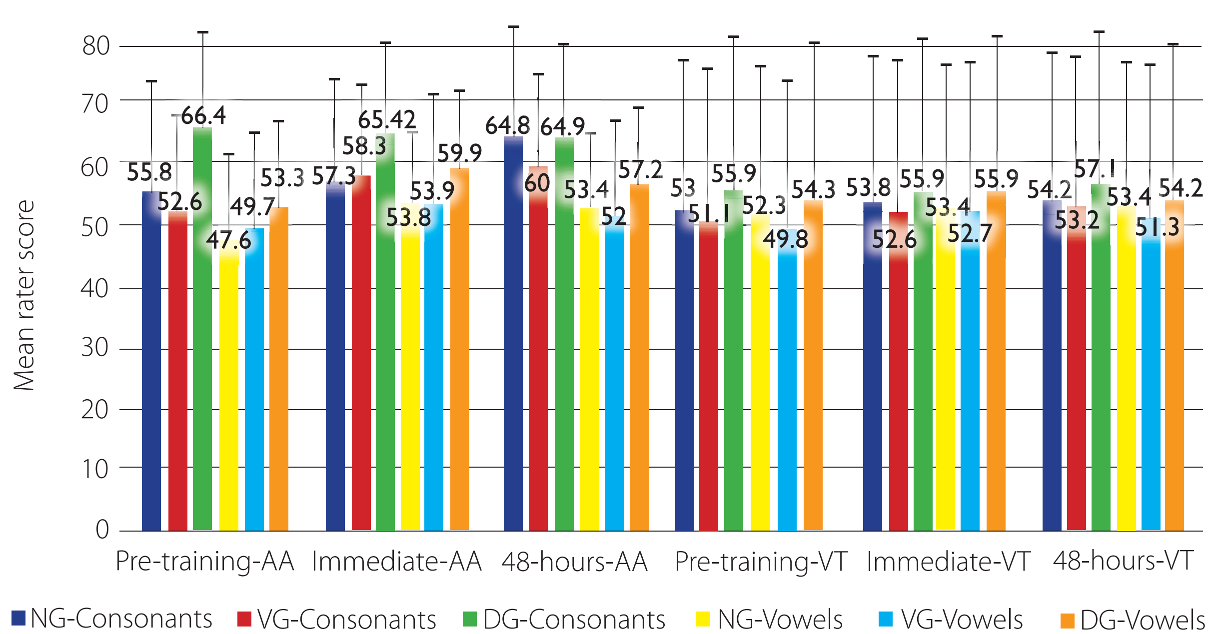
Research Question 1: Articulatory Accuracy and Vocal Technique of Consonants and Vowels

Articulatory Accuracy

The mean ratings for articulatory accuracy were averaged across the three expert listeners and analyzed for each of the three training conditions (No Gesture, View Gesture, View+-Do Gesture), the three time points (baseline/pre-testing, immediate post-testing, 48-hours post-testing), and type of sound (Figure 4). Means and standard deviations for articulatory accuracy ratings for all variables are found in Table 1 on the next page. A repeated measures ANOVA yielded several significant results. First, the data revealed differences between the ratings of articulatory accuracy between the three groups with a moderate effect size $F(2, 282) = 6.65$, $p < 0.001$, $\eta_p^2 = 0.045$, such that the View+Do Gesture training group

Figure 4

Mean Ratings of Expert Listeners for Articulatory Accuracy (AA) and Vocal Technique (VT): Consonants and Vowels for No Gesture (NG), View Gesture (VG), and View+Do Gesture (DG) Groups



Note. Error bars show standard deviations.

Table 1
Articulatory Accuracy Descriptive Data (Means and Standard Deviations) for All Training Conditions, Sounds, and Time Points

			M (SD)
Baseline/Pre-training	Vowel	No Gesture	47.58 (14.24)
		View Gesture	49.73 (15.81)
		View+Do Gesture	53.33 (14.06)
		Total	50.22 (14.81)
	Consonant	No Gesture	55.77 (18.12)
		View Gesture	52.60 (18.44)
		View+Do Gesture	66.38 (15.67)
		Total	58.25 (18.32)
	Total	No Gesture	51.68 (16.73)
		View Gesture	51.17 (17.14)
		View+Do Gesture	59.85 (16.19)
		Total	54.23 (17.11)
Immediate Post-training	Vowel	No Gesture	53.83 (11.65)
		View Gesture	53.90(17.77)
		View+Do Gesture	59.90 (12.47)
		Total	55.88 (14.41)
	Consonant	No Gesture	57.27 (16.73)
		View Gesture	58.31 (16.87)
		View+Do Gesture	65.42 (14.74)
		Total	60.33 (16.43)
	Total	No Gesture	55.55 (14.44)
		View Gesture	56.10 (17.38)
		View+Do Gesture	62.66 (13.86)
		Total	58.10 (15.59)

Continued on the next page

48 hours Post-training	Vowel	No Gesture	53.40 (11.93)
		View Gesture	52.00 (16.04)
		View+Do Gesture	57.23 (12.45)
		Total	54.21 (13.68)
	Consonant	No Gesture	64.81 (15.17)
		View Gesture	60.02 (18.11)
		View+Do Gesture	64.88 (15.06)
		Total	63.24 (16.22)
	Total	No Gesture	59.10 (14.74)
		View Gesture	56.01 (17.48)
		View+Do Gesture	61.05 (14.27)
		Total	58.72 (15.65)

had the highest observed scores. Second, there was a significant difference in the ratings of articulatory accuracy for vowels and consonants, such that consonants had higher overall accuracy scores $F(1,282) = 19.36$, $p < 0.001$, $\eta_p^2 = 0.064$. There was no interaction between group and sound. Third, all three groups (No Gesture, View Gesture, View+Do Gesture) significantly improved their articulatory accuracy ratings following training according to the within analysis $F(1,282)= 36.67$, $p <0.001$, $\eta_p^2 = 0.11$. There were no significant interactions between training group conditions with time point or sound, indicating that the presence of gesture did not differentially impact ratings of articulatory accuracy.

Vocal Technique

The mean ratings for vocal technique were averaged across the three expert listeners and analyzed for each of the three independent variables, with data for these variables displayed in Figures 4. Means and standard deviations of vocal technique ratings for all variables are found in Table 2 on the next page. A repeated measures ANOVA revealed a main effect for group (No Gesture, View Gesture, View+Do Gesture), with ratings of vocal technique significantly improved following training according to the within analysis $F(1,282)= 21.28$, $p <0.001$, $\eta_p^2 = 0.07$. There were no other significant main effects for training group or sound. A significant interaction of sound by time point was found such that consonants demonstrated greater improvement over time compared to vowels $F(1,282)=3.64$, $p <0.03$, $\eta_p^2 = 0.013$. Aligned with the articulatory accuracy analyses, there were no significant interactions between training group condition with time point or sound, indicating that the presence of gesture did not differentially impact ratings of articulatory accuracy.

Table 2
Vocal Technique Descriptive Data (Means and Standard Deviations) for All Training Conditions, Sounds, and Time Points

			M (SD)
Baseline/Pre-training	Vowel	No Gesture	52.33 (23.99)
		View Gesture	49.79 (24.36)
		View+Do Gesture	54.27 (25.90)
		Total	52.13 (24.66)
	Consonant	No Gesture	52.98 (23.56)
		View Gesture	51.08 (24.95)
		View+Do Gesture	55.88 (25.05)
		Total	53.31 (24.43)
	Total	No Gesture	52.66 (23.65)
		View Gesture	50.44 (24.53)
		View+Do Gesture	55.07 (25.35)
		Total	52.72 (24.51)
Immediate Post-training	Vowel	No Gesture	53.42 (23.25)
		View Gesture	52.67 (24.61)
		View+Do Gesture	55.94 (25.54)
		Total	54.01 (24.35)
	Consonant	No Gesture	53.75 (22.90)
		View Gesture	52.58 (24.39)
		View+Do Gesture	56.88 (24.88)
		Total	54.40 (23.97)
	Total	No Gesture	53.58 (22.96)
		View Gesture	52.63 (24.37)
		View+Do Gesture	56.41 (25.08)
		Total	54.20 (24.12)

Continued on the next page

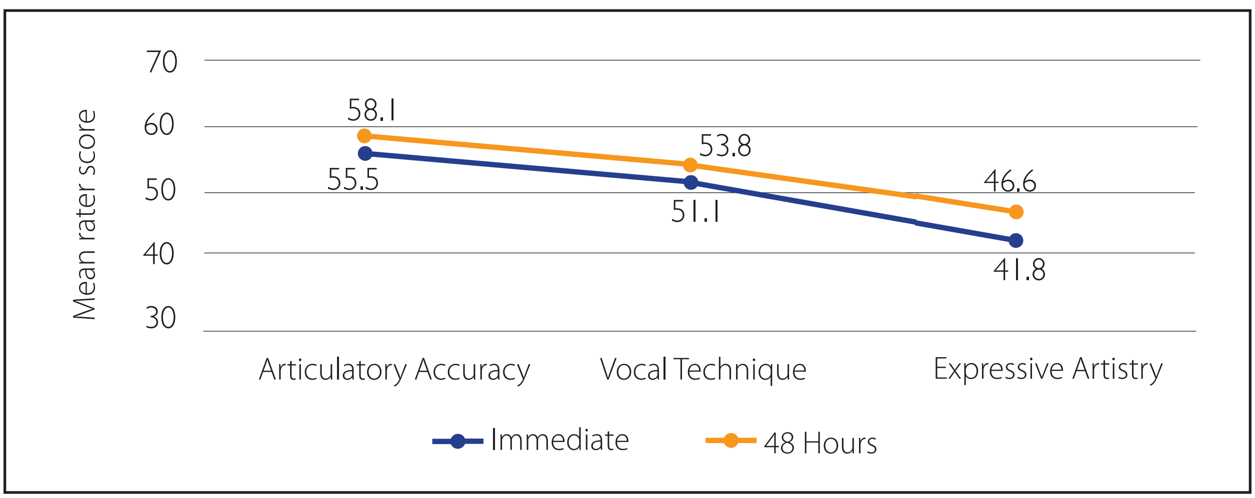
48 hours Post-training	Vowel	No Gesture	53.35 (23.75)
		View Gesture	51.35 (25.23)
		View+Do Gesture	54.23 (25.60)
		Total	52.98 (24.72)
	Consonant	No Gesture	54.25 (23.82)
		View Gesture	53.19 (24.61)
		View+Do Gesture	57.10 (24.87)
		Total	54.85 (24.32)
	Total	No Gesture	53.80 (23.67)
		View Gesture	52.27 (24.81)
		View+Do Gesture	55.67 (25.14)
		Total	53.91 (24.50)

Research Question 2: Articulatory Accuracy, Vocal Technique, and Expressive Artistry of Phrases

Similar to the analyses for individual sounds, the mean ratings for articulatory accuracy, vocal technique, and the additional variable, expressive artistry, of phrases were averaged across the three expert listeners and analyzed for each of the three training conditions (No Gesture, View Gesture, View+Do Gesture) and two time points (immediate posttesting and 48-hours posttesting) as displayed in Figure 5. Means and standard deviations for articula-

Figure 5

Mean Ratings of Expert Listeners for Articulatory Accuracy, Vocal Technique, and Expressive Artistry for Phrases Sung Immediately and 48-Hours after Training



tory accuracy ratings for all variables are found in Table 3. The highest mean ratings were for articulatory accuracy, followed by vocal technique, and then expressive artistry for both immediately and 48-hours following training. Retention was noted for all three dependent variables in improved ratings from immediate post-training to 48-hours post-training. The phrase data show small to large effect sizes from immediate post-training to 48-hour after training with ratings of expressive artistry demonstrating the largest effect size (Table 4).

Table 3

Descriptive Data (Means and Standard Deviations) for Articulatory Accuracy, Vocal Technique, and Expressive Artistry for Phrases Sung Immediately and 48-Hours After Training

	M	SD
Articulatory Accuracy: Immediate Post-Training	55.53	15.522
Articulatory Accuracy: 48-Hours Post-Training	58.08	14.066
Vocal Technique: Immediate Post-Training	51.08	24.945
Vocal Technique: 48-Hours Post-Training	53.85	25.482
Expressive Artistry: Immediate Post-Training	41.85	27.577
Expressive Artistry: 48-Hours Post-Training	46.57	27.479

Table 4

Paired T-test Analysis and Effect Size Measures

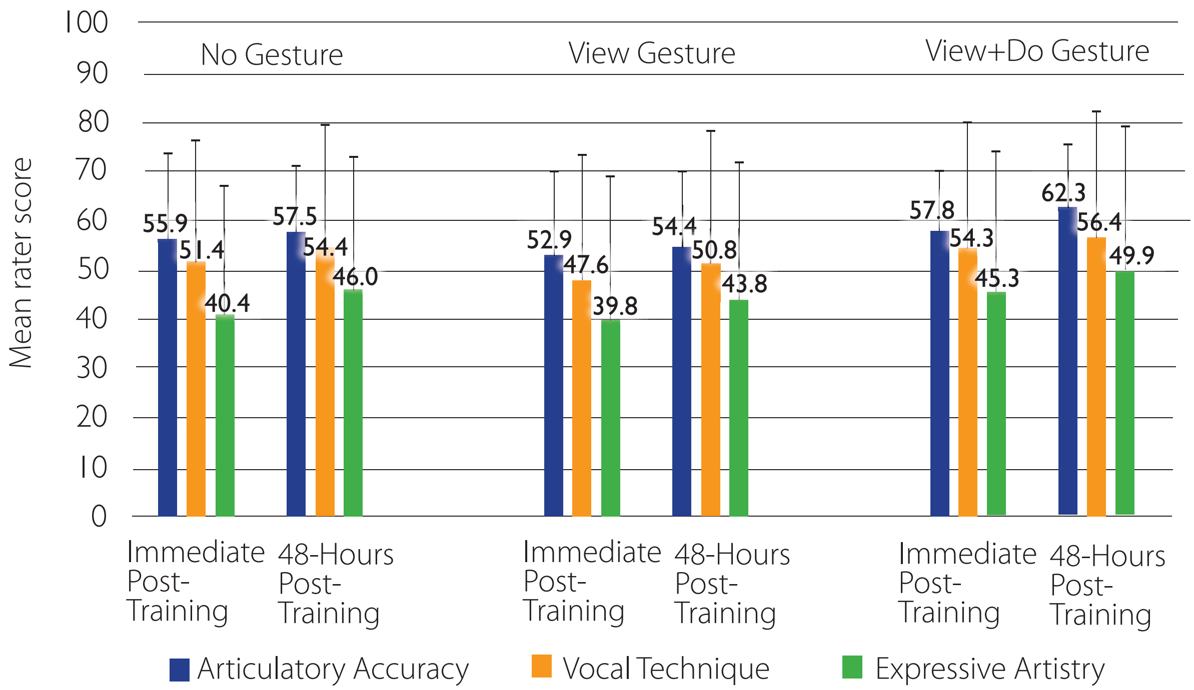
	t-value	df	*p*	Cohen's *d*
Articulatory Accuracy: Immediate Post – 48-Hours Post-Training	-2.817	71	p < .01	0.33
Vocal Technique: Immediate Post – 48-Hours Post-Training	-5.283	71	p <. 01	0.62
Expressive Artistry: Immediate Post – 48-Hours Post-Training	-6.246	71	p <.01	0.73

Articulatory Accuracy

A repeated measures ANOVA, 3 (Group: No Gesture, View Gesture, View+Do Gesture) x 2 (Time: Immediate Post-Training, 48-Hours Post-Training), was used to further examine ratings of articulatory accuracy in the sung phrases. This analysis indicated no meaningful difference between groups $F(2, 69) = 1.22$ and no interaction between groups and time $F(2, 69) = 1.21$(Figure 6). The within-component results indicated all groups significantly improved from immediate post-training to 48-hours after training $F(1, 69) = 38.37$, $p = <0.001$, $\eta_p^2 = 0.034$.

Figure 6

Mean Ratings of Expert Listeners for Articulatory Accuracy, Vocal Technique, and Expressive Artistry of Sung Phrases Immediately and 48-Hours After Training



Note. Error bars show standard deviations.

Vocal Technique

A repeated measures ANOVA, 3 (Group: No Gesture, View Gesture, View+Do Gesture) x 2 (Time: Immediate Post-Training, 48-Hours Post-Training), was used to examine ratings of vocal technique in the sung phrases. The between-group results indicated no meaningful difference between groups $F(2, 69) = <1$ as displayed in Figure 6 and no interaction between groups and time $F(2, 69) <1$. The within-component results indicated that all groups improved from immediate post-training to 48-hours after training $F(1, 69) = 27.43$, $p = <0.001$, $\eta_p^2 = 0.28$.

Expressive Artistry

A repeated measures ANOVA, 3 (Group: No Gesture, View Gesture, View+Do Gesture) x 2 (Time: Immediate Post-Training, 48-Hours Post-Training), was used to examine ratings of expressive artistry of the sung phrases. As with the other two dependent measures, the between-group results indicated no meaningful difference between groups $F(2, 69) < 1$ and no interaction between groups and time $F(2, 69) < 1$, see Figure 6. Once again, the within-component results revealed change over time, such that all groups were rated higher for expressive artistry at 48-hours post-training compared to immediate post-training $F(1, 69) = 38.37$, $p = < 0.001$, $\eta_p^2 = 0.36$. Because a main effect was only found for time, paired t-tests (Table 4) were completed to evaluate the two sets of data (i.e., immediate post-training and 48-hours post-training) for articulatory accuracy ($p < .01$), vocal technique ($p < .001$), and expressive artistry ($p < .001$).

Discussion

The present investigation addressed whether training with and without hand gestures that mimic the spatiotemporal configurations of the vocal apparatus (manual mimicry gestures) affected the learning and retention of German speech sounds in singing. Four German phonemes were assessed, the palatal and velar fricatives spelled with "ch" (/ç/ and /x/), and the mixed vowels spelled with an "ü" (/y/ and /Y/). Additionally, these four sounds were embedded into a sung German phrase to assess implementation in a more naturalistic context. Twenty-four participants were assigned to three control groups: No Gesture, View Gesture, and View+Do Gesture. The four sounds were rated by expert listeners at three time points (pre-testing, immediate post-testing, and 48-hours post-testing) to assess articulatory accuracy and vocal technique. The sung phrase was evaluated at two time points (immediate post-testing and 48-hours post-testing) to assess articulatory accuracy, vocal technique, and expressive artistry.

Participants in all control groups improved their sung production of German vowels, consonants, and phrases after their brief training, as rated by expert listeners. There was evidence of retention of learning for the dependent measures of articulatory accuracy and vocal technique at 48-hours post-training. Similar findings were found across different sounds and dependent measures, which is notable given the randomized presentation of the sung items and the blinded status of the expert raters. The No Gesture and View Gesture groups demonstrated higher perceived articulatory accuracy of the palatal and velar fricatives (consonants) at 48-hours post instruction than for the mixed vowels. For all three groups, vocal technique of the palatal and velar fricatives was rated higher at 48-hours post-instruction than for the mixed vowels.

These findings indicate that the palatal and velar fricatives may be easier for singers to learn and retain than the mixed vowels. Mixed vowels involve the dual coordination of a tongue vowel and a lip vowel, and as such, may be more difficult for singers to replicate initially and to consistently produce over time. The instructor used a two-handed gesture for the mixed vowels (one hand for tongue, one hand for lips), and this gesture may have proved to be more difficult to replicate than the one-handed gesture used with the fricatives, or caused a multitasking scenario that complicated the task. The impact of different parameters (e.g., one- two-hand, proximity of gesture to the articulators, size and synchronization of the gesture) of manual mimicry gestures on the speech mechanism is an area for future investigations. The largest effect of learning and retention was noted for the additional dependent measure of expressive artistry in the sung phrase. This finding may indicate that once participants developed greater confidence with the pronunciation of German sounds, they exhibited a more natural expression in their singing.

The use of gestures during training did not result in greater ratings of articulatory accuracy and vocal technique for individual sounds. Likewise, ratings of articulatory accuracy, vocal technique, and expressive artistry were not enhanced for sung phrases for individuals who used gestures during training. It is important to note that the participants in the View+Do Gesture group rated higher than the other two groups for all metrics, including at baseline. Participants were randomized to groups with consideration of level of education (year of study and undergraduate/graduate status), however the overall skill of the participants was not a factor for group assignment. The View+Do Gesture group was also at or above the mean ratings for the other two groups at their highest post-instruction ratings. These data were found for both individual sounds and phrases. Thus, it is possible that the View+Do Gesture group experienced a ceiling effect. In other words, there potentially was greater room for improvement in the No Gesture and View Gesture Groups and the individuals in the View+Do Gesture group were collectively performing near the peak of their performance for this task in the current paradigm.

Limitations

The study took place in a constrained experimental context, outside of the naturalistic context of the voice lesson or choral rehearsal, where diction is typically taught. An attempt was made to integrate the single phonemes into a musical phrase to better contextualize participants' learning. In a voice lesson or choral rehearsal, however, singers would have additional supports for learning, such as ongoing aural, facial, and/or gestural feedback from an instructor, and multiple sessions to reinforce text pronunciation. Additionally, in a choral rehearsal, singers have the benefit of aural reinforcement from other singers to adjust and refine their vocal articulation. Future investigations may include lengthier training with more opportunities for practice in multiple word and phrase units.

Singers were excluded from the study if they had formal German language instruction, but prior experience with singing in German was not assessed. Singers enter undergraduate and graduate study with variable levels of experience with German diction, based on their prior choral and solo vocal instruction. For those students with more experience singing in German, baseline ratings may have been higher. Future investigations could choose a language that is less common in choral instruction or increase the number of participants to reduce the overall variability of participants (e.g., skill level and prior experience). A repeated measures design, with each participant learning to produce different non-native sounds, some with the use of gestures and some without, could allow for participants to act as their own controls and reduce potential variability.

The raters for this study were experts in their field and were distributed across the United States. The three variables, articulatory accuracy, vocal technique, and expressive artistry were chosen as three dimensions of diction pedagogy. Choral directors and voice teachers may focus on a specific phoneme to assist in phonetic accuracy (articulatory accuracy), to improve vocal sound (vocal technique), or to add expressivity (expressive artistry). Definitions for these three variables were not given to the expert listeners to establish a shared understanding of the terms, however the expert's ratings were descriptively consistent for each participant. In other words, if a trial produced by a participant was rated low for articulatory accuracy by one rater, the other raters also tended to rate the trial in a similar fashion. The raters were instructed to listen under headphones in an area free of distractions and to take breaks as needed, but the equipment (e.g., computers and selected headphones) used to listen to the recorded productions and the environments of the raters were not the same. Future investigations should use standardized procedures for perceptual ratings, if feasible. Likewise, testing the raters hearing would reduce any potential confounds of hearing ability. Lastly, although the randomized perceptual ratings were completed by three expert listeners who were blinded to the purpose of the study, future studies may employ acoustical analyses to examine the effect of training conditions on the acoustic parameters of consonants and vowels.

Implications for Choral Practice

Choral educators must find effective ways to teach an ever-increasing range of languages found in the repertoire, including languages that introduce non-native phonemes (Chris, 2019). German introduces velar and palatal fricatives and mixed vowels to non-native speakers. In her survey of collegiate vocal diction instructors, Mahaney (2006) found that mixed vowels in German and French were among the sounds that were consistently difficult to teach and learn. An embodied approach to teaching diction could exploit the benefits of learning with movement, including reduced cognitive load and improved retention of learning (Goldin-Meadow et al., 2001). The use of gestures to teach diction could also capitalize on the conductor's gestural influence, affecting singer's vocalism (Grady, 2014; Manternach, 2016). Gestural experiences in choir can also be translated into the conductor's gesture as a visual reminder of a felt sensation (Chagnon, 2001; McCoy, 1996).

Manual mimicry gestures are a type of physiological gesture that mirror the specific movements of articulators in forming vowels and consonants (Nafisi, 2013). For vowels, they can reflect the height, advancement, and tension of the tongue. For consonants, they can indicate place, manner, and voicing (Daley & Rusiewicz, 2021). Used in the vocal context, the specificity of these gestures may allow vocal instructors to (a) identify and isolate new or non-native phonemes, (b) migrate a phoneme to a neighbor phoneme for improved resonance or projection (i.e., /I/ to /i/ or /g/ to /k/), or (c) assist to unify articulation among a group of singers, as in the ensemble context (Daley & Rusiewicz, 2021). Although this study did not provide a strong rationale for their use in teaching non-native sounds in German in a constrained context, future studies are needed to assess the impact of these types of gestures in a more naturalistic context, such as a vocal lesson or choral rehearsal.

To replicate the work of Xi et al. (2020), assessing the efficacy of the gesture to the desired phonetic target is essential as not all consonantal sounds are accurately represented in hand gestures. A future study might assess the efficacy of various gestures for the same phonetic target (i.e., more than one gesture for /x/). Another important consideration in gestural teaching is the perception of the effect of gestures on the learner, including when gestures are completed by the learner. Future studies could investigate the qualitative response of the learner in producing non-native sounds for singing with gestures. Lastly, gestures that mimic the spatiotemporal configurations of the vocal apparatus could be adopted for use in conductors' performance in a concert, to mirror how the singers close a consonant, for example. A future study could assess the efficacy of implementing diction-related gestures into conducting performance.

Teaching diction is an everyday task for the choral conductor, voice teacher, and vocal coach. The present investigation invites further exploration into the relationship between gesture and vocalism, and specifically, the intersections between hand gestures and non-native text learning in singing. More research is needed on how gestures might facilitate these outcomes for singers and inform best practices in related disciplines like speech-language pathology and pronunciation training, as well as add to the broader literature base on gesture production.

References

Bahr, D., & Rosenfeld-Johnson, S. (2010). Treatment of children with speech oral placement disorders (OPDs): A paradigm emerges. *Communication Disorders Quarterly, 31*(3), 131–138. https://doi.org/10.1177%2F1525740109350217

Benson, J. S. (2011). *A study of three choral pedagogues and their use of movement in the choral rehearsal* [Doctoral dissertation, The Florida State University]. ProQuest Dissertations and Theses.

Caldwell, T. (1995). *Expressive singing: Dalcroze Eurhythmics for voice.* Englewood Cliffs.

Chagnon, R. D. (2001). *A comparison of five choral directors' use of movement to facilitate learning in rehearsals* [Doctoral dissertation, Arizona State University]. ProQuest Dissertations and Theses.

Chris, G. E. (2019). Diction pedagogy: *A survey of New England high school choral directors* [Masters dissertation, University of Maryland]. ProQuest Dissertations and Theses.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development, 84*(6), 1863–1871. https://doi.org/10.1111/cdev.12097

Crosby, A. (2008). Dalcroze's Eurhythmic techniques for choral rehearsal: Moving to "O Magnum Mysterium". *Choral Journal, 48*(11), 30–41. http://www.jstor.org/stable/23556909.

Daley, C. (2013). *Moved to learn: Dalcroze applications to choral pedagogy and practice* [Doctoral dissertation, University of Toronto]. ProQuest Dissertations and Theses.

Daley, C. & Rusiewicz, H. L. (2021). Manual mimicry in the teaching and learning of phonemes for singing. *Journal of Singing, 77*(4), 521–533. https://www.nats.org/_Library/JOS_On_Point/JOS_077_04_2021_521.pdf

Daugherty, J. F., & Brunkan, M. C. (2013). Monkey see, monkey do? The effect of nonverbal conductor lip rounding on visual and acoustic measures of singers' lip postures. *Journal of Research in Music Education, 60*(4), 345–362. https://doi.org/10.1177/0022429412464039.

DeThorne, L. S., Johnson, C. J., Walder, L., & Mahurin-Smith, J. (2009). When "Simon Says" doesn't work: Alternatives to imitation for facilitating early speech development. *American Journal of Speech-Language Pathology, 18*(2), 133–145. https://doi.org/10.1044/1058-0360(2008/07-0090).

Dutoit, C-L. (1971). *Music movement therapy.* Dalcroze Society.

Ehmann, W. (1968). *Choral directing.* Augsburg Press.

Galván, J. (2008). Movement and the choral rehearsal – part one: The use of movement in the choral rehearsal. In M. Holt and J. Jordan (Eds.), *The school choral program: Philosophy, planning, organizing, and teaching* (pp. 515–532). GIA Publications.

Grady, M. L. (2014). *The effects of nonverbal behaviors exhibited by multiple conductors on the timbre, intonation, and perceptions of three university choirs, and assessed relationships between time spent in selected conductor behaviors and analysis of the choirs' performances* [Doctoral dissertation, University of Kansas]. ProQuest Dissertations and Theses.

Goldin-Meadow S., Nusbaum H., Kelly S. D., & Wagner S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Gentilucci, M., Benuzzi, F., Gangitano, M., & Grimaldi, S. (2001). Grasp with hand and mouth: Kinematic study on healthy subjects. *Journal of Neurophysiology, 86*(4), 1685–1699. https://doi.org/10.1152/jn.2001.86.4.1685

Hammer, D. W. (Host). (2006). *Treatment strategies for childhood apraxia of speech* [DVD]. CASANA.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience, 17,* 1123–1128. https://doi.org/10.1046/j.1460-9568.2003.02530.x

Hibbard, T. T. (1994). *The use of movement as an instructional technique in choral rehearsals* [Doctoral Dissertation: University of Oregon]. ProQuest Dissertations and Theses.

Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research, 57*(6), 2090–2101. https://doi.org/10.1044/2014_JSLHR-S-14-0049

Jaques-Dalcroze, É. (1921). *Rhythm, music and education.* (H. F. Rubenstein, Trans.). The Dalcroze Society. https://doi.org/10.1177%2F002205742109401204.

Koshut, C., Rusiewicz, H. L., Leech, S., Perciavalle, G., Cummings, A., & Walsh, G. (2016, November). Manual mimicry gestures in the treatment of CAS: A case study of two young children. Poster presented at the American Speech-Language-Hearing Association Annual Convention, Philadelphia, PA.

Krause, J. C., Pelley-Lopez, K. A., & Tessler, M. P. (2011). A method for transcribing the manual components of Cued Speech. *Speech Communication, 53*(3), 379–389. https://doi.org/10.1016/j.specom.2010.11.001

Liao, M.-Y., & Davidson, J. W. (2016). The effects of gesture and movement training on the intonation of children's singing in vocal warm-up sessions. *International Journal of Music Education, 34*(1), 4–18. https://doi.org/10.1177/0255761415614798.

Mahaney, C. L. (2006). *Diction for singers: A comprehensive assessment of books and sources* [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations and Theses.

Manganello, R. B. (2011). *Movement in the middle school choral rehearsal* [Doctoral dissertation, Boston University]. ProQuest Dissertations and Theses.

Manternach, J. N. (2016). Effects of varied conductor prep movements on singer muscle engagement and voicing behaviors. *Psychology of Music, 44*(3), 574–586. https://doi.org/10.1177/0305735615580357

McCoy, C. W. (1986). *The effects of movement as a rehearsal technique on performance, meter discrimination ability, and attitude of members of high school choral ensembles* [Doctoral dissertation, University of Iowa]. Proquest Dissertations and Theses.

Nafisi, J. (2013). Gesture and body-movement as teaching and learning tools in the classical voice lesson: A survey into current practice. *British Journal of Music Education, 30,* 347–367. http://dx.doi.org/10.1017/S0265051712000551.

Oney, T. N. (2017). Dalcroze eurhythmics: An application to voice pedagogy. *Journal of Singing, 74*(1), 37–43. https://link.gale.com/apps/doc/A624031673/AONE?u=pl9014&sid=googleScholar&xid=89959510

Pierce, A. (2007). *Deepening musical performance through movement: The theory and practice of embodied interpretation.* Indiana University Press. https://doi.org/10.1525/mts.2012.34.2.137.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*(5), 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0

Rusiewicz, H. L. (2020). Show of hands: Using manual mimicry in speech and voicetherapy. *Voice Foundation Newsletter, 25*(1), 9–11. https://voicefoundation.org/wp-content/uploads/2020/04/Spring-Newsletter-2020.pdf

Rusiewicz, H. L., & Rivera, J. L. (2017). The effect of hand gesture cues within the treatment of /r/ for a college-aged adult with persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology, 26*(4), 1236–1243. https://doi.org/10.1044/2017_AJSLP-15-0172

Schreiber, J., & Asner-Self, K. (2011). *Educational research: The interrelationship of questions, sampling, design, and analysis.* Wiley.

Shenenberger, M. (2008). Using Dalcroze Eurhythmics in the choral rehearsal. In M. Holt and J. Jordan (Eds.), *The school choral program: Philosophy, planning, organizing, and teaching* (pp. 533–560). GIA Publications.

Sieck, S. (2013). Factors in teaching foreign language choral music to students. *Choral Journal, 54*(3), 24–35. http://www.jstor.org/stable/23561363

Square, P. A., Namasivayam, A. K., Bose, A., Goshulak, D., & Hayden, D. (2014). Multi-sensory treatment for children with developmental motor speech disorders. *International Journal of Language & Communication Disorders, 49*(5), 527–542. https://doi.org/10.1111/1460-6984.12083.

Smith, A., McFarland, D. H., & Weber, C.M. (1986). Interactions between speech and finger movements: An exploration of dynamic pattern perspective. *Journal of Speech and Hearing Research, 29*(4), 471–480. https://doi.org/10.1044/jshr.2904.471

Wis, R. M. (1993). *Gesture and body movement as physical metaphor to facilitate learning and to enhance musical experience in the choral rehearsal* [Doctoral dissertation, Northwestern University]. ProQuest Dissertations and Theses.

Wis, R. M. (1999). Physical metaphor in the choral rehearsal: A gesture-based approach to developing vocal skill and musical understanding. *Choral Journal, 40*(3), 25–33. http://www.jstor.org/stable/23553144

Xi, X., Li, P., Baills, F., & Prieto, P. (2020). Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research, 63*(11), 3571–3585. https://doi.org/10.1044/2020_JSLHR-20-00084