Seeing voices: the role of multimodal cues in vocal learning

Varkevisser, J.M.

Chapter 3

Multimodality during live tutoring is relevant for vocal learning in zebra finches

Judith Varkevisser, Ezequiel Mendoza, Ralph Simon, Maëva Manet, Wouter Halfwerk, Constance Scharff & Katharina Riebel

This chapter is published in Animal Behaviour:
https://doi.org/10.1016/j.anbehav.2022.03.013

Abstract

In many songbird species, young birds learn their song from adult conspecifics. Like much animal communication, birdsong is multimodal: singing is accompanied by beak and body movements. We hypothesized that these visual cues could enhance vocal learning thus partly explaining the reduced learning from unimodal audio playbacks compared to multimodal live social tutoring observed in many birdsong studies. To test this, juvenile zebra finches, *Taeniopygia guttata*, were tutored in a yoked design where replicate tutoring groups of three male–female dyads were exposed to the same live tutor simultaneously in three different ways. (1) Tutees were housed with the tutor in a central compartment; hence they could hear, see and interact with their tutor ('live'). (2) Tutees placed in one of two adjacent compartments could hear but not see the same tutor from behind a black loudspeaker cloth ('audio-only'). (3) Tutees could likewise hear the tutor through loudspeaker cloth but could also see the tutor through a one-way mirror ('audiovisual'). Comparisons of subadult and adult song showed more changes in the audio-only than in the audiovisual or live tutored tutees, suggesting the audio-only group's song development was delayed. According to (blinded) human observer similarity scoring, the audio-only tutees' singing was least similar and the live tutees' singing most similar to their tutor's singing, while the audiovisual tutees showed an intermediate level of similarity, but the between-treatment differences in similarity were not significant. Conversely, the audio-only group showed the highest similarity values with their father's song, which they only heard before the experimental tutoring. Given that the quantity and quality of the tutor song input were the same across treatments within tutoring groups, the results support the hypothesis that visual in addition to auditory exposure to a tutor can affect the timing and possibly also the amount of vocal learning.

Introduction

Songbirds are well-known vocal learners (Catchpole & Slater, 2003; Doupe & Kuhl, 1999). For the majority of species studied, learning from conspecific social tutors is crucial to develop fully functional species-specific song (Catchpole & Slater, 2003). In some species, hearing adult song from playback provides birds with sufficient input to develop their song, but in a considerable number of the species studied, playing tutor song back via loudspeakers (so-called tape tutoring) resulted in lower tutor song copying accuracy than from a live conspecific as tutor (reviewed in Baptista & Gaunt, 1997; Soma, 2011).

The zebra finch, *Taeniopygia guttata*, one of the commonest animal models for studies on vocal learning (Griffith & Buchanan, 2010; Mello, 2014), is a species that learns better from a live conspecific than from passive tutor song exposure (reviewed in Derégnaucourt, 2011; Slater, Eales & Clayton, 1988). It is usually concluded that this difference is due to a lack of social interaction with the tutor in the tape–tutor condition (Chen et al., 2016; Derégnaucourt et al., 2013; Eales, 1989), but there are more dimensions that differ between live and tape tutoring (for discussion see Nelson, 1997). For example, live tutoring offers multimodal tutor exposure, as tutees can hear and see their tutor, while tape tutoring only offers auditory, unimodal song exposure. The majority of studies comparing social versus nonsocial tutoring used live and tape tutors and were thus also comparing multi- versus unimodal tutoring, meaning that these issues have been confounded in previous studies (Varkevisser et al., in preparation).

Several lines of evidence suggest that multi- rather than unimodal presentation of song might increase the salience of the stimulus regardless of a social component. Zebra finch song, like birdsong in general and many signals in animal communication (Halfwerk et al., 2019; Higham & Hebets, 2013; Partan & Marler, 1999), is a multimodal signal, as auditory song production is accompanied by visual cues such as beak and body movements (Goller, Mallinckrodt, & Torti, 2004; Ullrich, Norton, & Scharff, 2016; Williams, 2001). Visual stimulation together with an auditory stimulus can facilitate learning of an auditory stimulus as has been demonstrated in domestic chickens, *Gallus g. domesticus,* in the context of filial imprinting (van Kampen & Bolhuis, 1991; van Kampen & Bolhuis, 1993) and in young nightingales, *Luscinia megarhynchos*, that learned songs from audio playbacks combined with light flashes better than those presented as audio-only playbacks (Hultsch et al., 1999). Zebra finches also seem to attend to visual information during song learning. First, tutees show beak and dance movements that are highly similar to the individual-specific movements produced by their tutors, while they differ from those of unfamiliar males (Williams, 2001). Second, the visual appearance of a bird plays a role in tutor choice (Clayton, 1988; Mann et al., 1991; Mann & Slater, 1995). Moreover, improved song learning was found when zebra finch tutees received visual stimulation contingent on their song production (Carouso-Peck & Goldstein, 2019). There are also a number of studies that used static, nonmoving taxidermic mounts as a visual stimulus, but found no improved learning in birds briefly seeing this visual stimulus either right before, during or after song presentation (Bolhuis et al., 1999; Houx & ten Cate, 1999). Overall, it thus seems worth investigating more systematically whether the visual cues associated with song production facilitate song learning.

Our aim in the present study was to create an experimental situation that would allow us to compare tutees receiving either multi- or unimodal exposure to the same live tutor to see whether the additional visual cues, independent of visual interaction, would facilitate song learning in zebra finches. Tutees in our experiment could all hear and vocally interact with the same tutor, since in each set-up a tutor and three male tutees were kept in the same cage which was separated into three compartments (Fig. 1). In the middle compartment, a tutee was housed together with a tutor so that they could interact visually, acoustically and physically. This represents a situation in which normally high levels of song copying occur (Derégnaucourt, 2011). In one of the adjacent compartments, a tutee was housed and separated from the tutor in the middle compartment with acoustically transparent cloth, but the tutee could see the tutor through a small one-way mirror. This provided tutees in this treatment group with multimodal tutor exposure, but it prevented visual tutor–tutee interaction, as the tutees could see the tutor, but the tutor could not see them. In the other side compartment, a tutee was also separated from the tutor in the central compartment with loudspeaker cloth, but without a one-way mirror. This group could hear, but not see the tutor, thus receiving only unimodal audio exposure to the tutor. We provided all male tutees with a juvenile female as a social companion. This was to prevent social isolation, which was a confound in previous studies comparing live tutoring and audio-only song exposure. In addition, the experiment was run simultaneously in two locations with birds from two different breeding colonies using the same paradigm, but with slightly different technical realization on location.

Birds were tutored experimentally between 35 and 65 days posthatching (DPH). The peak of the sensitive phase for song learning in zebra finches is 35–65 DPH, but sensory learning starts as early as 20–25 days DPH (reviewed in Gobes et al., 2017). Zebra finches normally primarily learn the motif of the tutor they can socially interact with between 35 and 65 DPH, but if learning conditions are suboptimal at that time, they might incorporate syllables heard before that age in their song (Böhner, 1986; Eales, 1989; Jones, ten Cate, & Slater, 1996; reviewed in Gobes et al., 2017). As our tutoring methods had not been tried before, we assessed the similarity of tutee's songs with both their father (the first encountered model) and the tutors they encountered during the experimental tutoring phase (35–65 DPH). Song development entails not only learning to sing specific syllables but also the ordering and timing of syllables, as well as the stereotypy of song delivery, all aspects of song that can differ substantially between individual male zebra finches (Helekar et al., 2000; Holveck et al., 2008; Hyland Bruno & Tchernichovski, 2019; Scharff & Nottebohm, 1991). We thus assessed a number of these parameters in addition to song similarity scores. If visual cues play a role in song learning, then song learning in the birds with multimodal tutor exposure should be better than in the birds with unimodal tutor exposure. However, if visual cues do not play a role, and live tutors facilitate song learning mainly because of their 'sociality', then equal song copying is expected in the birds with audiovisual and audio-only tutor exposure.

Methods

Subjects and housing

Subjects for this study were 13 adult male tutors and 44 male and 90 female juvenile domesticated wild-type zebra finches from two different breeding colonies. One colony was located at Leiden University (contribution to experiment: N = 9 adult male tutors and 27 male and 54 female juveniles) and the other at the Free University Berlin (contribution to experiment: N = 4 adult male tutors and 18 male and 36 female juveniles). In Leiden, subjects were bred in several rounds: 12 male and 24 female juveniles (four tutor groups, see below) hatched in March 2017, three male and six female juveniles (one tutor group) hatched in August 2017 and 12 male and 24 female juveniles (four tutor groups) hatched in November 2017. In Berlin, breeding in the colony was continuous and juveniles for the first tutor group hatched in January 2017 and for the last tutor group in November 2018. All young birds were the offspring of established breeding pairs and were housed in breeding cages (Leiden: 80 x 40 cm and 41 cm high; Berlin: 180 x 42 cm and 33 cm high) until chicks were 35 DPH (age of chicks was determined as the median hatching day of all chicks within the nest). In Berlin, the father remained in the same cage but was separated from the juveniles at 23 DPH by a wire mesh covered in paper allowing vocal communication but not visual or physical contact. In Leiden, the father remained in the breeding cage with the tutees until 35 DPH, at which age (mean ± SD = 35.3 ± 1.2 days) young males and females were assigned to tutor groups that were exposed to the song of the same unrelated (coefficient of relation < 0.125) adult male (the 'tutor'). The adult tutors had been housed in same-sex aviaries prior to the experiment (Leiden, age range at the start of the experiment 120–806 days, mean ± SD: 509 ± 300; Berlin, range 1919–2945 days, mean ± SD: 2482 ± 424). Next to one adult tutor, each tutor group consisted of three male and three female tutees for the tutoring period and an additional three female companions that were cohoused with the tutees after the tutoring phase. For each tutoring group, whenever possible, we chose three males from the same nest. This was possible in two tutor groups in Leiden and all six tutor groups in Berlin (in the other groups in Leiden, two siblings were spread over the different treatment conditions across tutor groups). Each tutor group was then complemented with three female tutees always chosen from a different brood as females develop a preference for the song they hear early in life, and might guide male song development to a certain degree based on this preference (Jones & Slater, 1993). Therefore, whenever possible, these three females originated from the same brood and had heard the same tutor song early in life (N = 8 tutor groups). In the other groups, two siblings were spread over the different treatment conditions across tutoring groups. Tutors, together with one male and female tutee (live condition), were placed in the middle compartment of an experimental cage consisting of three compartments (Fig. 1; Leiden: 150 x 40 cm and 50 cm high, located in a larger bird room where other birds were audible, but not visible; Berlin: 90 x 33 cm and 42 cm high, located in a sound-attenuated chamber 91 x 150 cm and 78 cm high).

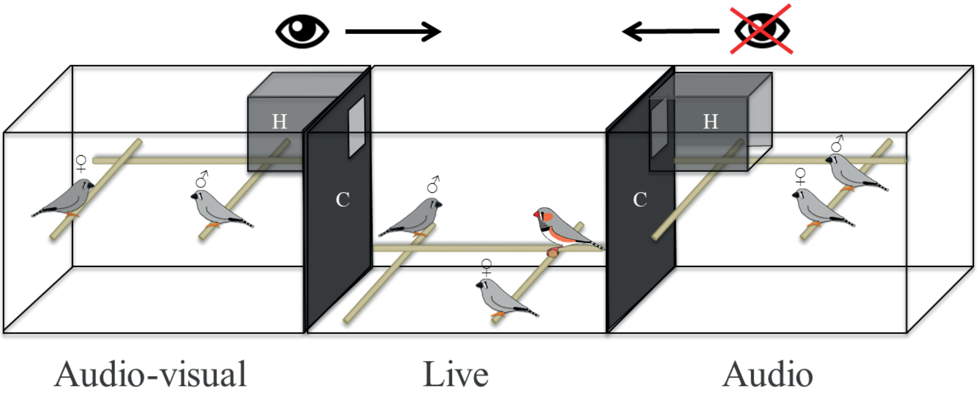


Figure 1. Schematic front view of the experimental set-up in which the song tutoring took place. C = separation made from loudspeaker cloth, H = observation hut. Eye and crossed-out eye symbol show through which one-way mirror the tutees could and could not see the tutor, respectively.

A male/female tutee dyad was also placed into each of the compartments to the left and right. The compartments were separated from each other by opaque (black), but acoustically transparent loudspeaker cloth. One of the two side compartments was assigned to the audio-only condition: tutees could only hear (though the loudspeaker cloth) but not see the tutor in the central compartment. The other compartment was designated to the audiovisual treatment: as in the audio-only treatment the tutor could be heard via the cloth, but in addition, the tutee could watch the tutor via a small one-way mirror (5x8 cm), when perched on the upper central perch. The audio-only compartment had an identical mirror, but it was rendered opaque (by gluing together two one-way mirrors with a piece of white paper in between). The assignment of audio-only or audiovisual compartment to the left or right compartment was balanced across set-ups. As the one-way mirrors only function if there is a difference in light intensity between the two sides of the mirror, we built 'observation huts', consisting of black painted wood and opaque loudspeaker cloth (Leiden) or black plastic (Berlin), around the one-way mirrors, such that a bird perching in front of the mirror would find itself in a dimly illuminated space. Perches were arranged in such a way that birds in the side compartments could easily reach the one-way mirrors, while the perches in the middle cage were positioned lower so that it was more difficult to reach the mirrors. In Leiden, we noticed that some of the tutors were nevertheless flying up to the one-way mirrors and clinging to the cloth next to it. To avoid the birds in the middle compartments coming close to the one-way mirrors and seeing their own appearance, we glued transparent plastic hemispheres (10 cm diameter) around the one-way mirrors. We characterized sound propagation through the cages and the loudspeaker cloth as follows. A loudspeaker playing pure tones between 200 Hz and 18 kHz in 200 Hz steps was placed at different positions in the central compartment. We had a microphone in the same compartment and one microphone at different positions in the neighbouring side compartment. We found that at the positions where we made the recordings in the side compartment, the frequency response was similar to that in the central compartment but was attenuated (probably due to spreading loss and atmospheric attenuation) between 3dB for lower frequencies and up to 10 dB at higher frequencies (see Fig. A1). A measurement with a microphone installed in the 'observation hut' showed that the one-way mirror, the hut and the position of the hut probably influenced the frequency response. There was almost no effect for frequencies below 4 kHz, but between 6 kHz and 10 kHz a higher attenuation compared to the other location was measured, indicating that the song of a tutor gets somewhat filtered when tutees listen in the observation hut. Note, however, that the audio-only conditions were the same in the two side cages and that the audio-only and audiovisual tutees thus had the same audio conditions.

When tutees reached 65 DPH, tutors were removed from the experimental set-up. Female tutees were also moved to form all-female groups in aviaries as their development was followed separately. The females were moved to prevent them from learning from the male tutees that start singing adult-like song around this time (Immelmann, 1969). All male tutees that had remained in the set-up received a new female companion each; these were of the same age and from the same breeding round. These females had been housed in sets of three with an adult tutor pair between 35 and 65 DPH to be then moved into the experimental set-up to replace the female tutees. For every tutor group, the three new female companions had been housed with the same adult tutor pair before 65 DPH. In addition to this, in Leiden, all the mirrors were now covered with cardboard. Male tutees remained in the experimental cages (in Berlin and Leiden) except for a brief recording session at 65 days and then until their song was recorded after 100 DPH. Then, males were moved into large all-male group aviaries. See Fig. 2 for a timeline of the experimental procedure.

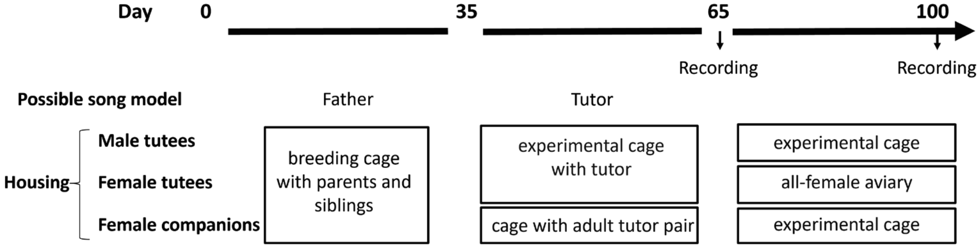


Figure 2. Time line of the experimental procedure. Note that in Berlin, the father was removed from the breeding cage at around 23 days posthatching.

Throughout, birds were housed on a 13.5/10.5 h (Leiden) or 12/12 h (Berlin) light/dark cycle, at 20–25 °C and 45–65% humidity. Within tutor groups, the tutees from the different treatments were housed in the same bird room (Leiden) or in the same soundproof box (Berlin); thus, within groups tutees always experienced the same temperature and humidity conditions. Birds had ad libitum access to a commercial tropical seed mixture (Leiden: Beyers, Belgium; Berlin: Teurlings, Germany), cuttlebone, grit and drinking water. This diet was supplemented twice a week with hardboiled eggs, germinated tropical seeds, vegetables and fruit.

In total, 15 tutor groups were raised in the experimental set-ups: nine in Leiden and six in Berlin. One tutor group (Berlin) consisted only of three male tutees and one tutor, because there were no same-age females available at the start of song tutoring.

Usage of observation huts by tutees

To investigate whether the tutees in the side compartments (audiovisual and audio-only treatments) were using the observation huts, for the first four tutor groups raised in Leiden, we filmed the tutees in the two treatment groups (GoPro Hero 3+ camera, San Mateo, CA, U.S.A.) two mornings a week throughout the tutoring period. For each group, we analysed 2–4 h of video recorded during one or two mornings (recordings started between 0900 and 1000 hours) in the second week of tutoring, because we expected the birds to be familiar with the experimental set-up by then. For these videos, every 30 s the positions (inside or outside the hut) of the male and female tutee were scored separately. The proportion of observations during which the male and female tutee of the audiovisual and audio-only treatments were inside the observation huts was then calculated. Zebra finches have a wide visual field (each eye around 170° in the horizontal plane; Bischof, 1988) and can look through the window while their body or head is not directed towards it. From our video recordings, we could therefore only assess whether the birds were in the hut, but not when the tutees were looking through the one-way mirror. However, the proportion of observations where the tutees were inside the huts does give an indication of the total time for which the tutees could have watched the tutor.

Song recording

Both in Berlin and in Leiden, for all song recordings (fathers, tutors and male tutees), birds were moved the afternoon before a recording day to acclimate in a sound-attenuated chamber (Leiden: 125 x 300 cm and 240 cm high; Berlin: 60 x 60 cm and 73 cm high) that contained a small recording cage (Leiden: 76 x 45 cm and 45 cm high; Berlin: 46 x 29 cm and 48 cm high), and then recorded continuously the next morning with a microphone suspended from above the cage (Leiden: Sennheiser MKH40 microphone, Wedemark, Germany connected to a TASCAM DR-100MKiii recorder (TEAC Corp., Los Angeles, CA, U.S.A.), sampling at 96 kHz, 16 bits; Berlin: Earthworks SRO microphone (Milford, NH, U.S.A.) connected to a PC using the recording function of the Sound Analysis Pro software (SAP; Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000), sampling at 44.1 kHz, 16 bits). The tutees' fathers were recorded after successful breeding and after their offspring (the tutees) had been moved into the experimental set-up. All song tutors had been recorded prior to moving them into the experimental set-ups. All tutees were recorded twice: once at 65 DPH (mean ± SE: 66 ± 1.4 days) and once as young adults after 100 DPH (mean ± SE: 130 ± 11.3 days). The recording at 65 DPH took place while tutee song was still developing, but when most syllables of the final song are usually present (Slater, Eales, & Clayton, 1988). At 65 DPH, in Leiden, male and female tutees were placed together in the recording cage (as in Leiden, many birds had to be recorded around the same time, and males are more likely to sing the first day if housed with a female). In Berlin, young male tutees were recorded without their female companions. At >100 DPH, in both Leiden and Berlin male tutees were recorded while they were temporarily housed individually in the recording cage. After tutoring had started, it turned out that one of the audiovisual groups in Berlin had two females instead of a male and female tutee due to misidentification. From the remaining 44 male tutees, 33 birds could be recorded at 65 DPH. At >100 DPH, 41 birds produced more than 20 songs. Only song of these birds was used in the song analysis. The father from one of the tutor groups could not be recorded, so the song of the tutees from this group was only compared to the tutor song.

Song analysis

Comparison of the relative success of tutoring methods has been hampered by the many different analysis methods used in zebra finch song research. Studies up until 1999, including many relevant for this study, mainly used visual inspection of spectrograms by human observers (e.g. Bolhuis et al., 1999; Eales, 1989; Houx & ten Cate, 1999) to assess the similarity between the tutees' song and possible model songs. Zebra finch song studies since 2000 have regularly used automated digital measurement methods, such as SAP (Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000) and Luscinia (Lachlan et al., 2010). To make our results comparable to these previous studies and to our recent study employing a video tutoring method (Varkevisser et al., 2021), we used all three of these methods (human observers, SAP and Luscinia) in this study. To aid cross-study comparison and interpretation, we also assessed the correlation between the three methods for our complete data set and for the subset of live tutored tutees separately, as previous method comparisons have only taken into account similarity assessment for the song of tutors and their live tutored tutees (Lachlan et al., 2010; Tchernichovski et al., 2000), a condition known to lead to high tutor song similarity (Derégnaucourt, 2011). Next to tutor–tutee similarity, an additional set of structural dimensions (see Table 1) were analysed, again using the same parameters as Varkevisser et al. (2021).

Song and motif selection

Following Sossinka and Boehner (1980), we defined a song syllable as a unit of sound separated from another sound by a silent interval of at least 5 ms and a motif as an individual-specific sequence of syllables. The term 'song' refers to a series of motifs separated from other sounds by more than 2 s of silence or a series of motifs preceded by multiple introductory notes (Sossinka & Böhner, 1980). Selection of songs and sound editing were conducted by visual inspection of combined spectrograms and amplitude waveform displays with Praat software (v. 6.0.19, Boersma & Weenink, 2008; spectrogram settings: fast Fourier transformations with 1000 time and 250 frequency steps, 0.005 s window length, dynamic range 55 dB, Gaussian window). The spectrograms of all audio-recording sessions were visually screened and digitally parsed into songs to be saved as separate audio files using one folder per recording session (65 DPH and >100 DPH) of each male. From each folder, 20 songs were picked at random (with custom-written software by Niklas J. Tralles) and from these songs, the motif encountered most often was selected and termed 'the typical motif'. In addition, we selected the motif with the highest number of different syllables (the 'full motif') from the adult (>100 DPH) recordings. We also selected a random subset of 10 motifs by first randomly selecting 10 of the 20 songs and then selecting one motif from each of the 10 songs with a random number generator (http://www.random.org). These motifs were digitally cut from the recordings, band stop filtered from 0 to 420 Hz, and amplitude normalized with the 'scale peak' function (all with Praat Software, v. 6.0.19). Introductory notes that were part of each motif occurrence were kept, but all additional introductory notes were cut off before further analysing these 10 motifs with the SAP and Luscinia software (see below).

Song structure and performance

For the typical and full motifs, one of the authors (J.V.) visually inspected the spectrograms and labelled all different syllables with different letters (see Fig. 3, using the Praat software and settings as described above). For each tutee, we counted the syllables in the typical motif as well as the number of unique syllables in the full motif. We then calculated sequence linearity and consistency (Scharff & Nottebohm, 1991), by assessing the different transitions between the syllables for the sample of 20 randomly selected undirected crystallized songs of each male. Sequence linearity is the total number of different syllables divided by the number of different transitions between syllables and higher scores indicate a more stereotyped syllable order of the different motifs within a song. We determined sequence consistency by first noting all transitions and then determining the most frequent ('typical') transition for each syllable in the 20 songs. We divided the total number of occurrences of typical transitions by the total number of transitions encountered in the 20 randomly selected songs. As with sequence linearity, higher scores indicate more stereotyped songs. As an additional song stereotypy measure, we conducted within-subject comparisons by comparing each of the 10 randomly selected motifs with each other in SAP and Luscinia, using the same settings as for the similarity scores (see below). We continued analyses with the median SAP similarity score and the median 1-d Luscinia distance value, so that for both scores higher scores indicate a higher stereotypy. These values are referred to as the 'SAP stereotypy score' and 'Luscinia stereotypy score'.

Similarity to tutors' and fathers' song

Human observer similarity scoring

We followed the procedures from Houx and ten Cate (1999a) for the human observer similarity scoring, but with the difference that we chose syllables (see above and Fig. 3) instead of elements as units. We opted for syllables, because based on the literature we expected poor tutor song copying and isolate-like song in the experimental groups (Eales, 1989; Price, 1979) which can make determining element boundaries problematic because of the variance in the frequency patterns being higher than in normal song (Price, 1979). Identifying syllable boundaries is less of a problem, as syllables can be recognized by the short silent intervals that delineate them. Three observers (Ph.D. candidates from the Leiden lab), blinded with respect to birds' IDs and treatments (and with some but varying experience with spectrogram analyses), independently assessed syllable similarity between the song models and the tutees. Each observer scored the complete set of spectrograms (see Fig. 3 and Fig. A2) while working through a PowerPoint presentation on a personal computer. Each new slide presented a new set of spectrograms: one of a tutee's full motif (labelled 'tutee') on top and directly underneath a second spectrogram labelled 'model' (which unknown to the observer was from either the tutor or the father of the tutee). The observers had to compare each tutee with two models: the tutor and the father. They were asked to indicate for each tutee syllable the most similar syllable of the model by paying attention to a syllable's frequency pattern, duration, overall shape and sequential position and to score the degree of similarity on a scale from 0 to 3 (0 = 'no similarity at all', 1 = 'slight similarity', 2 = 'moderate similarity' and 3 = 'very strong similarity'). Interobserver reliability was calculated after normalizing individual observer scores by subtracting the mean of the observer's scores from each score and then dividing it by the standard deviation. Using these normalized similarity values as the response variable and tutee ID as a factor, we then conducted a one-way ANOVA to calculate repeatability (Lessells & Boag, 1987) which was high for all comparisons (Tutor–Tutee: $F_{2,38}$= 12.92, $P < 0.01$, r ± SE = 0.80 ± 0.05; Tutee–Tutor: $F_{2,38}$ = 10.18, $P < 0.01$, r ± SE = 0.75 ± 0.06; Father–Tutee: $F_{2,38}$= 7.07, $P < 0.01$, r ± SE = 0.67 ± 0.07; Tutee–Father: $F_{2,38}$ = 5.17, $P < 0.01$, r ± SE = 0.58 ± 0.08). While this indicates relatively high agreement, it also shows that observers differed; this was mainly because observers varied in how strict they were regarding the syllables with lower similarity to the tutor syllables. To capture this best, for further analysis, we decided to combine the individual scores of all observers by first summing them and then dividing them by the maximum score a bird could have received from three observers. This resulted in one similarity score for a particular model–tutee comparison, while correcting for between-individual differences in the number of syllables in the motif, thus providing a measure that combines the proportion of syllables copied with a weighting of their similarity.

Tutees can differ in the proportion of copied versus improvised syllables (e.g. Tchernichovski et al., 2021; Williams, 1990), which means that the direction of comparison can affect the scores for syllable sharing. For example, if a tutee has accurately copied the syllables ABC from a tutor with the motif ABCDE, this tutee would score higher on the tutee–model comparison (ABC = 100% of the tutee's syllables are shared with the tutor) than on the model–tutee comparison (ABC = only 60% of the tutee's syllables were copied from ABCDE). Conversely, another tutee singing motif ABCDEFG, where ABCDE are copied and F and G improvised, would score higher on the model–tutee comparison (all tutor syllables, i.e. 100%, copied) than on the tutee–model comparison (only ABCDE, but not F and G are shared, thus this yields only 71%). As the types and direction of effects (i.e. poor copying or improvisation) for this new type of tutoring could not be predicted from the literature, we assessed song similarities between tutors and tutees to capture both overlap and level of improvisation by looking at (1) the proportion and similarity of the model's syllables that the tutee has copied ('similarity score model–tutee') and (2) the proportion and similarity of the tutee's syllables that are shared with the model ('similarity score tutee–model'). For the similarity score model–tutee, for each model syllable, we noted the ID and similarity score of the tutee syllable that received the highest score and then summed these scores. If two or more tutee syllables received the same score, we noted this score once, but for the similarity score tutee–model, the scores for all tutee syllables were included (see Table 1 for full formula).

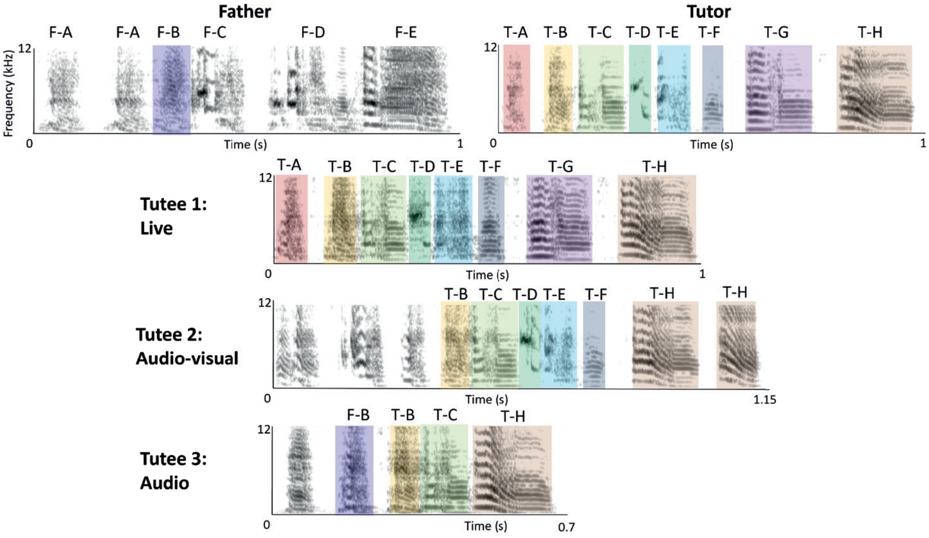


Figure 3. Spectrograms of the songs of the father, tutor and three male tutees from one tutor

group (full motif). Syllables are labelled by two letters indicating the song model (F, T) combined with a second letter indicating the syllable identity. Human observers scored the similarity between two syllables on a scale from 0 to 3. Syllables shaded in the same colour were judged as the most similar syllable by at least two observers. Note, however, that this binary categorization of shared/nonshared syllables is for illustration purposes; it does not reflect the continuous scoring of similarity in the analyses which combined the scores of all three observers and corrected the score for total motif length (see parameters 'human observer similarity score model-tutee' and 'human observer similarity score tutee-model' in Table 1).

Automated similarity scoring (SAP and Luscinia)

For the automatic, quantitative song comparisons, we used Luscinia (version 2.16.10.29.01) and SAP (MxN comparison, default settings tuned for zebra finch, per tutor–tutee pair amplitude thresholds were adjusted for correct syllable segmentation, version 2011.104) to compare each of the 10 randomly selected tutee motifs to each of the 10 randomly selected father's and tutor's motifs. For each possible comparison, we assessed the asymmetric time courses SAP similarity score for the model to tutee and tutee to model comparisons (SAP similarity score model–tutee and tutee–model). Both values indicate the percentage of sounds of one song (tutee or model) observed in the other. As the quantitative measure of similarity, for each individual we used the median value of the scores resulting from the comparisons between the 10 randomly selected motifs per individual. We used the median, because our sample size was too small to create a good-fitting model for all similarity scores and because the SAP scores did not follow a normal distribution and were bound between 0 and 100. For the acoustic distance calculations between model–tutee pairs in Luscinia, which uses a dynamic time warping (DTW) algorithm, we selected the acoustic features 'mean frequency', 'fundamental frequency' and 'fundamental frequency change' (following Lachlan, van Heijningen, ter Haar, & ten Cate, 2016) and we added the feature 'time', which allows for flexible comparison of motifs of different duration. The DTW analysis results in one distance measure (d) between 0 and 1 for each possible motif pair. In contrast to the human observer and SAP similarity scores, this measure is symmetric, meaning that it is the same for the model to tutee and the tutee to model comparisons. In Luscinia, a smaller distance value means a higher similarity, but because the other two methods express higher similarities with higher values, we simply calculated the inverse of the median distance score (1 - d, henceforth 'Luscinia similarity score'), to aid comparison across all three methods.

Structural changes in the typical motif between 65 and 100 DPH

For each tutee, the syllables of the typical motif produced at 65 DPH were compared with those at the second recording at > 100 DPH by visually inspecting the spectrograms to assess the number of changes (i.e. syllable deletions, repetitions or insertions). For this analysis, spectrograms were saved under a code number and then inspected by one of the authors (J.V.) without knowing a tutee's treatment group or whether syllables were improvised or copied from model song.

Table 1. All song analysis parameters and the formulas and sample sizes for calculation.

Parameter	Definition	Sample per bird
Typical motif	Most frequently produced motif	20 random songs
Full motif	Motif with highest # different syllables in bird's repertoire	20 random songs
Total number of syllables	# syllables in a tutee's typical motif	Typical motif
Number of unique syllables	# unique syllables in a tutee's full motif	Full motif
Linearity	(# different syllables/song)/(#transition types/song)	20 random songs
Consistency	(total # typical transitions)/(total # of transitions)	20 random songs
Human observer similarity score model-tutee	(Σ similarity scores (all observers) for all model syllables)/(# model syllables*3 (max score)*# observers)	Full motif
Human observer similarity score tutee-model	(Σ similarity scores (all observers) for all tutee syllables)/(# tutee syllables*3 (max score)*# observers)	Full motif
SAP similarity score model-tutee	Median SAP similarity scores comparing tutor's/father's to tutee's motifs	10 random motifs
SAP similarity score tutee-model	Median SAP similarity scores comparing tutee's to tutor's/father's motifs	10 random motifs
Luscinia similarity score	Median 1 – Luscinia distance score for tutor/father motifs compared to tutee motifs	10 random motifs
SAP stereotypy score	Median SAP similarity scores within-tutee motif comparisons	10 random motifs

Luscinia stereotypy score	Median 1 – Luscinia distance scores within-tutee motif comparisons	10 random motifs
Changes 65 to > 100 dph	# changes in motif produced at 65 and >100 dph	Typical motif (65 and 100 dph)

dph: days posthatching. All samples analysed were from the 100 dph recordings, except the sample used to calculate Changes 65 to > 100 dph. For that parameter, the typical motifs recorded at 65 and >100 dph were analysed. The parameters and definitions listed here are identical to those used by Varkevisser et al. (2021).

Statistical analysis

We used RStudio (R version 3.5.1, http://www.rstudio.com/) to build linear mixed-effects models (LMMs) to compare whether treatment groups differed in number of unique syllables, the sequence linearity and sequence consistency scores, and the human observer, SAP and Luscinia scores. Human observer, SAP and Luscinia scores are bounded distributions and were therefore arcsine square-root transformed prior to analyses to meet model assumptions. Generalized linear mixed-effect models (GLMMs) with a Poisson distribution and log-link function were used to assess whether tutees from different treatments differed in the total number of syllables and the number of changes from 65 to > 100 DPH (package lme4: Bates, Mächler, Bolker, & Walker, 2014). For every parameter, we first ran a null model including 'Tutor group' (Number of the tutor group, 15 tutor groups in total) as a random intercept and 'Location' (Leiden or Berlin) as a fixed factor. We always included 'Location' as the locations differed in the technical realization of the experiment (see above). We used ANOVAs to compare this null model to a model that included 'Treatment' (Live, Audiovisual or Audio-only) as a fixed factor. We used a Shapiro–Wilk test to assess whether the model's residuals followed a normal distribution. Post hoc tests with Tukey adjustment for multiple comparisons were conducted for between-treatment comparisons if the model with 'treatment' was significantly better than the model without 'treatment' as a fixed factor (package emmeans: Lenth, Singmann, Love, Buerkner, Herve, 2018). The three similarity scores (human observers, SAP and Luscinia) for all tutees and the live tutored tutees only were compared with Pearson correlation coefficients after the human observer scores were square-root transformed to meet normality assumptions.

Ethical note

We adhered to the ASAB/ABS Guidelines for the Use of Animals in Research and the European and Dutch legislation on animal experimentation. At all stages in the experiment birds had ad libitum access to food and water and were cohoused with at least one other bird (apart from the short song recording sessions). The manipulation of the size and composition of social groups (in accordance with general housing procedures) is not considered a procedure in the Experiments on Animals Act (Wet op de Dierproeven, 2014) which is the applicable legislation in the Netherlands in accordance with the European guidelines (EU directive no. 2010/63/EU) regarding the protection of animals used for scientific purposes. We also had no indication that the described procedures induced distress or impaired welfare. At all times, all birds were housed and cared for in accordance with these regulations and internal guidelines concerning care of the animals and licensing and skill of personnel, including review and monitoring by the Leiden University Animal Welfare Body and following their advice to ensure the wellbeing of all animals at the facility (with or without a licence).

Results

Usage of observation huts by tutees

To assess whether tutees came near the one-way mirrors, tutee position (inside or outside the observation hut) was scored for the audio-only and audiovisual tutees of four tutor groups (N = 16: four male and four female audio-only and audiovisual tutees). All tutees used the perches in the observation hut more often than expected by chance. Although only 14.9% of the total perch area was inside the observation hut, the average percentage of observations (mean ± SD) inside the observation huts was almost double of what was expected: in the audiovisual group males spent 28 ± 19% (N = 4 birds, 15 observation-hours) and females 26 ± 18% (N = 4 birds, 28.5 observation-hours) of observations in the huts. Interestingly, for the audio-only tutees, the percentage of observations where the tutees were inside the hut was similar to that for the audiovisual groups for both males (26 ± 18%, N = 4 birds, 32 observation-hours) and females (25 ± 17%, N = 4 birds, 32 observation-hours) suggesting that the huts were a preferred area for all birds, regardless of whether it allowed them to see the tutor.

Song structure and performance

The parameters used to assess song structure and performance (total number of syllables, number of unique syllables, linearity and consistency) did not vary significantly between tutoring treatments (models including ‘treatment’ as fixed factor were not significantly better than the models without ‘treatment’, see Table 2 and the details of the models with treatment in Table 3). To test whether the tutees from the different treatments differed in between-motif stereotypy, we compared the 10 randomly selected tutee motifs to each other in SAP and Luscinia. There was no significant effect of treatment on the SAP or Luscinia stereotypy scores (adding 'treatment' as fixed factor did not significantly improve the null model: SAP stereotypy score: N = 41, χ_2 = 2.18, P = 0.34; Fig. 4a, Table 4; Luscinia stereotypy score: N = 41, χ_2 = 0.15, P = 0.93; Fig. 4b, Table 4).

Table 2. Mean values for the song structure and performance parameters per treatment group and details on ANOVA comparing the null model with a model including 'treatment' as a fixed effect (both models included 'tutor group' as random factor and 'location' as fixed factor). We did not include the tutor data in the models.

	Tutor (not in models)	**Live** 🐦👁👂	**Audio-visual** 👁👂	**Audio** 👂	**ANOVA null model and model with 'treatment'**		
	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	N	χ^2	p
Total nr syllables	*7.1 ± 2.0*	7.2 ± 1.5	7.3 ± 3.4	6.4 ± 2.3	41	0.88	0.64
Nr unique syllables	*6.2 ± 1.1*	5.5 ± 1.7	6.1 ± 1.7	5.9 ± 1.7	41	0.87	0.65
Linearity	*0.55 ± 0.12*	0.43 ± 0.09	0.47 ± 0.12	0.47 ± 0.08	41	1.19	0.55
Consistency	*0.93 ± 0.06*	0.92 ± 0.05	0.92 ± 0.06	0.91 ± 0.06	41	0.17	0.92

Table 3. Details of models with treatment as fixed factor for the song structure and performance parameters

Response variable	***Model term***	***Level***	***Estimate***	***SE***	***z/t***
Total no. of syllables[1]	Intercept		1.92	0.12	15.55
	Treatment				
		Audio-visual	0.12	0.15	0.81
		Live	0.12	0.14	0.82
	Location				
		Leiden	-0.11	0.12	-0.96
No. of unique syllables[1]	Intercept		6.62	0.53	12.39
	Treatment				
		Audio-visual	0.15	0.61	0.24
		Live	-0.36	0.57	-0.63
	Location				
		Leiden	-1.21	0.55	-2.21

Linearity[2]	Intercept		0.46	0.03	14.32
	Treatment				
		Audio-visual	0.004	0.04	0.12
		Live	-0.03	0.04	-0.86
	Location				
		Leiden	0.003	0.03	0.09
Consistency[2]	Intercept		0.89	0.02	50.95
	Treatment				
		Audio-visual	0.01	0.02	0.36
		Live	0.01	0.02	0.28
	Location				
		Leiden	0.04	0.02	2.46

[1] GLMM with a Poisson distribution and 'Tutor group' as a random factor

[2] LMM with 'Tutor group' as a random factor

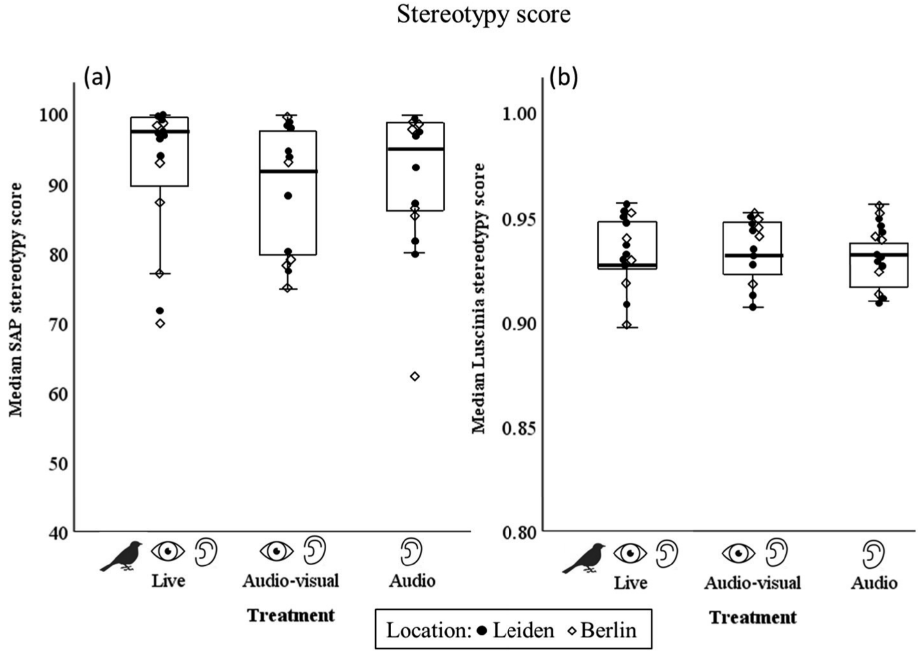


Figure 4. (a) SAP and (b) Luscinia stereotypy scores for the 10 randomly selected tutee motifs recorded at 100 days posthatching in the three treatments. Box plots indicate the median (midline), interquartile range (box) and 1.5 times the interquartile range (whiskers).

Table 4. Details of models with 'treatment' as fixed factor for the (arcsine square-root transformed) SAP and Luscinia stereotypy scores

Response variable[1]	Model term	Level	Estimate	SE	t
SAP stereotypy score	Intercept		1.24	0.06	21.87
	Treatment				
		Audio-visual	-0.05	0.07	-0.77
		Live	0.04	0.06	0.66
	Location				
		Leiden	0.09	0.06	1.61
Luscinia stereotypy score	Intercept		0.93	0.006	165.83
	Treatment				
		Audio-visual	0.002	0.005	0.33
		Live	0.002	0.005	0.33
	Location				
		Leiden	-0.002	0.006	-0.32

[1] LMM with 'Tutor group' as random factor.

Similarity to tutors' and fathers' song

Comparison of different similarity assessment methods

For the similarity scores of all tutees, we found a significant correlation between the human observer and the Luscinia similarity scores and between the human observer and the SAP similarity scores for the father–tutee comparison. There was no correlation between the Luscinia and SAP similarity scores (see Table 5). When looking at the similarity scores of the live tutored tutees only (to enable comparison with earlier studies comparing Luscinia or SAP with human observer similarity scores for live tutored tutees), we only found a significant correlation between the Luscinia and the human observer similarity scores (see Table 5). To find out whether the low correlation between the SAP and the human observer similarity scores was related to the different song samples used to calculate these (one typical motif for the human observer scores and 10 randomly selected motifs per tutee for the SAP scores), we repeated the SAP similarity score calculations with the same sample that was used for the human observer similarity scores (one typical motif for each tutee compared to one typical motif of each tutor, the same motifs that were used for the human observer similarity scores). This led to a significant correlation between the

SAP and human observer scores for the tutor–tutee comparison, but not for the father–tutee comparison (see Table 5). None of the correlation coefficients were very high (all below 0.73), suggesting that the three methods measured different aspects of song similarity. Below, we present the data from all three similarity assessment methods.

Table 5. Pearson correlation coefficients for the human observer similarity scores (square-root transformed), the median SAP and the median Luscinia similarity scores

	Tutor-Tutee		Father-Tutee	
Comparison[1]	r	p	r	p
All tutees				
Human – SAP	0.15	0.37	**0.32**	**0.04**
Human – Luscinia	**0.73**	**<0.01**	**0.47**	**<0.01**
SAP - Luscinia	0.10	0.53	0.04	0.82
Live tutored tutees only				
Human – SAP	0.41	0.15	0.21	0.47
Human – Luscinia	**0.69**	**<0.01**	**0.59**	**0.03**
SAP - Luscinia	0.04	0.89	-0.11	0.72
All tutees: one motif of each tutee				
Human – SAP	**0.40**	**<0.01**	0.26	0.12

Sample sizes: all tutees: tutor–tutee: N = 41; father–tutee: N = 39; live-tutored tutees only: tutor–tutee: N = 14; father–tutee: N = 13. Significant values are given in bold, for all tutees, for the live tutees only and for the SAP and human observer similarity scores calculated for the same sample of one typical motif of each tutee compared to one typical motif of the tutor.

[1] Human = human observers similarity score, SAP = SAP similarity score, Luscinia = Luscinia similarity score.

Similarity between tutees' and their tutors' songs

For the human observer similarity scores calculated by comparing the tutor's syllables to the tutee's syllables (tutor–tutee comparison), adding 'treatment' as fixed factor to the null model did not lead to a significant improvement (N = 41, χ_2 = 2.78, P = 0.25). Similarity was highest for the tutees in the live treatment group (model estimates LMM: mean ± SE: 0.68 ± 0.04; Table 6, Fig. 5a), followed by the audiovisual (mean ± SE: 0.59 ± 0.04) and the audio-only tutees (mean ± SE: 0.58 ± 0.04). Likewise, for the tutee–tutor comparison, adding ‘treatment’ as fixed factor to the null model did not lead to a significant improvement (N = 41, χ_2 = 1.08, P = 0.58; Table 6). Again, similarity was highest in the live group (model estimates LMM: mean ± SE: 0.84 ± 0.03; Table 6, Fig. 5b), intermediate in the audiovisual group (mean ± SE: 0.80 ± 0.04) and lowest in the audio-only group (mean ± SE: 0.73 ± 0.03).

For the comparison of the tutor’s and tutee’s songs in SAP, there was no significant effect of tutoring treatment in the tutor–tutee or tutee–tutor comparison; the model including ‘treatment’ as fixed factor was not significantly better than the null model for the SAP tutor–tutee similarity scores (N = 41, χ_2 = 0.44, P = 0.80; Table 6, Fig. 5c) and the SAP tutee–tutor similarity scores (N = 41, χ_2 = 2.49, P = 0.29; Table 6, Fig. 5d).

Treatment had a significant effect on Luscinia similarity scores for the comparison between tutees and their tutors’ songs (adding ‘treatment’ as fixed factor significantly improved the null model: N = 41, χ_2 = 8.72, P = 0.01; Fig. 5e, Table 6): this score was higher in the live than in the audiovisual tutees and there was a nonsignificant trend for the live tutored tutees also having higher scores than the audio-only tutees (for post hoc test results see Table 6).

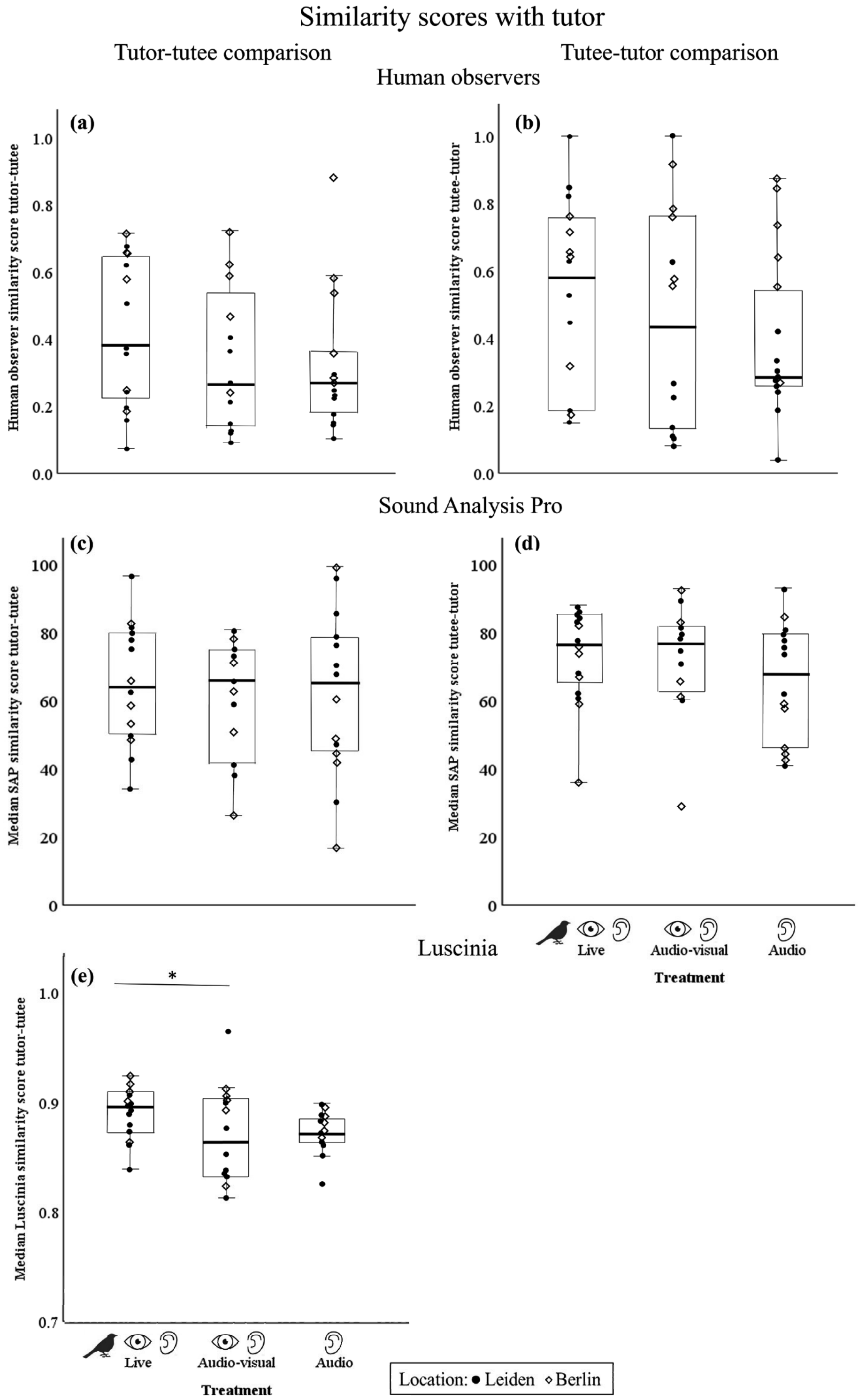

Similarity scores with tutor
Tutor-tutee comparison
Tutee-tutor comparison
Human observers
(a)
Human observer similarity score tutor-tutee
0.0
0.2
0.4
0.6
0.8
1.0
(b)
Human observer similarity score tutee-tutor
Sound Analysis Pro
(c)
Median SAP similarity score tutor-tutee
0
20
40
60
80
100
(d)
Median SAP similarity score tutee-tutor
Live
Audio-visual
Audio
Treatment
Luscinia
(e)
*
Median Luscinia similarity score tutor-tutee
0.7
0.8
0.9
1.0
Live
Audio-visual
Audio
Treatment
Location: ● Leiden ◇ Berlin

Table 6 Details of models with 'Treatment' as fixed factor for the arcsine square-root transformed human observer, SAP and Luscinia similarity scores for the comparison of tutor and tutee song

			Tutor-tutee			*Tutee-tutor*		
Response variable	***Model term***	***Level***	***Estim.***	***SE***	***t***	***Estim.***	***SE***	***t***
Human observers sim. scores[1]	Intercept		0.74	0.07	11.19	0.87	0.10	8.47
	Treatment							
		Audiovisual	0.01	0.07	0.19	0.07	0.12	0.59
		Live	0.10	0.07	1.52	0.12	0.12	0.99
	Location							
		Leiden	-0.28	0.07	-4.06	-0.24	0.10	-2.40
SAP sim. scores[1]	Intercept		0.87	0.08	10.35	0.87	0.06	14.85
	Treatment							
		Audiovisual	-0.03	0.07	-0.36	0.08	0.06	1.36
		Live	0.02	0.07	0.28	0.08	0.06	1.38
	Location							
		Leiden	0.09	0.10	0.93	0.15	0.06	2.30
Lusc. sim. scores[2]	Intercept		0.09	0.0005	203.8			
	Treatment							
		Audiovisual	-0.0004	0.0005	-0.74			
		Live	0.001	0.0005	2.28			
	Location							
		Leiden	-0.001	0.0005	-2.73			

[1] LMM with 'Tutor group' as random factor.

[2] LMM with 'Tutor group' as random factor. Post hoc comparisons: audiovisual versus live: estimate = -0.001, SE = 0.0005, t = -2.92, P = 0.02; audio-only versus live: estimate = -0.001, SE = 0.0005, t = -2.28, P = 0.08; audio-only versus audiovisual: estimate = 0.0004, SE = 0.0005, t = 0.737, P = 0.74.

Similarity between tutees' and their fathers' songs

We also checked whether birds had learned from the father with which they had been housed before the experimental tutoring (Böhner, 1990). For the human observer similarity scores in the comparison of the father's syllables to the tutee's syllables (father–tutee comparison), adding 'treatment' as fixed factor did not significantly improve the null model (N = 39, χ_2 = 3.38, P = 0.18), but as above we kept the experimental ‘treatment’ as fixed factor in the final model (Table 7). The similarity scores for the father–tutee comparison were highest in the group that learned the least during the experimental phase, namely the audio-only group (model estimates LMM: mean ± SE: 0.72 ± 0.03; Table 7, Fig. 6a), followed by the live (mean ± SE: 0.64 ± 0.02) and the audiovisual group (mean ± SE: 0.58 ± 0.03). For the human observer similarity scores in the tutee–father comparison, adding ‘treatment’ as fixed factor also did not significantly improve the null model (N = 39, χ_2 = 0.20, P = 0.91). The human observer similarity scores for this comparison were highest in the audio-only group (mean ± SE: 0.60 ± 0.09; Table 7, Fig. 6b) compared to the audiovisual (mean ± SE: 0.57 ± 0.10) and live groups (mean ± SE: 0.56 ± 0.09).

For the comparison of the father’s song and tutee’s song in SAP, there was no significant effect of tutoring treatment in the father–tutee or tutee–father comparison (model including ‘treatment’ as fixed factor was not significantly better than the null model for the SAP father–tutee comparison (N = 39, χ_2 = 2.07, P = 0.35; Table 7, Fig. 6c) and the SAP tutee–father comparison (N = 39, χ_2 = 0.23, P = 0.89; Table 7, Fig. 6d).

Treatment did not significantly affect Luscinia similarity scores for the comparison between tutees and their fathers’ songs (model with ‘treatment’ as fixed factor was not significantly better than the null model: N = 39, χ_2 = 3.31, P = 0.19; Table 7, Fig. 6e).

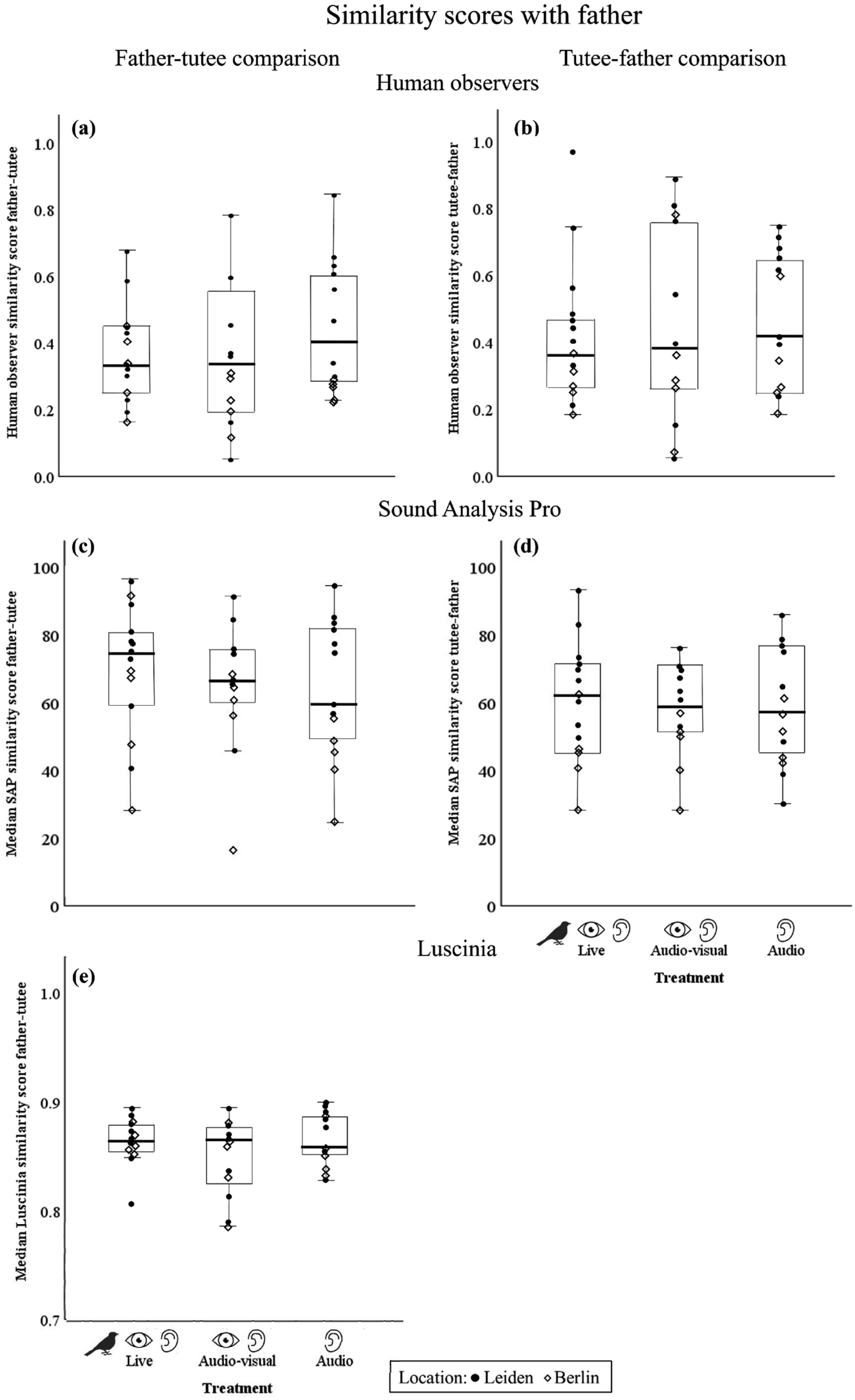

Similarity scores with father
Father-tutee comparison
Tutee-father comparison
Human observers
(a)
Human observer similarity score father-tutee
0.0
0.2
0.4
0.6
0.8
1.0
(b)
Human observer similarity score tutee-father
Sound Analysis Pro
(c)
Median SAP similarity score father-tutee
0
20
40
60
80
100
(d)
Median SAP similarity score tutee-father
Live
Audio-visual
Audio
Treatment
Luscinia
(e)
Median Luscinia similarity score father-tutee
0.7
0.8
0.9
1.0
Live
Audio-visual
Audio
Treatment
Location: ● Leiden ◇ Berlin

Table 7 Details of models with 'Treatment' as fixed factor for the arcsine square-root transformed human observer, SAP and Luscinia similarity scores for the comparison of father and tutee song

			Father-tutee			***Tutee-father***		
Response variable	***Model term***	***Level***	***Estim.***	***SE***	***t***	***Estim.***	***SE***	***t***
Human observers sim. scores[1]	Intercept		0.61	0.06	9.76	0.60	0.09	6.84
	Treatment							
		Audio-visual	-0.13	0.07	-1.74	-0.03	0.10	-0.28
		Live	-0.08	0.07	-1.18	-0.04	0.09	-0.43
	Location							
		Leiden	0.18	0.06	3.00	0.23	0.09	2.69
SAP sim. scores[1]	Intercept		0.78	0.08	10.04	0.74	0.05	14.50
	Treatment							
		Audio-visual	0.003	0.06	0.05	-0.006	0.06	-0.10
		Live	0.07	0.06	1.28	0.02	0.06	0.33
	Location							
		Leiden	0.23	0.09	2.60	0.20	0.05	4.11
Lusc. sim. scores[1]	Intercept		0.09	0.0005	174.6			
	Treatment							
		Audio-visual	-0.0009	0.0006	-1.59			
		Live	-0.0001	0.0006	-0.12			
	Location							
		Leiden	0.0006	0.0005	1.13			

[1] LMM with 'Tutor group' as random factor.

Changes between 65 and > 100 dph

The amount of changes in the typical motif between 65 and >100 DPH differed per treatment group: there were more changes in the motif of the audio-only tutored birds than the live tutored birds and there was a nonsignificant trend for the audio-only birds showing more changes in their motif than the audio-visual birds (model with 'treatment' significantly better than the null model: N = 33, χ_2 = 9.29, P < 0.01; Fig. 7, for post hoc test results see Table 8). Of the 12 birds in the audio-only group that we could record at both 65 and >100 DPH, three did not change anything, five had added one or more syllables to their typical motif and four birds had deleted one or more syllables from their typical motif between 65 and >100 DPH.

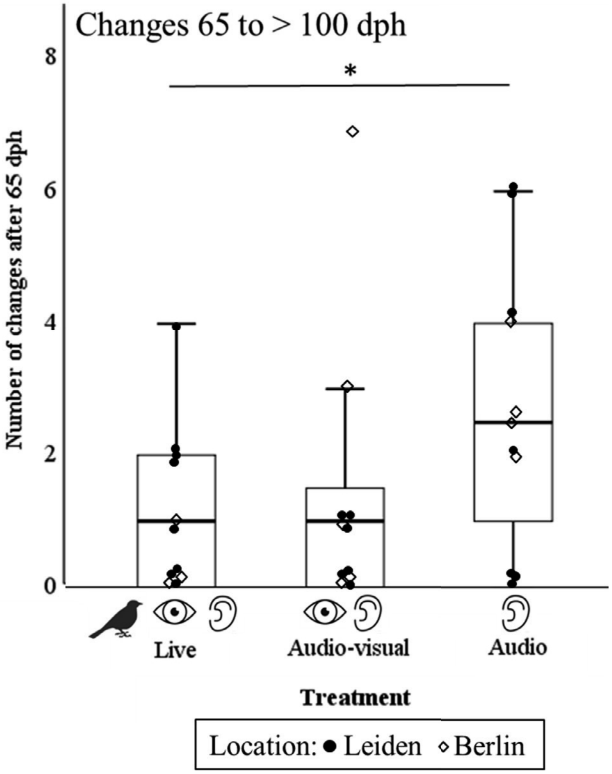


Figure 7. Number of changes in the typical motif of the tutees in the three treatment groups between 65 and >100 days posthatching. Box plots indicate the median (mid-line), interquartile range (box) and 1.5 times the interquartile range (whiskers). *P < 0.05, GLMM see Table 8.

Table 8. Details of best model (GLMM) for the number of changes between 65 and > 100 days posthatching (response variable)

Model term	***Level***	***Estimate***	***SE***	***z***	***p***
Intercept		1.01	0.38	2.66	<**0.01**
Treatment					
	Audio-visual	-0.77	0.33	-2.33	**0.02**
	Live	-0.91	0.36	-2.53	**0.01**
Location					
	Leiden	-0.25	0.45	-0.57	0.57

GLMM with Poisson distribution and 'Tutor group' as random factor. Post hoc comparisons: audio-only versus live: estimate = 0.91, SE = 0.36, z = 2.53, P = 0.03; audio-only versus audio-visual: estimate = 0.77, SE = 0.33, z = 2.32, P = 0.052; audiovisual versus live: estimate = 0.20, SE = 0.40, z = 0.49, P = 0.88. Significant P values are given in bold.

Discussion

The aim of this study was to test whether audiovisual exposure to a live tutor would facilitate song learning in zebra finches in comparison to auditory exposure only. To test this hypothesis, birds within each tutor group (with 15 replicates) were simultaneously tutored by an adult male in one of three conditions: cohousing with the tutor ('live' tutoring group), audio-only exposure to the same tutor via an acoustically transparent loudspeaker cloth ('audio') or both auditory (loudspeaker cloth) and visual (through a one-way mirror) exposure to the same tutor ('audiovisual'). Overall, the findings in this study suggest improved song learning in the tutees with audiovisual tutor exposure ('live' and 'audiovisual' group) compared to the group with audio-only tutor exposure ('audio'). First, the tutees with audiovisual exposure showed a different developmental trajectory: they made fewer changes between the subadult and adult recordings than the tutees with audio-only exposure. Second, their songs tended to have a higher similarity to the tutor's song and a lower similarity to the father's song (to which they were exposed before the peak of the sensitive period for song learning) than the tutees with audio-only exposure, but these between-treatment differences in similarity were not significant.

Birds from the audio-only condition differed from the birds in the other treatments in how their song developed. During song development, tutees start producing highly variable subsong, which becomes more stereotyped in structure and sequence over time. In socially reared zebra finches, around 60 DPH almost all syllables of the final song are produced and often in the same sequence as in adulthood (Arnold, 1975; Slater, Eales, & Clayton, 1988). A higher number of changes after 65 DPH might thus indicate a delay in song development, compared to birds with fewer changes after 65 DPH. More song plasticity after the peak of the sensitive period for song learning has been found in zebra finches housed in social isolation between 35 and 120 DPH compared to zebra finches housed in peer groups during this period (Jones et al. 1996). In the study described here, there were more changes in the typical motif of the audio-only birds than in that of the live birds after 65 DPH, which is in line with earlier findings showing that zebra finch tutees that were only auditorily exposed to adult conspecifics during the sensitive period for song learning change their song up until a later age than control birds reared in aviaries together with adult conspecifics (Morrison & Nottebohm, 1993). This earlier study concluded that the closing of the sensitive period depends on whether a bird was able to have visual social interactions with a tutor. Here, however, we did not find a difference in the number of changes between the live group and the audiovisual group, while tutees in this latter group could not have visual social interactions with the tutor. This suggests that the timing of song development might be influenced not only by visual social interaction, but also by mere visual exposure to the tutor. Our results thus provide direct support for the hypothesis that seeing as well as hearing a tutor has a facilitating effect on the timing of song learning, independent of visual tutor–tutee interaction.

The zebra finch literature generally reports less learning from audio-only than from live tutors (e.g. Derégnaucourt et al., 2013; Eales, 1989; reviewed in Derégnaucourt, 2011). In contrast, in the current study, audio-only and live tutored birds did not differ significantly in how much they learned from the tutor. In our experiment, other than in earlier tape versus live tutor comparisons, audio-tutored birds could vocally interact with the live tutor and also had a nonsinging female as a social companion to avoid the potential confound of social isolation on song development in the audio-only and audiovisual groups, which could explain why audio-only and live tutored tutees showed comparable song learning in our study. It is also possible that being housed with a female companion resulted in fewer syllables copied from the tutor in all conditions, including the live condition, because tutees incorporated calls from the female in their song (see e.g. Price, 1979) or were reinforced by their female companions' behaviour to retain specific syllables that resembled the song of the female's father, which was different from that of the male tutee's father or tutor (Carouso-Peck et al., 2020; Carouso-Peck & Goldstein, 2019; Jones & Slater, 1993). We can compare the absolute scores directly with those from Phan et al. (2006), who computed SAP similarity scores in a similar fashion to us for live tutored tutees, but their tutees only had their father as a tutor and were housed continuously with their parents and siblings. These tutees had a higher average similarity to their tutor's song (71±4) than the tutees in the current study (62±3). Derégnaucourt et al. (2013) computed SAP similarity scores in a similar fashion for an experiment also involving a live and audio-only condition, but where all tutees' fathers had been removed at 25 days and where live and audio-only tutees were housed without a female companion. Derégnaucourt et al. (2013) also found a higher average SAP similarity score for the live tutored tutees (76±4) than we found in the current study (62±3), but a comparable average similarity score for the audio-only tutored tutees (60±4; this study: 61±3). This suggests that in the current study the female companion, the prolonged exposure to the father's song in most tutees or the vocal interaction with the other tutees behind the loudspeaker cloths (Honarmand et al., 2015) could have contributed to the lack of a significant difference between the live and audio-only tutored tutees.

The highest tutor song copying rates in the live group and the lowest tutor song copying rates in the audio-only group are in line with earlier studies that have shown that reduced quality of tutor access during the peak of the sensitive period for song learning (35–65 DPH) can lead to increased copying of song heard before this period (reviewed in Gobes et al., 2017), which in our experiment was the father's song. The audio-only group indeed showed the highest and the live tutored group the lowest father's song copying rate, although this was not significantly different. This suggests that causes other than social isolation or a lack of vocal isolation might play a role in the lower song copying by audio-only tutees as well. The lowest overall song copying in this group is in line with a lack of visual cues being one possible cause of poorer song copying in audio-only tutees.

The study included tutees raised in Leiden and Berlin. Compared to the Leiden Song similarity with tutor song was highest in the birds from the live group. Multiple factors may have contributed here: unlike the other tutees, live tutees could visually and physically interact with and physically approach the tutor (Liu et al., 2021), which might have contributed to song learning success. Besides, due to our experimental set-up, extrapolating from the acoustic transmission properties, unless the tutor was sitting close to the loudspeaker cloth, the tutor song was louder (and likely to be so on average) in the central compartment than in the other two compartments, which, if amplitude affects learning, might have contributed to the higher similarity scores of the live tutees. Our results are in line with other observations of improved learning from live tutors, but the question of which factors contribute to the improved song learning from live versus different types of audio-only tutors (ranging from stereotyped playbacks to visually occluded tutors, e.g. Baptista & Gaunt, 1997; Beecher, 2017; Houx & ten Cate, 1999; Nelson, 1997, 1998) is an ongoing discussion.

Earlier studies addressed the question whether additional visual stimulation improved song learning but have not found an effect. Presentations of a stationary taxidermic mount of a zebra finch male as a visual stimulus right before, during or after tutor song presentation or presentation of a video of a singing tutor synchronized with tutor song did not lead to more song copying than tape tutoring only (Bolhuis et al., 1999; Houx & ten Cate, 1999; Varkevisser et al., 2021). In the study presented here, the tutees with visual tutor exposure (the live and audiovisual groups) tended to have a higher similarity to the tutor's song and a lower similarity to the father's song than the birds without visual tutor exposure (the audio-only group), but this difference was not significant.

The study included tutees raised either in Leiden or Berlin. Compared to the Leiden tutees, the Berlin tutees produced songs more similar to their tutor's song and less similar to their father's song. It is difficult to pinpoint the reason for this at this stage, as the differences could be stochastic or arise from a number of differences in the technical realization of the experiment at the two locations. While all tutees in Leiden and Berlin were moved to the tutoring set-up at day 35, tutees in Berlin were in a different compartment of the cage than their father between day 23 and day 35, so that they could hear the father but not see him or interact with him physically. In Leiden the father remained in the same space with the juveniles until day 35. The tutees in Leiden may therefore have picked up more from their father than the tutees in Berlin. This will, however, need systematic study as there were also other differences between Berlin and Leiden. For example, in Berlin, the three compartment cages were positioned in a soundproof box, while in Leiden they were in a room with other birds present. This probably made the tutor in Berlin more audible to the tutees in the side compartments than in Leiden. Conversely, the tutees in the compartments adjacent to the tutor might also have been more audible to the tutor which could have led to more vocal interactions. Individual rather than population differences might also have contributed: there were only 15 tutors in total. If in our colonies some tutors are copied more and more readily by (related and unrelated) young birds as reported for other colonies (Tchernichovski et al., 2021) and some of these preferentially and better copied tutors were better represented in one location, this could be mistaken for a location or population effect.

The three similarity assessment methods used in this study (human observers, Luscinia and SAP) differed in whether they picked up a significant treatment effect. In previous studies, Luscinia and SAP were both found to be highly correlated with human observer similarity scores (Luscinia: r = 0.96, N = 18, Lachlan et al. 2010; SAP: r = 0.91, N =10, Tchernichovski et al., 2000). In the current study, in all comparisons the human observers' scores were significantly correlated with the Luscinia but not the SAP scores. Our study design had several features that differed from the previous comparison between SAP and human observer similarity scoring (Tchernichovski et al., 2000) that might have led to this lower correlation. For instance, the previous study used only live tutoring and one tutor, and therefore tutees in the previous comparison probably copied more from the tutor than the tutees in the current study. As the correlation between human observers and SAP was weaker in the larger sample including all groups than in the smaller sample only involving the live tutored birds, SAP or the human observers might have more difficulty assessing similarity between model song and poorly copied tutee song, than between model song and well-copied song. Human visual scoring was used to validate the two automated methods and is considered a suitable method for assessing song similarity if multiple independent observers that are blind to the expected outcome of the comparisons are used (Jones, ten Cate, & Bijleveld, 2001), which was the case in our study, and which is why our conclusions are mainly based on the results from the human observer similarity scoring.

Overall, our findings suggest that birds with multimodal exposure developed their adult song faster and tended to produce songs that were more similar to the tutor's song than birds with unimodal tutor exposure, which is in line with our hypothesis that visual exposure to a singing tutor has a facilitating effect on zebra finch song learning. There are various ways in which visual exposure to the tutor might have facilitated song learning in this experiment. First, as we hypothesized, the beak and throat movements associated with song production might have made song easier to detect and remember, which would be in line with a study showing that visual stimulation matched in rhythm to auditory song presentation can facilitate song learning (Hultsch et al., 1999) and studies showing that stimuli with multiple components (in one or multiple modalities) are easier to detect and remember than unicomponent stimuli (reviewed in Hebets & Papaj, 2005; Rowe, 1999). On the other hand, the facilitating effect might not necessarily have to do with the coupling between visual and auditory song exposure. It might also be that the tutor provided visual feedback in reaction to songs produced by the tutees. Young zebra finches that received contingent visual feedback (a video of a female conspecific) on their immature song production copied more tutor song than birds that received noncontingent visual feedback (Carouso-Peck, & Goldstein, 2019) and an observational study showed that the number of fluff-ups performed by the mother before, during or after tutee song production was positively correlated with tutee song learning success (Carouso-Peck et al., 2020). In this observational study, no visible behaviour of the father was studied as possible feedback to juvenile song production (Carouso-Peck et al., 2020). From our tutoring experiment, we cannot rule out the possibility that the tutor provided visual feedback to the tutees, which might have facilitated song learning in the birds with visual access to the tutors. Follow-up studies, for example where no vocal interaction is possible between the tutor and tutees, could help find out which mechanism underlies the effect that visual exposure to a tutor has on zebra finch song learning.

In this study, we disentangled the effect of multimodal exposure to a tutor from that of social visual interaction with a tutor on the song learning process in zebra finches. Our results suggest that multimodal exposure to a tutor affects zebra finch song development and might be one of the factors involved in the difference in song learning success from live and tape tutors. Follow-up studies are necessary to get more insight into the mechanism through which multimodal exposure to a tutor facilitates song learning. This can give more insight into the factors involved in the vocal learning process.

Acknowledgements

Funding for this research was provided by a Human Frontier Science Program Grant (No RGP0046/2016). We would like to thank Carel ten Cate for comments on an earlier version of this manuscript and members of the Behavioural Biology group in Leiden for discussion. We would like to thank Jing Wei, Quanxiao Liu and Zhiyuan Ning for the visual comparison of the spectrograms.

References

Arnold, A. P. (1975). The effects of castration on song development in zebra finches (*Poephila guttata*). Journal of Experimental Zoology, 191(2), 261–278. https://doi.org/10.1002/jez.1401910212

Baptista, L. F., & Gaunt, S. L. L. (1997). Social interaction and vocal development in birds. In C. T. Snowdon & M. Hausberger (Eds.), Social influences on vocal development (pp. 23–40). Cambridge, Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using *lme4*. 67(1). https://doi.org/10.18637/jss.v067.i01

Bischof, H. J. (1988). The visual field and visually guided behavior in the zebra finch (*Taeniopygia guttata*). Journal of Comparative Physiology A, 163, 329–337. https://doi.org/10.1007/BF00604008

Böhner, J. (1986). Der zeitliche Verlauf des Gesangerwerbs beim Zebrafinken. Ver handlungen Der Deutschen Zoologischen Gesellschaft, 79.

Bolhuis, J., van Mil, D., & Houx, B. (1999). Song learning with audiovisual compound stimuli in zebra finches. Animal Behaviour, 58, 1285–1292. https://doi.org/10.1006/anbe.1999.1266

Carouso-Peck, S., & Goldstein, M. H. (2019). Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. Current Biology, 29, 631–636. https://doi.org/10.1016/j.cub.2018.12.026

Carouso-Peck, S., Menyhart, O., DeVoogd, T. J., & Goldstein, M. H. (2020). Contingent parental responses are naturally associated with zebra finch song learning. Animal Behaviour, 165, 123–132. https://doi.org/10.1016/j.anbehav.2020.04.019

Catchpole, C. K., & Slater, P. J. B. (2003). Bird song: biological themes and variations. Cambridge, Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. Proceedings of the National Academy of Sciences, 201522306. https://doi.org/10.1073/pnas.1522306113

Clayton, N. S. (1988). Song tutor choice in zebra finches and bengalese finches: the relative importance of visual and vocal cues. Behaviour, 104, 281–299.

Derégnaucourt, S. (2011). Birdsong learning in the laboratory, with especial reference to the song of the zebra finch (*Taeniopygia guttata*). Interaction Studies, 12, 324–350. https://doi.org/10.1075/is.12.2.07der

Derégnaucourt, S., Poirier, C., van der Kant, A., & van der Linden, A. (2013). Comparisons of different methods to train a young zebra finch (*Taeniopygia gut tata)* to learn a song. Journal of Physiology, 107, 210–218. https://doi.org/10.1016/j.jphysparis.2012.08.003

Doupe, A. J., & Kuhl, P. K. (1999). Bird song and human speech: common themes and mechanisms. Annu. Rev. Neurosci., 22, 567–631. https://doi.org/10.1146/annurev.neuro.22.1.567

Eales, L. A. (1989). The influences of visual and vocal interaction on song learning in zebra finches. Animal Behaviour, 37, 507–508. https://doi.org/10.1016/0003-3472(89)90097-3

Gobes, S. M. H., Jennings, R. B., & Maeda, R. K. (2017). The sensitive period for auditory-vocal learning in the zebra finch: consequences of limited-model availability and multiple-tutor paradigms on song imitation. Behavioural Processes, 163, 5–12. https://doi.org/10.1016/j.beproc.2017.07.007

Goller, F., Mallinckrodt, M. J., & Torti, S. D. (2004). Beak gape dynamics, during song in the zebra finch. Journal of Neurobiology, 59(3), 289–303. https://doi.org/10.1002/neu.10327

Griffith, S. C., & Buchanan, K. L. (2010). The zebra finch : the ultimate Australian supermodel. Emu, 110, v–xii. https://doi.org/10.1071/MUv110n3ED

Halfwerk, W., Varkevisser, J., Simon, R., Mendoza, E., Scharff, C., & Riebel, K. (2019). Toward testing for multimodal perception of mating signals. Frontiers in Ecology and Evolution, 7, 2013–2019. https://doi.org/10.3389/fevo.2019.00124

Hebets, E. A., & Papaj, D. R. (2005). Complex signal function: Developing a framework of testable hypotheses. Behavioral Ecology and Sociobiology, 57(3), 197–214. https://doi.org/10.1007/s00265-004-0865-7

Helekar, S. A., Marsh, S., Viswanath, N. S., & Rosenfield, D. B. (2000). Acoustic pattern variations in the female-directed birdsongs of a colony of laboratory-bred zebra finches. Behavioural Processes, 49(2), 99–110. https://doi.org/10.1016/S0376-6357(00)00081-4

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. Behavioral Ecology and Sociobiology, 67(9), 1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Holveck, M. J., Vieira De Castro, A. C., Lachlan, R. F., ten Cate, C., & Riebel, K. (2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. Behavioral Ecology, 19(6), 1267–1281.

https://doi.org/10.1093/beheco/arn078

Houx, B. B., & ten Cate, C. (1999). Do stimulus-stimulus contingencies affect song learning in zebra finches (*Taeniopygia guttata*)? Journal of Comparative Psychology, 113(3), 235–242. https://doi.org/10.1037/0735-7036.113.3.235

Hultsch, H., Schleuss, F., & Todt, D. (1999). Auditory-visual stimulus pairing enhances perceptual learning in a songbird. Animal Behaviour, 58, 143–149. https://doi.org/10.1006/anbe.1999.1120

Hyland Bruno, J., & Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. Behavioural Processes, 163, 53–59. https://doi.org/10.1016/j.beproc.2017.11.001

Immelmann, K. (1969). Song development in the zebra finch and other estrildid finches. In R. A. Hinde (Ed.), Bird vocalizations. Cambridge, England: Cambridge University Press,.

Jones, A. E., & Slater, P. J. B. (1993). Do young male zebra finches prefer to learn songs that are familiar to females with which they are housed. Animal Behaviour, 46, 616–617. https://doi.org/10.1006/anbe.1993.1233

Jones, A. E., ten Cate, C., & Bijleveld, C. C. J. H. (2001). The interobserver reliability of scoring sonagrams by eye: A study on methods, illustrated on zebra finch songs. Animal Behaviour, 62(4), 791–801. https://doi.org/10.1006/anbe.2001.1810

Jones, A. E., ten Cate, C., & Slater, P. J. B. (1996). Early experience and plasticity of song in adult male zebra finches (*Taeniopygia guttata*). Journal of Comparative Psychology, 110(4), 354–369. https://doi.org/10.1037/0735-7036.110.4.354

Lachlan, R. F., Verhagen, L., Peters, S., & ten Cate, C. (2010). Are there species-uni versal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). Journal of Comparative Psychology, 124(1), 92–108. https://doi.org/10.1037/a0016996

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: estimated marginal means, aka least-squares means.

Liu, W. chun, Landstrom, M., Schutt, G., Inserra, M., & Fernandez, F. (2021). A memory-driven auditory program ensures selective and precise vocal imitation in zebra finches. Communications Biology, 4(1). https://doi.org/10.1038/s42003-021-02601-4

Mann, N. I., & Slater, P. J. B. (1995). Song tutor choice by zebra finches in aviaries. Animal Behaviour, 49(3), 811–820. https://doi.org/10.1016/0003-3472(95)80212-6

Mann, N. I., Slater, P. J. B., Eales, L. A., & Richards, C. (1991). The influence of visual stimuli on song tutor choice in the zebra finch, *Taeniopygia guttata*. Animal Behaviour, 42(2), 285–293. https://doi.org/10.1016/S0003-3472(05)80560-3

Mello, C. V. (2014). The zebra finch, *Taeniopygia guttata*: An avian model for investigating the neurobiological basis of vocal learning. Cold Spring Harbor Protocols, 2014(12), 1237–1242. https://doi.org/10.1101/pdb.emo084574

Morrison, R. G., & Nottebohm, F. (1993). Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. Journal of Neurobiology, 24(8), 1045–1064.

Partan, S., & Marler, P. (1999). Communication goes multimodal. Science, 283, 1272–1274. https://doi.org/0.1126/science.283.5406.1272

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. Journal of Comparative and Physiological Psychology, 93(2), 260–277. https://doi.org/10.1037/h0077553

Rowe, C. (1999). Receiver psychology and evolution of multicomponent signals. Animal Behaviour, 58, 921–931. https://doi.org/10.1006/anbe.1999.1242

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning. Journal of Neuroscience, 11(9), 2896–2913. https://doi.org/10.1523/JNEUROSCI.11-09-02896.1991

Slater, P. J. B., Eales, L. A., & Clayton, N. S. (1988). Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. Advances in the Study of Behaviour, 18, 1–34. https://doi.org/10.1016/S0065-3454(08)60308-3

Soma, M. F. (2011). Social factors in song learning: a review of Estrildid finch research. Ornithological Science, 10(2), 89–100. https://doi.org/10.2326/osj.10.89

Sossinka, R., & Böhner, J. (1980). Song types in the zebra finch. Zeitschrift Für Tierpsychologie, 53, 123–132. https://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Tchernichovski, O., Eisenberg-Edidin, S., & Jarvis, E. (2021). Balanced imitation sustains song culture in zebra finches. Nature Communications, 1–21. https://doi.org/10.1038/s41467-021-22852-3

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. Animal Behaviour, 59(6), 1167–1176. https://doi.org/10.1006/anbe.1999.1416

Ullrich, R., Norton, P., & Scharff, C. (2016). Waltzing Taeniopygia: integration of courtship song and dance in the domesticated Australian zebra finch. Animal Behaviour, 112, 285–300. https://doi.org/10.1016/j.anbehav.2015.11.012

van Kampen, H. S., & Bolhuis, J. J. (1991). Auditory learning and filial imprinting in the chick. Behaviour, 117, 303–319. https://doi.org/10.1163/156853991X00607

van Kampen, H. S., & Bolhuis, J. J. (1993). Interaction between auditory and visual learning during filial imprinting. Animal Behaviour, 45, 623–625. https://doi.org/10.1006/anbe.1993.1074

Varkevisser, J. M., Simon, R., Mendoza, E., How, M., van Hijlkema, I., Jin, R., Liang, Q., Scharff, C., Halfwerk, W. H., & Riebel, K. (2021). Adding colour-realistic video images to audio playbacks increases stimulus engagement but does not enhance vocal learning in zebra finches. Animal Cognition. https://doi.org/10.1007/s10071-021-01547-8

Williams, H. (1990). Models for song learning in the zebra finch: fathers or others?

Animal Behaviour, 39(4), 745–757. https://doi.org/10.1016/S0003-3472(05)80386-0

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (*Taeniopygia guttata*). The Journal of Experimental Biology, 204, 3497–3506.

Appendix

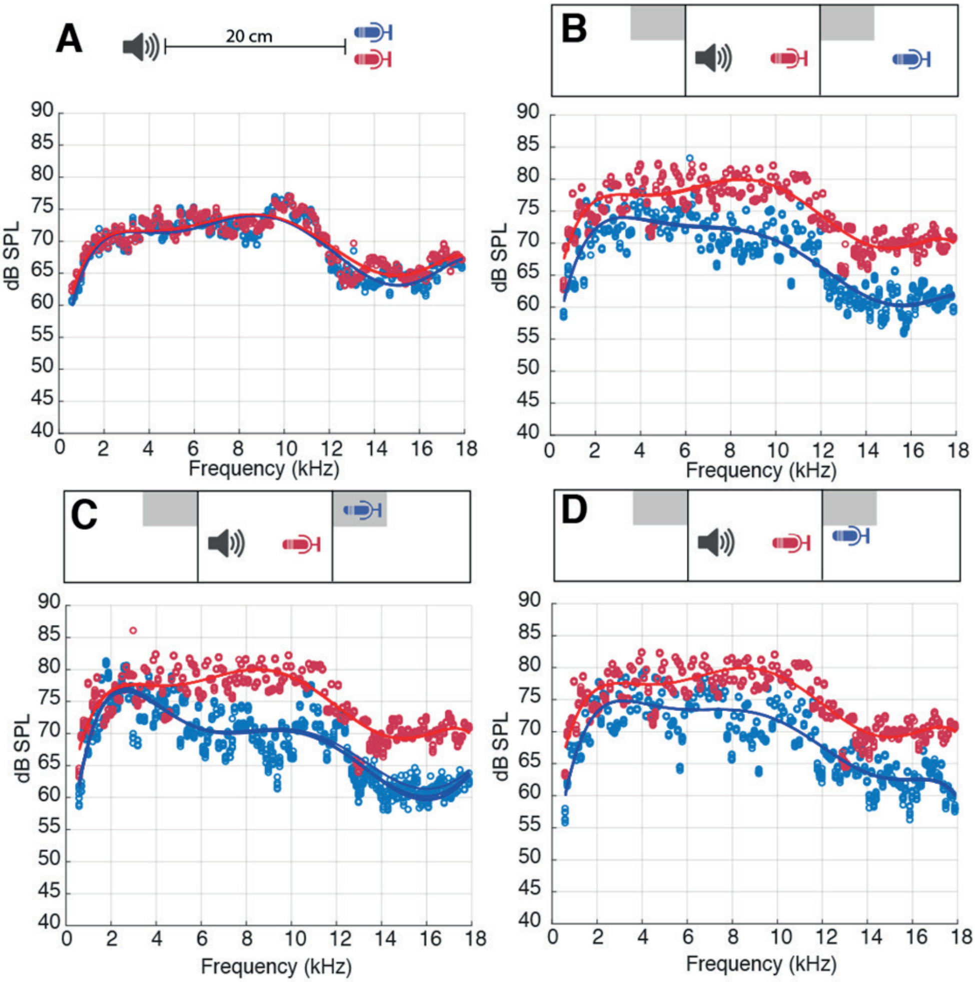
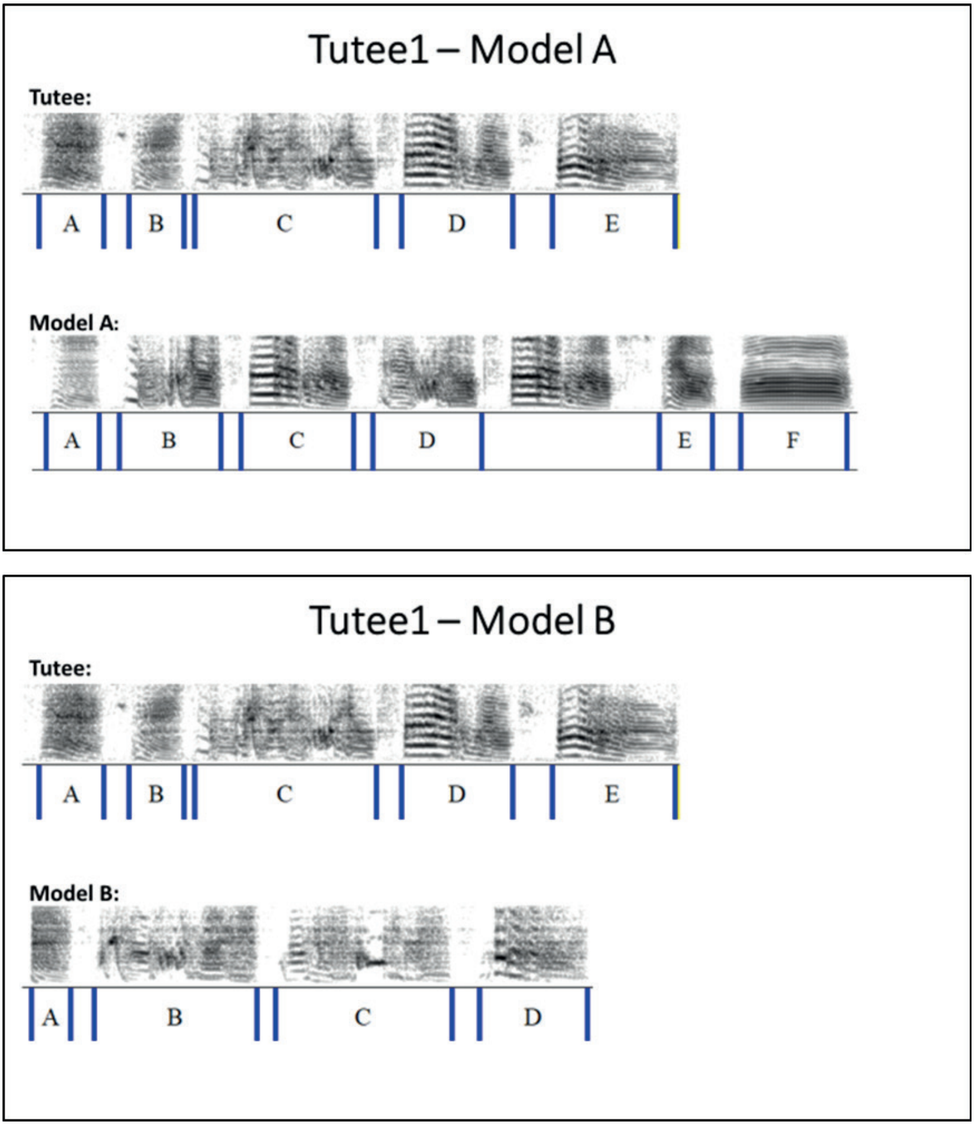


Figure A1. Acoustic transmission properties of the cages measured with two free-field microphones (40BF, preamplifier 26AB, power module 12AA; G.R.A.S. Sound & Vibration) and one speaker (Vifa, Viborg, Denmark) which played tones in frequency steps on 200 Hz. (A) Frequency response for a reference measurement where both microphones were installed in 20cm distance to the speaker. (B) Frequency response where one microphone (red) was in the central compartment and the other one (blue) was in the neighbouring compartment at the same height. (C) Frequency response where one microphone (red) was in the central compartment and the other one (blue) was in the neighbouring compartment in the 'observation hut' behind the one-way mirror. (D) Frequency response where one microphone (red) was in the central compartment and the other one (blue) was in the neighbouring compartment next to the observation hut not covered by the one-way mirror.



Tutee1 – Model A

Tutee:

Model A:

Tutee1 – Model B

Tutee:

Model B:

Figure A2. Example of slides used for human observer similarity scoring.