Seeing voices: the role of multimodal cues in vocal learning

Varkevisser, J.M.

Chapter 6

Thesis summary and general discussion

Bird song is one of the most thoroughly studied animal examples of a vocally learned signal (Catchpole and Slater 1995; Bradbury and Vehrencamp 2011) and often used as a model system for human speech development, because of the many parallels between speech and bird song (Doupe and Kuhl 1999; Bolhuis et al. 2010). Several songbird species learn less well from audio-only tutor song exposure (so called 'tape tutoring') than from live social tutors (reviewed in Baptista & Gaunt, 1997; Soma, 2011). This might be because live tutoring, unlike audio-only tutoring, enables social tutor-tutee interactions, which are thought to play an important role in the vocal learning process (e.g. Beecher & Burt, 2004; Goldstein, King, & West, 2003; Kuhl, 2003, 2007, but also see Nelson 1997). It is unclear, however, whether and to what extent live tutoring also facilitates song learning because it results in multimodal exposure to a tutor, as tutees can both see and hear their tutor, while audio-only tutoring results in unimodal exposure. In this thesis, I investigated the effect of audio-visual compared to audio-only exposure to a tutor on song learning in zebra finches, a songbird species often-cited for learning less well from audio-only tutors than from live tutors (Eales 1989; Derégnaucourt et al. 2013; Chen et al. 2016). In this chapter, I will summarize and discuss the results of the studies described in this thesis and indicate what future research can further improve our knowledge on the effect of multimodal tutor exposure on vocal learning.

Open issues from previous (zebra finch) song tutoring studies

To get more insight into the factors playing a role in the vocal learning process, the effect of different tutoring paradigms on birdsong learning has been studied extensively, especially in zebra finches. Like multiple other songbird species, zebra finches learn more from a social, live tutor than from audio-only exposure to tutor song (Eales 1989; Derégnaucourt et al. 2013; Chen et al. 2016). Several studies have investigated the effect of specific dimensions on zebra finch song learning in order to find out what facilitates song learning from live tutors compared to audio-only song exposure (e.g. Adret, 1993; Bolhuis, van Mil, & Houx, 1999; Houx & ten Cate, 1999a). Based on the outcomes of these studies, it is now often hypothesized that social interaction with a tutor is the key facilitating aspect of live compared to audio-only tutoring (e.g. Chen, Matheson, & Sakata, 2016; Derégnaucourt, Poirier, Kant, & Linden, 2013; Slater, Eales & Clayton, 1988). In Chapter 2, previous zebra finch song tutoring studies were reviewed to find out whether they have systematically controlled for multi- and unimodal tutoring while studying the importance of social interaction for zebra finch song learning. In almost all studies, tutees with multimodal tutor exposure could socially interact with their tutor, while tutees with unimodal tutor exposure could not socially interact with their tutor (Chapter 2). Studies

thus usually confounded 'multimodal' and 'social' tutoring. Social tutoring tends to lead to improved song learning compared to non-social tutoring, but as social and multimodal were confounded, this might partly be due to a facilitating effect of multimodal exposure to a tutor. Another systematic difference between live and audio-only tutoring studies was the social environment of the tutees during tutoring. While audio-only tutored birds were usually housed in social isolation during tutoring, live tutored birds had the tutor as a social companion. This makes it unclear whether the lower amount of song copying from audio-only tutors might partly be attributed to an adverse effect of social isolation on song learning in the tape tutored tutees (Chapter 2). The song tutoring experiments described in this thesis were therefore aimed at testing the effect of multi- versus unimodal tutor exposure, while tutees in the different tutoring conditions were housed in comparable social environments during tutoring.

Song tutoring experiments comparing audio and audio-visual tutor exposure
The first tutoring experiment of this thesis, described in Chapter 3, was designed to investigate whether multi- compared to unimodal exposure to a live tutor would facilitate zebra finch song learning. To this end, zebra finch tutees were offered visual exposure to an adult tutor through a one-way mirror, in addition to auditory tutor exposure. Song learning in these tutees was compared to that in tutees that were raised in the same cage as the tutor and in tutees that were only auditorily exposed to the tutor. All tutees in this experiment were housed with a female companion. The tutees with multimodal exposure were expected to show improved tutor song copying compared to the tutees with unimodal exposure. The song analysis suggested that the unimodally tutored tutees had copied less tutor song than the tutees from the other groups, although the difference was not significant. I also found that the multimodally tutored tutees differed in their song ontogeny from the unimodally tutored tutees: more changes occurred after 65 days post-hatching in the song of the audio-only tutored birds than in that of the live tutored birds, while the audio-visually tutored birds did not differ from the live tutored birds. Although these results could be interpreted to support that multimodal tutor exposure facilitates song learning, an alternative explanation could be that visual feedback from the tutor in response to the tutees' vocalizations had facilitated song learning. To offer multimodal tutor exposure without the possibility of the tutor providing visual feedback to the tutees, I used artificial tutors in the other tutoring experiments described in chapter 4 and 5 of this thesis.

The tutoring experiment described in Chapter 4, investigated song learning in tutees that could see a video of the tutor singing the song that they were at the

same time auditorily exposed to. I compared these tutees to tutees that were only auditorily exposed to song and to tutees that heard song while they were exposed to the same tutor video, but here the pixels were randomized and the frames were played in reversed order. Again, all tutees were housed with a female companion. I expected that the tutees that were exposed to the normal video in addition to song playback would show improved song learning compared to the audio with the pixelated video and audio-only tutoring conditions. The results, however, did not show that the tutor videos led to improved song learning, even though the tutees in the condition with the normal tutor video were attracted most by the stimulus presentation. The videos used in this experiment were adjusted to zebra finch vision with state-of-the-art techniques, but it might be that certain properties of the videos, such as the brightness, negatively affected the birds' acceptance of the videos as conspecific tutors. Additionally, the lack of three-dimensionality in the videos might have made the visual cues less salient. Therefore a three-dimensional robotic zebra finch (Robo- Finch) was used for the visual stimulation in Chapter 5.

In the experiment described in Chapter 5, I investigated song learning in tutees that were exposed to the playback of pre-recorded tutor song, while a RoboFinch was simultaneously producing the beak and head movements that normally accompany the production of this song. These tutees were compared to tutees exposed to the same tutor song without a RoboFinch present and to tutees exposed to a RoboFinch that started moving after song playback had finished. These tutees were all housed in social isolation, and to investigate whether that affected their song learning outcomes, I also included a condition in which tutees were housed with a female companion while they were only auditorily exposed to the tutor song. The expectation was that the visual cues that were synchronized with the auditory song presentation would lead to improved song learning compared to the other tutoring conditions. However, I did not find any significant effects of the visual cues on song learning success. There was an effect of the social companion during tutoring on song learning outcomes: the tutees that had only auditorily been exposed to tutor song while housed with a social companion sang with a higher between-motif stereotypy than the tutees that had been housed solitarily throughout song tutoring.

In the following paragraphs, I will discuss what the results of these song tutoring experiments suggest about the effect of a social companion and the effect of audio-visual versus audio-only tutor exposure on song learning.

The effect of having a social companion during tutoring on song learning
In previous zebra finch song tutoring studies, significantly higher song learning success was found in live than in audio-only tutored tutees. However, in these studies the live tutored tutees had the tutor as a social companion, while the audio-only tutored tutees where housed in social isolation (e.g. Eales 1989; Derégnaucourt et al. 2013; Chen et al. 2016). In the song tutoring experiment described in Chapter 3, I compared song learning in audio-only and live tutored tutees that were both socially housed with a female companion (who does not sing) during song tutoring. With all tutees socially housed, the song of the live tutored tutees was not significantly more similar to the tutor song than the song of the tutees that were only auditorily exposed to the tutor. This suggests that the social isolation of the audio-only tutored tutees in previous studies might have contributed to the difference in song learning success between audio-only and live tutored tutees. The tutoring conditions did not lead to significant differences between the groups, but out of the three tutoring conditions in Chapter 3, the tutees from the live tutoring condition copied most from the tutor, which is in line with previous studies showing more learning from live than audio-only tutors and which suggests that the previously found difference between live and audio-only tutored tutees cannot solely be attributed to the difference in the social environment of the tutees during tutoring.

In the experiment described in Chapter 5, song learning from pre-recorded audio-only song playback was compared in male tutees that were housed in social isolation and in male tutees that were housed with a female companion during the tutoring period. The amount of tutor song copied did not differ between these tutees. Between-motif stereotypy, however, was higher in the tutees tutored with a female companion than in the tutees that were tutored in social isolation (Chapter 5). Song learning outcomes can thus be affected by whether zebra finches are housed with a social companion or in social isolation during tutoring. In future studies, it is therefore important to make sure that birds tutored in different tutoring conditions are housed in comparable social environments during the tutoring phase.

Comparing audio-visual and audio-only tutoring conditions
To investigate song learning from audio-visual and audio-only tutors, three tutoring experiments were conducted in which tutees in an audio-only condition were presented with tutor song auditorily only, while tutees in an audio-visual condition received the exact same song exposure auditorily while being visually exposed to either the live tutor producing this song (Chapter 3), a two-dimensional video of the tutor producing this song (Chapter 4) or a three-dimen-

sional robot tutor producing the beak and head movements accompanying the production of this song (Chapter 5). The birds that thus received audio-visual tutoring were unable to have visual social interactions with their tutors, and therefore the effect of audio-visual tutor exposure could be investigated independent of the effect of social tutor-tutee interactions.

In these tutoring studies, song learning success in the different treatments was assessed by comparing the adult song of the tutees to the song of their tutor. The findings in the experiment described in Chapter 3 suggested that tutees with audio-visual exposure to a live tutor tended to have a higher tutor song learning success than tutees with audio-only exposure to a live tutor: the song of the audio tutees tended to show the lowest, and the song of the live tutees the highest similarity with the tutor song, while the audio-visual tutees showed an intermediate level of similarity. Conversely, the audio group tended to show the highest similarity with the song of their father, which they were exposed to before the experimental tutoring. In the tutoring experiments described in Chapter 4 and Chapter 5, the audio-visual tutoring conditions did not lead to improved tutor song copying compared to the audio-only tutoring conditions. Across the three experimental methods described in this thesis, multimodal exposure to a live tutor thus seemed to have induced higher song learning success than unimodal exposure, while multimodal exposure to artificial tutors did not lead to improved song learning success compared to unimodal exposure.

To study the effect of audio-visual or audio-only tutoring on the timing of song learning, tutee song was recorded at two different moments in time: once at 65 days post-hatching, which is still during song development, and once after 100 days post-hatching, when song is normally crystallized (Gobes et al. 2017). To find out whether tutor groups differed in how 'developed' song already was at 65 days post-hatching when compared to song at 100 days, I recorded the motifs produced by the tutees at these two moments. In the tutoring experiment described in Chapter 3, more changes after 65 days occurred in the audio-only tutored birds than in live tutored birds, while the audio-visually tutored birds did not differ from live tutored birds in this respect. This is in line with an earlier finding demonstrating that zebra finch tutees that were exposed to a tutor only auditorily change their song up to a later age than tutees reared together with tutors in aviaries (Morrison & Nottebohm, 1993). The conclusion of this earlier study was that the closing of the sensitive period depends on whether a bird was able to have visual social interaction with a tutor. In the experiment in Chapter 3, however, I did not find a difference in the amount of changes between the live and the audio-visually tutored group, even though the audio-vis-

ually tutored group could not have visual social interaction with the tutor. This suggests that the timing of song development might not only be influenced by visual social interaction, but also by mere visual exposure to the tutor. The song produced by the tutees from the experiments in Chapter 4 and 5 were unfortunately still too variable at 65 days to use it for further analyses. This might have had to do with the tutoring through passive play-back of pre-recorded tutor song in these studies instead of the tutoring by a live conspecific in Chapter 3 that enabled vocal tutor-tutee interaction, which might affect song learning (Chapter 2). It is possible that the tutoring conditions in Chapter 4 and 5 did not lead to differences in the amount of tutor song copied by the tutees, but did affect the course of song development. Unfortunately, the current data do not allow a conclusion on whether this was the case. The effect of multi- versus unimodal tutoring on the timing of song development found in Chapter 3 shows that it is worthwhile to record both subadult and adult song of zebra finch tutees in tutoring experiments, as it can demonstrate whether different tutoring conditions affect the time course of vocal development. Future studies should address this, using a method that can assess how developed highly variable subadult song is.

In addition to effects on song learning, possible effects of uni- or multimodal tutoring on tutee behaviour were investigated. In Chapter 3, I tested how much time tutees in the audio and audio-visual condition spent in the observation huts, which was the location from which the tutees in the audio-visual, but not in the audio-only condition could see the tutor. Overall, tutees spent a higher proportion of time in the observation huts than expected. However, the tutees from the audio-visual and audio group spent a comparable amount of time in the huts. The possibility to see the tutor thus did not lead to an increase in hut visits. Although the video tutoring experiment described in Chapter 4 did not demonstrate a difference in song learning, it did show that tutee behaviour during song presentation was affected by the different tutoring conditions. The condition where auditory song presentation was accompanied by a video of the tutor producing the song was most salient to the tutees, but this did not lead to increased song learning, as mentioned before, possibly due to insufficient quality of the visual stimulus. In another tutoring study, zebra finch tutees spent more time on the perch next to a visual stimulus (a taxidermic mount of an adult male zebra finch) during than before its exposure, but the presentation of this visual stimulus also did not facilitate song learning (Houx and ten Cate 1999a). This suggests that visual stimulation presented together with auditory song presentation affects tutee behaviour, but not necessarily song learning success.

In the experiments described in Chapter 4 and 5, a control condition was included in which tutees were raised with visual stimulation that had no rhythmic correspondence with the auditory stimulation. These conditions were included to investigate whether non-social, non-sound-contingent visual stimulation would affect song learning differently than sound-contingent visual stimulation (namely the beak and body movements normally accompanying song production). In Chapter 4, for this condition a video was used in which the pixels were randomized and the frames were played in reversed order. In Chapter 5, tutees in this condition were raised with a RoboFinch that started moving after auditory song playback had finished. We expected that the synchronized audio-visual conditions would improve song learning more than these control conditions, for instance because nightingales show improved learning from song playbacks presented with a synchronously flashing stroboscope (Hultsch et al. 1999). Contrary to our expectation, the control conditions did not affect song learning outcomes differently from the 'normal' audio-visual conditions. However, in Chapter 4, tutees spent more time close to the normal tutor video than to the pixelated and reversed video, suggesting that social, sound-specific visual stimulation might be more salient to tutees than non-social, non-sound-specific visual stimulation. Likewise, in humans, sound-specific motor gestures have been found to attract the attention of infants more than unspecific gestures (Kuhl and Meltzoff 1982; Patterson and Werker 1999) and several other studies in animals have shown effects of correctly synchronized visual and acoustic information on perceptual salience (e.g. Taylor et al. 2011; Rȩk 2018).

Effects of audio-visual tutoring on vocal learning

Based on the literature, I hypothesized that audio-visual tutor exposure would lead to improved song learning compared to audio-only tutor exposure (reviewed in Chapter 2). Although the results of Chapter 3 were in line with this hypothesis, this hypothesis was not supported by the results of Chapter 4 and 5. However, Chapter 3 used a live conspecific tutor, while Chapter 4 and 5 used artificial tutors that had not been used in previous tutoring studies. It is thus unclear to what extent methodological decisions, such as the amount and timing of song playback and the stereotypy of song presentation, affected song learning outcomes. It is also unclear whether the visual quality of these tutors was sufficient to affect song learning. However, in the context of imprinting, learning of an auditory signal in chickens was enhanced when simultaneously with the presentation of the auditory signal a rotating box was shown (van Kampen and Bolhuis 1991, 1993). Moreover, young nightingales learn songs from audio playbacks combined with stroboscope light flashes better than

songs presented as audio-only playbacks (Hultsch et al. 1999). This suggests that other bird species can show improved learning of auditory signals when these are paired with any moving visual stimulation. Another difference between the experiment in Chapter 3 on the one hand, and the experiments described in Chapter 4 and 5 on the other, is that the latter experiments were carried out in sound attenuated chambers in which tutees did not hear anything except for their own vocalizations and the tutor song. One of the advantages of multi- compared to unimodal signalling is that multimodal signals are more likely to be detected by receivers than unimodal signals (reviewed in Rowe, 1999). In the experiments in the sound attenuated chambers, it was very unlikely that the tutees did not detect the tutor song. It might thus be that the facilitating effect of the visual cues in addition to auditory song presentation would have been stronger in a noisier environment, in which the detection probability of tutor song would be lower. Likewise, for human speech, visual exposure to speakers' mouth movements contributes to speech intelligibility especially in noisy environments (Sumby and Pollack 1954; Middelweerd and Plomp 1987).

The results of Chapter 3, however, suggest that visual exposure to a tutor can affect song development. From this chapter it is unclear by which mechanism visual tutor exposure might have affected song learning. For instance, it is possible that the tutor gave visual feedback to tutee vocalizations. In other studies, visual feedback contingent on tutee vocalizations was found to improve zebra finch song development (Carouso-Peck and Goldstein 2019). It is, however, also possible that exposure to the visual cues accompanying song production, such as beak and throat movements, affected song learning. For instance, exposure to these visual cues might have drawn the tutee's attention to the auditory signal, as the detectability of a signal can be enhanced if it is presented at the same time as an additional stimulus in another sensory modality (Feenders, Kato, Borzeszkowski, & Klump 2017; reviewed in Rowe 1999). Likewise, in second language learning in human adults, audio-visual training, where mouth and lip movements associated with unfamiliar speech sounds are visible, improves the perception and production of these speech sounds more than audio-only training (e.g. Badin, Tarabalka, Elisei, & Bailly, 2010; Hazan, Sennema, Iba, & Faulkner, 2005; Hirata & Kelly, 2010; Liu, Massaro, Chen, Chan, & Perfetti, 2007; Wang, Hueber, & Badin, 2014). Unlike tape tutors, live tutors can provide visual feedback to tutee vocalizations and provide exposure to sound-production accompanying visual cues. This suggests that besides social tutor-tutee interaction, other mechanisms might play a role in the vocal learning process and might contribute to the difference in song learning suc-

cess from live and audio-only tutors.

Several songbird species learn less well from audio-only than from live social tutors and in many taxonomic groups, the simultaneous presentation of two stimuli in different modalities has been shown to improve signal perception compared to the presentation of one stimulus (reviewed in Rowe, 1999). This suggests that in general, audio-visual exposure to a vocalizing tutor might facilitate vocal learning compared to audio-only exposure. It is important to note, however, that not all songbird species learn less well from tape tutors than from live tutors (reviewed in Baptista & Gaunt, 1997). Future research could investigate whether there is a correlation between the ecology or song characteristics of different songbird species and whether these species learn less well from audio-only playback than from live tutors. For instance, as suggested by Slater et al. (1988), visual cues might be mainly of importance in species with quiet vocalizations, that can only be perceived when tutees are close to a tutor. This type of research might help in forming hypotheses concerning why certain species learn equally well from audio-only exposure to vocalizations as from live tutors, while others do not.

Suggestions for further research

During this research, I identified several open questions that I think should be addressed in further studies. First of all, in the tutoring studies described in this thesis, song learning success in the different tutoring conditions was assessed by determining the similarity between tutee and tutor song. This similarity was calculated using three different methods (visual spectrogram comparisons by human observers and similarity assessment by Luscinia and Sound Analysis Pro software), that all have previously been used to assess song learning success in zebra finches. Up till now, however, these three methods had not been used and compared with the same dataset. The results of the different methods were not very highly correlated, suggesting that the methods pick up different aspects of song similarity. Future research should look into these differences and aim to find out which method best represents sound similarity perception by zebra finches. In future song tutoring studies, that method should then be used to assess song learning success. This thesis mainly focussed on the effect of multi- or unimodal tutor exposure on the auditory component of song production. Future studies could investigate whether multi- or unimodal tutoring affects the visual component of song production. For instance, it could be assessed whether the previously found similarity between the beak movements of tutees and tutors (Williams, 2001) is affected by whether tutees had audio-only or audio-visual exposure to their tutor during the sensitive phase for song

learning.

Second, the artificial tutoring paradigms used in this thesis offer many possibilities for future research. However, first more research is needed into the effect of different methodological choices concerning these artificial tutors on song learning outcomes. For instance, the RoboFinch used in the robot tutoring experiment (Chapter 5) offers many possibilities for further research into multimodal communication and social interactions. The robotic zebra finch can also be used to study the process of multimodal integration in zebra finches, by offering a slight spatial or temporal mismatch between the auditory and visual information and investigating whether this affects zebra finch behaviour compared to a situation without a mismatch (as has been done in dart-poison frogs: Narins, Grabul, Soma, Gaucher, & Hödl, 2005 and pied currawongs: Lombardo, MacKey, Tang, Smith, & Blumstein, 2008).

This thesis focussed on the effect of visual cues on song production learning in male zebra finches. Song production learning only occurs in males, but both male and female zebra finches develop a preference for songs heard early in life over unfamiliar songs, no matter whether they have heard this song from a live (Riebel, Smallegange, Terpstra, & Bolhuis, 2002) or tape tutor (Holveck & Riebel, 2014; Houx & ten Cate, 1999a, b; Riebel, 2000). So far, however, no studies have investigated whether visual cues that are presented in addition to auditory song presentation affect song preference learning, for instance when it comes to the strength of the preference for a particular song. Carrying out the experiments described in this thesis with both male and female tutees, and assessing both song production and preference learning, can shed light on whether visual cues affect both processes equally.

Conclusions

To conclude, the studies in this thesis have demonstrated that multi- versus unimodal exposure to a live tutor can affect the timing of vocal development and possibly also the amount of vocal learning. Multimodal exposure to artificial tutors affected tutee behaviour and made stimulus presentation more salient, but did not affect the song learning outcomes assessed in the experiments in this thesis. These were, however, the first studies using these artificial tutors and future studies should, therefore, further investigate how properties of these artificial tutors affect song learning. I also found that song learning outcomes can be affected by the social environment in which tutees are housed during tutoring. Multi- versus unimodal tutoring and social housing versus social isolation during tutoring might have played a role in the difference in song learning out-

comes found in previous studies comparing live and tape tutoring paradigms. Future studies should be aware of the possible influences of multimodal tutor exposure and the social context on vocal development.

References

Adret P (1993) Operant conditioning, song learning and imprinting to taped song in the zebra finch. Anim Behav 46:149–159

Badin P, Tarabalka Y, Elisei F, Bailly G (2010) Can you "read" tongue movements? Evaluation of the contribution of tongue display to speech understanding. Speech Commun 52:493–503. https://doi.org/10.1016/j.specom.2010.03.002

Baptista LF, Gaunt SLL (1997) Social interaction and vocal development in birds. In: Snowdon CT, Hausberger M (eds) Social influences on vocal development. Cambridge, Cambridge University Press, pp 23–40

Beecher MD, Burt JM (2004) The role of social interaction in bird song learning. Curr Dir Psychol Sci 13:224–228. https://doi.org/10.1111/j.0963-7214.2004.00313.x

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

Bolhuis JJ, Okanoya K, Scharff C (2010) Twitter evolution: Converging mechanisms in birdsong and human speech. Nat Rev Neurosci 11:747–759. https://doi.org/10.1038/nrn2931

Bradbury JW, Vehrencamp SL (2011) Principles of animal communication. Sinauer Associates, Sunderland

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

Catchpole CK, Slater PJB (1995) How song develops. In: Catchpole CK, Slater PJB (eds) Bird Song: Biological Themes and Variations. Cambridge: Cambridge University Press., pp 45–69

Chen Y, Matheson LE, Sakata JT (2016) Mechanisms underlying the social enhancement of vocal learning in songbirds. Proc Natl Acad Sci 201522306. https://doi.org/10.1073/pnas.1522306113

Derégnaucourt S, Poirier C, van der Kant A, van der Linden A (2013) Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. J Physiol 107:210–218. https://doi.org/10.1016/j.jphyspar is.2012.08.003

Doupe AJ, Kuhl PK (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 LA (1989) The influences of visual and vocal interaction on song learning in zebra finches. Anim Behav 37:507–508. https://doi.org/10.1016/0003-3472(89)90097-3

Feenders G, Kato Y, Borzeszkowski KM, Klump GM (2017) Temporal ventriloquism effect in european starlings: evidence for two parallel processing pathways. Behav Neurosci 131:337–347. https://doi.org/10.1037/bne0000200

Gobes SMH, Jennings RB, Maeda RK (2017) The sensitive period for auditory-vocal learning in the zebra finch: consequences of limited-model availability and multiple-tutor paradigms on song imitation. Behav Processes 163:5–12. https://doi.org/10.1016/j.beproc.2017.07.007

Goldstein MH, King AP, West MJ (2003) Social interaction shapes babbling: testing parallels between birdsong and speech. Proc Natl Acad Sci U S A 100:8030–5. https://doi.org/10.1073/pnas.1332441100

Hazan V, Sennema A, Iba M, Faulkner A (2005) Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. Speech Commun 47:360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hirata Y, Kelly SD (2010) Effects of lips and hands on auditory learning of second-language speech sounds. J Speech Lang Hear Res 53:298–310. https://doi.org/10.1044/1092-4388(2009/08-0243)

Holveck MJ, Riebel K (2014) Female zebra finches learn to prefer more than one song and from more than one tutor. Anim Behav 88:125–135. https://doi.org/10.1016/j.anbehav.2013.11.023

Houx BB, ten Cate C (1999a) Do stimulus-stimulus contingencies affect song learning in zebra finches (*Taeniopygia guttata*)? J Comp Psychol 113:235–242. https://doi.org/10.1037/0735-7036.113.3.235

Houx BB, ten Cate C (1999b) Song learning from playback in zebra finches: is there an effect of operant contingency? Anim Behav 57:837–845. https://doi.org/10.1006/anbe.1998.1046

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

Kuhl PK (2007) Is speech learning "gated" by the social brain? Dev Sci 10:110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

Kuhl PK (2003) Human speech and birdsong: communication and the social brain. Proc Natl Acad Sci U S A 100:9645–9646. https://doi.org/10.1073/pnas.1733998100

Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. Science (80-) 218:1138–1141. https://doi.org/10.1126/science.7146899

Liu Y, Massaro DW, Chen TH, et al (2007) Using visual speech for training chinese pronunciation: an in-vivo experiment. SLaTE Work Speech Lang Technol Educ ISCA Tutor Res Work Summit Inn, Farmington, Pennsylvania USA 29–32

Lombardo SR, MacKey E, Tang L, et al (2008) Multimodal communication and spatial binding in pied currawongs (*Strepera graculina*). Anim Cogn 11:675–682. https://doi.org/10.1007/s10071-008-0158-z

Middelweerd MJ, Plomp R (1987) The effect of speechreading on the speech-recep-

tion threshold of sentences in noise. J Acoust Soc Am 82:2145–2147. https://doi.org/10.1121/1.395659
Narins PM, Grabul DS, Soma KK, et al (2005) Cross-modal integration in a dart-poison frog. Proc Natl Acad Sci U S A 102:2425–2429. https://doi.org/10.1073/pnas.0406407102
Nelson D (1997) Social interaction and sensitive phases for song learning: A critical review. In: Snowdon CT, Hausberger M (eds) Social influences on vocal development. Cambridge, Cambridge University Press, pp 7–22
Patterson ML, Werker JF (1999) Matching phonetic information in lips and voice is robust in 4.5-month-old infants. Infant Behav Dev 22:237–247. https://doi.org/10.1016/S0163-6383(99)00003-X
Ręk P (2018) Multimodal coordination enhances the responses to an avian duet. Behav Ecol 29:411–417. https://doi.org/10.1093/beheco/arx174
Riebel K (2000) Early exposure leads to repeatable preferences for male song in female zebra finches. Proc R Soc London Ser B Biol Sci 267:2553–8. https://doi.org/10.1098/rspb.2000.1320
Riebel K, Smallegange IM, Terpstra NJ, Bolhuis JJ (2002) Sexual equality in zebra finch song preference: evidence for a dissociation between song recognition and production learning. Proc R Soc London Ser B Biol Sci 269:729–33. https://doi.org/10.1098/rspb.2001.1930
Rowe C (1999) Receiver psychology and evolution of multicomponent signals. Anim Behav 58:921–931. https://doi.org/10.1006/anbe.1999.1242
Slater PJB, Eales LA, Clayton NS (1988) Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. Adv study Behav 18:1–34. https://doi.org/10.1016/S0065-3454(08)60308-3
Soma MF (2011) Social factors in song learning: a review of Estrildid finch research. Ornithol Sci 10:89–100. https://doi.org/10.2326/osj.10.89
Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26:212–215. https://doi.org/10.1121/1.1907309
Taylor RC, Klein BA, Stein J, Ryan MJ (2011) Multimodal signal variation in space and time: how important is matching a signal with its signaler? J Exp Biol 214:815–820. https://doi.org/10.1242/jeb.043638
van Kampen HS, Bolhuis JJ (1993) Interaction between auditory and visual learning during filial imprinting. Anim Behav 45:623–625. https://doi.org/10.1006/anbe.1993.1074
van Kampen HS, Bolhuis JJ (1991) Auditory learning and filial imprinting in the chick. Behaviour 117:303–319. https://doi.org/10.1163/156853991X00607
Wang X, Hueber T, Badin P (2014) On the use of an articulatory talking head for second language pronunciation training: the case of Chinese learners of French. 10th Int Semin Speech Prod 449–452