The Company of **Biologists**

RESEARCH ARTICLE

Lingual articulation in songbirds

Roderick A. Suthers[1],*, John R. Rothgerber[1] and Kenneth Kragh Jensen[2]

ABSTRACT

Lingual articulation in humans is one of the primary means of vocal tract resonance filtering that produces the characteristic vowel formants of speech. In songbirds, the function of the tongue in song has not been thoroughly examined, although recent research has identified the oropharyngeal–esophageal cavity as a resonance filter that is actively tuned to the frequency of the song. In northern cardinals (*Cardinalis cardinalis*), the volume of this cavity is inversely proportional to the frequency of the song above 2 kHz. However, cardinal song extends below this range, leaving the question of whether and how the vocal tract is tracking these low frequencies. We investigated the possible role of the tongue in vocal tract filtering using X-ray cineradiography of northern cardinals. Below 2 kHz, there was prominent tongue elevation in which the tip of the tongue was raised until it seemed to touch the palate. These results suggest that tongue elevation lowers the resonance frequency below 2 kHz by reducing the area of the passage from the oral cavity into the beak. This is consistent with a computational model of the songbird vocal tract in which resonance frequencies are actively adjusted by both changing the volume of the oropharyngeal–esophageal cavity and constricting the opening into the beak.

KEY WORDS: Tongue, Birdsong, Vocal tract filter

INTRODUCTION

Birdsong provides a valuable model system in which to study the production and perception of complex vocalizations that have been acquired through vocal learning (Hultsch and Todt, 2004). In most species, little is known about the underlying articulatory movements that form the song motor patterns (Elemans, 2014). However, a recent computational acoustic model of the songbird vocal tract has highlighted the role of the air-filled oropharyngeal–esophageal cavity (OEC) as a resonance filter actively tuned to the frequency of the song (Fletcher et al., 2006; Riede et al., 2006). Building on this model, we here report the first measurements of lingual articulation from an intact songbird during spontaneous song. This articulation causes a change in the area of the opening from the OEC into the beak, which broadens the predicted frequency range of vocal tract resonance beyond that identified by OEC volume alone.

The avian vocal organ, the syrinx, lies at or near the junction between the two primary bronchi and the caudal end of the trachea (King, 1989). Respiratory muscles provide the driving force for phonation, as they do in terrestrial mammals (Elemans, 2014). In Oscine songbirds, sound is produced at the syrinx by flow-induced oscillation of tissue masses called labia (Doupe and Kuhl, 1999; Larsen and Goller, 2002; Riede and Goller, 2010) and each side of

the syrinx has independent motor control of ipsilateral sound production (Suthers, 1990). The rate of labial vibration determines the fundamental frequency (f_0) of the vocalization (Goller and Riede, 2013). The labia are analogous to vocal folds in mammals.

Sound generated in the syrinx travels to the beak through the resonance filters of the supra-syringeal vocal tract. Possible filters include the trachea, glottis, OEC and beak. By changing the dimensions of its resonance filters, the bird can modulate the spectral properties of the syringeal sound. The vocal tract may be manipulated to emphasize or suppress f_0, its harmonics, or the formant structure.

Lingual articulation is one method of altering the resonance properties of the vocal tract in animals. Human production of vowels is based on placement of the tongue (Titze, 1994). The two lowest formants are thought to determine how vowels are perceived. A high–low placement of the tongue changes the frequency of the first formant, and a front–back placement changes the frequency of the second formant. Although little is known about their importance in avian communication, formants or harmonic-rich formant-like structures are also observed in the vocalizations of various bird species, including zebra finches (Ohms et al., 2010), mynah birds (Klatt and Stefanski, 1974), African gray parrots (Pepperberg, 2010) and monk parakeets (Beckers et al., 2004; Ohms et al., 2012). The ability of mynah birds and parrots to modify formants contributes to their ability to imitate human speech, although the production mechanism may not be identical to that found in humans.

Based on his studies of phonation by orange-winged amazon parrots (*Amazona amazonica*), Nottebohm (1976) speculated that these birds might use their tongue to alter the resonating structures of their nasopharyngeal space to generate speech-like formants. Since then, several studies have looked at the effect that tongue placement has on formant structure in bird vocalizations. Research on speech-imitation by an African gray parrot has suggested that, similar to humans, the bird can change the front–back placement of the tongue to alter the formant properties of the vocalization, but not the high–low placement (Patterson and Pepperberg, 1994). In an experiment in which the syrinx of a euthanized monk parakeet was replaced with a speaker, Beckers et al. (2004) found that artificially manipulating front–back tongue placement caused changes in formant frequency and amplitude, and manipulating high–low placement caused changes in amplitude. Ohms et al. (2012) used X-ray cineradiography to measure tongue height in naturally vocalizing monk parakeets. Changes in tongue height were observed during the production of various natural call notes. For contact notes and greeting notes, a sustained decrease in tongue height was observed for the duration of the note. For chatter sounds, an initial decrease in tongue height was observed at the start of vocalization, followed by a gradual increase back to the original position. Although no direct relationship between tongue height and frequency was found, the patterns observed suggest that the tongue may play an important role in modulating vocal tract resonance.

To the best of our knowledge, there has been no research on the role of the tongue in songbird phonation. However, ongoing

[1]Medical Sciences, School of Medicine, Jordan Hall, Indiana University, 1001 East Third Street, Bloomington, IN 47405, USA. [2]Starkey Hearing Technologies, 6600 Washington Avenue S., Eden Prairie, MN 55344, USA.

*Author for correspondence (suthers@indiana.edu)

Journal of Experimental Biology

List of symbols and abbreviations	
$2f_0$	second harmonic
f_0	fundamental frequency
FM	frequency modulation
LV	distance between the larynx and the second cervical vertebra
OEC	oropharyngeal–esophageal cavity

experiments have provided evidence in support of a model of the songbird vocal tract in which the OEC acts as one of the primary resonance filters (Fletcher et al., 2006). X-ray cineradiography of singing birds has shown that expansion and contraction of the OEC during song is accompanied by cyclical movements of the hyoid skeleton and outward arching of the cervical vertebrae (Riede et al., 2006). Fletcher's model predicts that f_0 of the vocalization is inversely proportional to the square root of OEC volume, such that at low f_0 the OEC is expanded and at high f_0 the OEC is contracted. This has been observed in a number of species including northern cardinals (*Cardinalis cardinalis*; Riede et al., 2006), white-throated sparrows (Riede and Suthers, 2009) and zebra finches (Ohms et al., 2010), and experimental manipulation of the OEC in zebra finches has supported its role as a vocal filter (Riede et al., 2013).

Measurements using X-ray cineradiography of the OEC during song in northern cardinals revealed that OEC volume tracked f_0 within the range of ~2 to 9 kHz (Riede et al., 2006; R.A.S., unpublished data). However, cardinal song, which is characterized by upward or downward frequency modulation (FM) tonal sweeps, has a typical range extending as low as ~0.8 kHz. Between ~0.8 and ~2.0 kHz, the OEC remains relatively unchanged at its maximum volume, leaving the question of whether and how the cardinal vocal tract is tracking these low frequencies. According to Fletcher's OEC model, tongue placement, possibly influenced by beak gape, might act to reduce the effective cross-sectional area of the opening from the OEC to the beak and further reduce the resonance frequency.

In this study, we investigated lingual articulation in northern cardinals and its interaction with vocal tract filters. X-ray cineradiographic analysis revealed prominent tongue elevations at low f_0. We expand the OEC model by providing descriptive evidence that lingual articulation contributes to the resonance filtering properties of the songbird vocal tract at low f_0.

MATERIALS AND METHODS
Subjects
Experiments were conducted on seven adult male northern cardinals, *Cardinalis cardinalis* (Linnaeus 1758), ranging from 2 to 8 years old. These birds were removed from their nests when they were about 1 week post-hatch and raised in the laboratory where they were tutored with both live and digitally recorded song of adult cardinals. As young adults they were housed individually in an aviary where they could hear and see other cardinals, in addition to other species including brown thrashers (*Toxostoma rufum*), brown-headed cowbirds (*Molothrus ater*), white-throated sparrows (*Zonotrichia albicollis*), zebra finches (*Taenopygia guttata*) and Bengalese finches (*Lonchura striata*). Food and water were provided *ad libitum*.

All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee and the Radiation Safety Office of Indiana University and comply with the 'Principles of laboratory animal care', publication no. 86-23, revised 1985, of the National Institutes of Health.

X-ray cinematography and song recordings
Before an experiment, the bird was moved into a 35×38×43 cm cage, which was placed in front of the image intensifier of a Series 9800 mobile C-arm (OEC Medical Systems Inc., GE Healthcare, Piscataway, NJ, USA). The X-ray opaque wire mesh on the side of the cage that was closest to the image intensifier was replaced by an X-ray-transparent 2.5 mm thick sheet of acrylic plastic and the opposite wall of the cage was replaced by a piece of mist net that was invisible to the X-rays and did not reflect sound. The other two sides and top of the cage were wire mesh. A small wooden perch 10 cm long was placed about 15 cm in front of the image intensifier so that the bird's head was centered in the X-ray beam when he was on the perch.

Song was recorded with a directional condenser microphone (model AT835b, Audio Technica, Stow, OH, USA) positioned about 0.5 m from the bird. All walls of the experimental room were covered with 5 cm thick fiber glass core, mylar encapsulated sound absorbing panels (Acoustical Surfaces, Inc., Chaska, MN, USA) to minimize sound reflection.

A lateral view of the supra-syringeal vocal tract during song was recorded by cineradiography of the spontaneously singing bird as it sat on the perch in front of the C-arm image intensifier. It was not possible to record from a frontal view because of the X-ray opaque markers in the tongue (see below) not being visible against the dense skull and beak. During most experiments, the C-arm was operated in a digital cine mode at 30 frames s^{-1} at about 50 kVp and 24 mA. Each frame was produced by one 10 ms X-ray pulse and had an image resolution of 1 k×1 k pixels. In most experiments, the digital signal from the fluoroscope was recorded on a video recorder (Sony GVD-1000 Video Walkman, MiniDV format, 480×480 pixels) and the bird's vocalizations were recorded on the audio channel of the video recorder.

During an experiment, the experimenter was outside the room watching the bird on a video monitor and listening on headphones for song. If the bird sang while sitting in a position that gave a good lateral view of the head, the experimenter turned on the C-arm X-ray tube. Afterwards, relevant sequences of the X-ray movies were digitized and rendered at 30 frames s^{-1} (video) and concurrent vocalizations were digitized at 48 kHz sampling rate using the software Vegas Video, version 5.0 (Sonic Foundry, Madison, WI, USA). All data files were corrected for a recording delay of approximately 114 ms in the video relative to the audio.

In some experiments (birds 433, 483 and 512), the OEC model 9800 C-arm fluoroscope was retrofitted with a high-speed digital camera (Xcitex XC1-M, Xcitex Inc., Woburn, MA, USA) and a Nikon micro-Nikkor 60 mm f/2.8 AF lens (Nikon Corporation, Tokyo, Japan). This camera imaged the output of the image intensifier with a resolution of 1082×1082 pixels at up to 150 frames s^{-1}; 60 or 100 frames s^{-1} were used in the experiments reported here. The camera and video processing were controlled by the software Xcitex ProCapture (version 1.0.2.4). Audio was captured through a National Instruments BNC-2110 connector box connected to a National Instruments PCIe-6323 DAQ (National Instruments Corporation, Austin, TX, USA) and synchronized with the video in the ProCapture software. X-ray image and sound were recorded directly and simultaneously on a computer hard drive and there was no processing delay in the video relative to audio as confirmed by a prior calibration check.

Data analysis
Measurements of tongue elevation, beak gape and distance between the larynx and the second cervical vertebra (LV; used as a measure
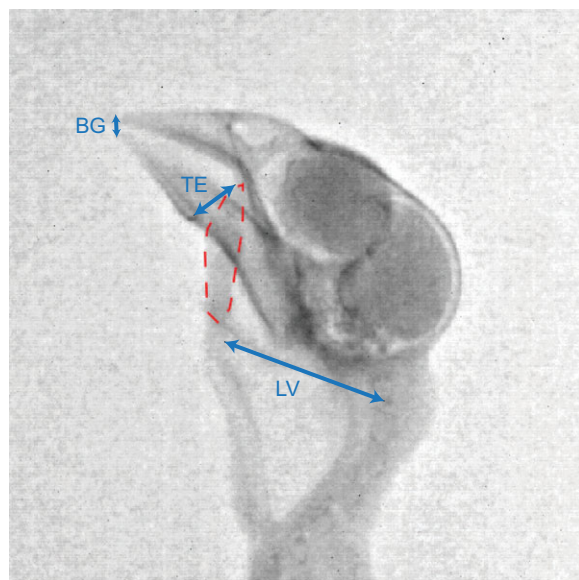
Fig. 1. Measurements of cardinal tongue movement and the upper vocal tract during song. The tongue is outlined by a dashed red line. Tongue elevation (TE) is the distance between the ventral edge of the lower mandible and the ventral edge of the tongue. The larynx–vertebra distance (LV) is the distance between the larynx and the second cervical vertebra. Beak gape (BG) is the distance between the tip of the maxilla and the tip of the lower mandible. This image is a lateral view of a singing cardinal in a frame of an X-ray movie.

measured from the distance between points placed on the tip of the maxilla and the tip of the lower mandible. LV was measured from the distance between points placed on the larynx and the second cervical vertebra. Fig. 1 shows the locations of these points and the distances that were measured. One of two methods was used to calibrate measured distances: (1) a 1.6 mm diameter stainless steel sphere was implanted subcutaneously in the bird's neck near the skull, or (2) the length of the maxilla was measured using calipers. Either of these methods provided a scale from which absolute distances could be computed in MaxTRAQ. To facilitate tracking of the tongue, two pieces of silver wire, about 0.2 mm in diameter and 1–2 mm long, were inserted into the tongues of five birds while under isoflurane anesthesia. These two wires were placed different distances from the tip of the tongue and served as markers that were visible in the X-ray images. Such markers were not necessary to measure tongue movements in experiments using the high-speed digital camera because of the increased resolution and higher frame rate of the camera.

There are several factors inherent in this measurement method that introduce a degree of error when trying to compare syllables across different video clips and across individuals. First, a small change in the angle of the bird's head relative to the camera may result in slightly different measurements for the various dimensions. While care was taken to choose video clips that provided a nearly lateral view, these may still have slight variations, especially when comparing different syllables from different video clips. Second, because of the difficulty of surgically inserting such small wires into the tongue, there was some variation in the placement of the X-ray opaque markers. These pieces of silver wire may also move a small amount within the tongue after being inserted. This could cause a shift in the measurements of one bird relative to another. Third, the determination of absolute distances as described above is prone
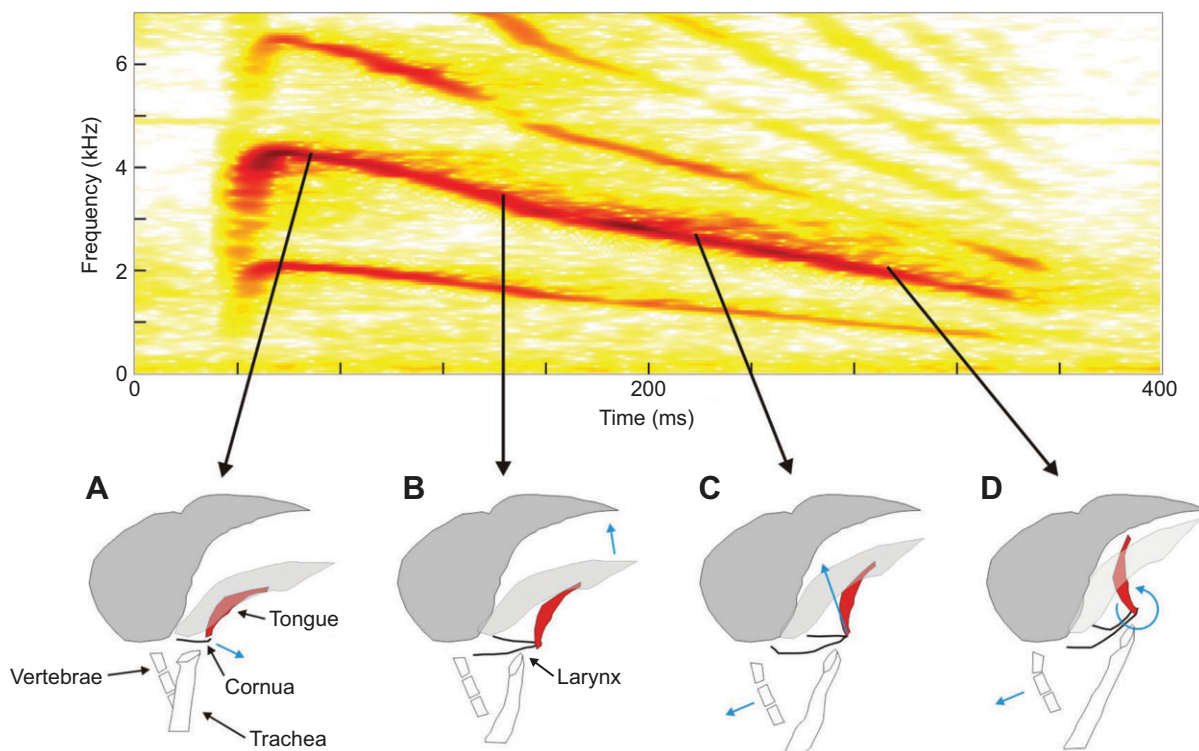
of OEC volume) were obtained using frame-by-frame manual point tracking in MaxTRAQ (version 2.4.0.2; Innovision Systems, Inc., Columbiaville, MI, USA). Tongue elevation was measured from the distance between points placed on the ventral edge of the lower mandible and the ventral edge of the tongue. Beak gape was

Fig. 2. Schematic diagram of vocal tract movements associated with TE during a downward-sweeping frequency modulation (FM) syllable. The vocal tract movements in A–D are for syllable 2 from bird 436 (436-2). Blue arrows indicate the direction of movement. For a description of the movements, see Results.
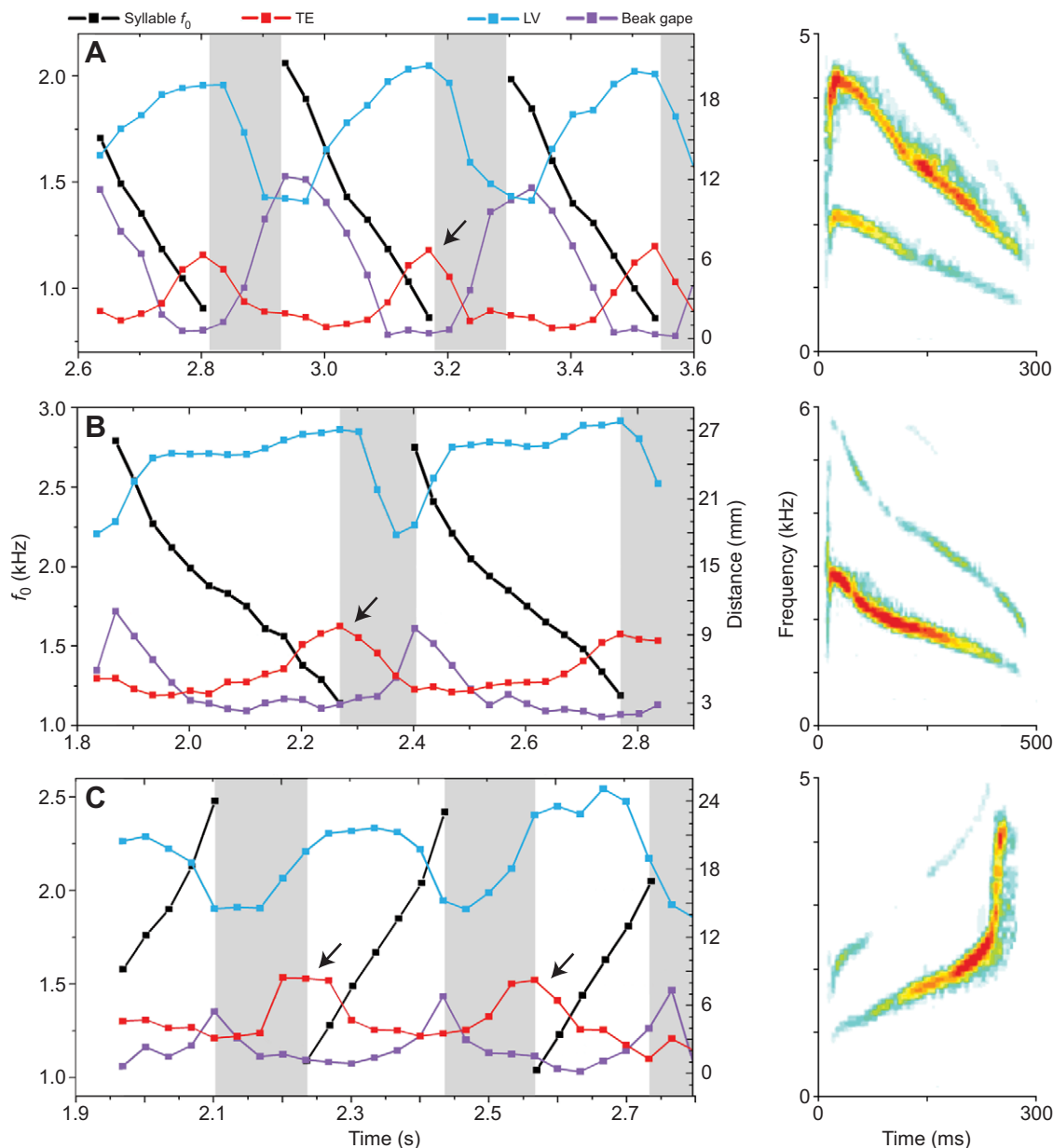
Fig. 3. Vocal tract motor patterns and spectrograms of three different syllables sung by bird 436. (A,B) The fundamental frequency (f_0) of downward-sweeping syllables; (C) the f_0 of an upward-sweeping syllable. The concurrent changes in the dimensions of the OEC are indicated by TE, LV and beak gape. TE, some examples of which are indicated by an arrow, does not occur in these syllables unless the f_0 is below ~1.5 or 2 kHz. This occurs at the end of downward-sweeping syllables (A,B) and at the beginning of upward-sweeping syllables (C). Note that the initial low-frequency part of the first syllable in C was not recorded. The gray shading indicates the recovery pause. A, syllable 436-2; B, syllable 436-4; C, syllable 436-5.

to slight variation because of the difficulty in precisely pinpointing edges in the low-resolution X-ray images; to minimize this variation, two researchers independently measured and then agreed upon the scale value. All measurements were checked by at least two researchers to verify the precision and accuracy of point placement. Nevertheless, the above constraints should be taken into account when comparing absolute values for distance measurements across syllables and across individuals.

f_0 values were calculated using Sound Analysis Pro 2011 (version 2011.100; http://www.soundanalysispro.com; Tchernichovski et al., 2000). Spectrograms were created in Sound Analysis Pro 2011 and Avisoft-SASLab Pro (version 5.2.07; Avisoft Bioacoustics, Glienicke, Germany).

Tongue elevation, LV and beak gape were plotted against f_0 and regression analysis was performed using SigmaPlot (version 11.0; Systat Software, Inc., San Jose, CA, USA). As it was predicted that tongue elevation would correlate with f_0 only below ~2 kHz, piecewise (segmented) linear regression analysis was performed using two segments. Breakpoints between segments were automatically generated in the software and were determined by which provided the best fit.

RESULTS

We examined 12 different syllable types recorded from seven male northern cardinals. Cineradiography of the upper vocal tract in singing male cardinals revealed prominent tongue movements
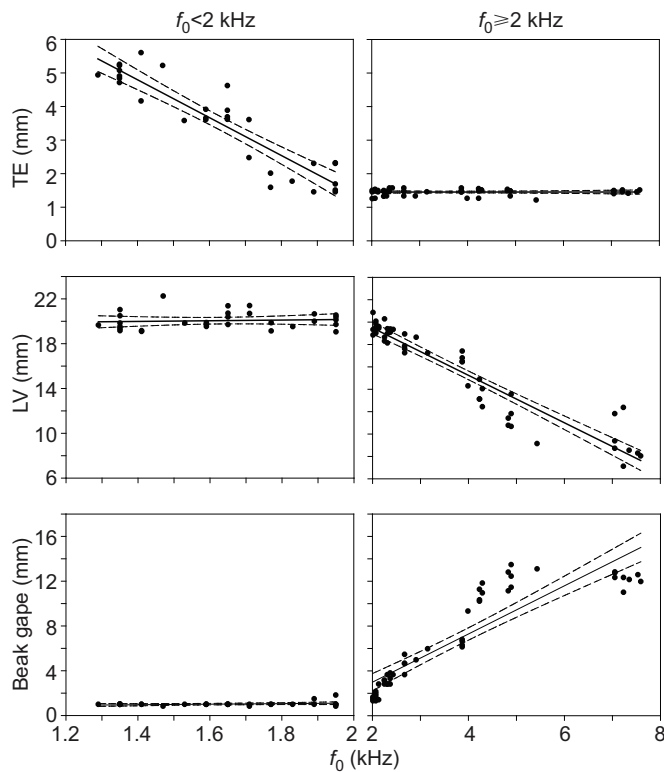
Fig. 4. TE, LV and beak gape as a function of f_0 above and below 2 kHz for syllable 512-2. TE is inversely correlated with f_0 below 2 kHz (r^2=0.85), while LV is inversely correlated with f_0 above 2 kHz (r^2=0.89) and beak gape is correlated with f_0 above 2 kHz (r^2=0.81). This indicates that TE is operating independently of these other mechanisms.

during the production of low-frequency sound: during most song, the tongue remains in the lower mandible, but if the f_0 drops below ~2 kHz, the tip of the tongue rises from its position on the floor of the beak and rotates to a nearly vertical position between the back of the beak and the OEC.

A schematic representation of this articulatory movement is shown in Fig. 2 for a downward-sweeping syllable sung by bird 436. This syllable is unusual in having a second harmonic ($2f_0$) with more acoustic energy than in f_0. This $2f_0$ emphasis does not affect the pattern of tongue movement, which is similar to that of most other syllables analyzed for this study. At the beginning of a downward-sweeping syllable, the beak is open and the tongue rests in the lower mandible (Fig. 2A). As the frequency gradually decreases, the beak gape is reduced and the hyoid apparatus, to which the tongue is attached, begins to move the larynx downward and forward (Fig. 2B). As these movements progress, the beak gape further decreases and the OEC dimensions are increased by the dorsal flexing of the cervical vertebrae and continued anterior movement of the larynx (Fig. 2C). This motor pattern continues as the $2f_0$ approaches a terminal value of 1.5 kHz. The OEC expands to its maximum dimensions and beak gape decreases until the beak is nearly closed. From ~2.0 kHz, the tongue elevates to a vertical orientation at the back of the beak, where it appears to touch, or nearly touch, the palate (Fig. 2D). See Movie 1 for an X-ray recording of a live cardinal displaying these motor patterns in action. The upward-sweeping syllables analyzed for this study follow the same pattern in reverse.

Tongue elevation, beak gape and LV were plotted together with f_0 over time to determine their interaction for three syllable types of

bird 436 (Fig. 3). These plots show that LV is smallest when f_0 is high and largest when f_0 is low, as would be expected if the OEC is tracking the vocal tract resonance (Fig. 3A). Conversely, beak gape is largest when f_0 is high and is nearly closed when f_0 is low (Fig. 3B). The relationship between LV and beak gape for these syllables is consistent with the initial description of cardinal vocal tract acoustics (Riede et al., 2006). In downward-sweeping syllables 436-2 and 436-4 (Fig. 3A,B), tongue elevation occurs at the end of the syllable after the f_0 drops below ~2 kHz. In upward-sweeping syllable 436-5 (Fig. 3C), tongue elevation begins prior to the start of vocalization and continues until the f_0 rises above about 1.5 kHz.

To further investigate the relationship between tongue elevation, OEC expansion, beak gape and f_0 above and below 2 kHz, a series of scatterplots was made for each syllable as in Fig. 4. Because of space constraints, only the linear regression r^2 values are reported here for the remaining syllables (Table 1). Most syllables showed a strong relationship below 2 kHz between tongue elevation and f_0 but no relationship between LV and f_0 or beak gape and f_0. While the tongue was elevating, both LV and beak gape generally remained constant, although a few syllables showed the beak beginning to open from around 1.8 kHz. Such beak gape may have an effect on tongue elevation, but the overlap with tongue elevation was minimal and no correlation was found between the two. Because the hyoid apparatus, which drives OEC expansion, is also responsible for some tongue movement, there is a concern that tongue elevation is simply a by-product of the maximum expansion of the OEC. However, these results show that, with a few exceptions to be discussed below, OEC dimensions do not change below 2 kHz. Syllables showed either no relationship or a weak relationship between tongue elevation and LV (Table 1), further suggesting that tongue elevation is operating independently of OEC expansion. It should be noted that, unlike previous studies on the OEC (e.g. Riede et al., 2006), a frontal X-ray view was not used for the current study because it was not possible to image the tongue from such an angle owing to the relative density of the beak and skull. Therefore, there may be some OEC expansion in the latero-lateral dimension that was not captured, although there is no hint of this in the recordings. Even if further OEC expansion was observed, it would not be possible to establish a causal relationship without experimental manipulation. Should it exist, it may be that such further expansion is a by-product of the elevating tongue, rather than the other way around, or that the two interact to lower the resonance frequency.

As predicted, LV showed a strong relationship with f_0 above 2 kHz for most syllables, while tongue elevation showed no relationship with f_0 above 2 kHz (Table 1). For some syllables, there was a relationship between beak gape and f_0 above 2 kHz, but this was less consistent than for the other variables being measured.

The rate of change of tongue elevation per kHz of f_0 below 2 kHz was calculated for each syllable in order to investigate the degree to which tongue elevation tracks f_0 (Table 2). Similarly, the rate of change of LV per kHz of f_0 above 2 kHz was also calculated (Table 2). The results showed a range from −1.90 to −7.08 mm kHz^{-1} (−4.51±1.79 mm kHz^{-1}, mean±s.d.) for tongue elevation, and −1.00 to −9.83 mm kHz^{-1} (−3.38±2.59 mm kHz^{-1}) for LV. This variation indicates that both tongue elevation and LV may not precisely track f_0. Individual characteristics of syllables, such as the degree of FM or repetition rate, may impose constraints, and there may be other undetected factors at work interacting with these mechanisms to influence the resonance frequencies of the vocal tract.

An example of a syllable that follows an unexpected pattern is given in Fig. 5A (407-4). An analysis of the particular

Table 1. Linear regression r^2 values below and above 2 kHz for all birds

Syllable	Type	f_0 range (kHz)	N	Below 2 kHz				Above 2 kHz			
				TE×f_0	LV×f_0	BG×f_0	LV×TE	TE×f_0	LV×f_0	BG×f_0	LV×TE
407-4	Upward	0.8–3.9	8	(−) **0.67**	(+) **0.60**	0.00	(−) 0.40	(−) 0.12	(−) **0.43**	(+) **0.53**	(+) 0.15
436-2	Downward	1.5–4.3*	10	(−) 0.35	(−) 0.01	(+) 0.32	(+) 0.35	0.00	(−) **0.89**	(+) **0.79**	0.00
436-4	Downward	1.0–2.8	5	(−) **0.87**	(−) 0.17	(+) 0.25	(+) 0.23	(+) 0.29	(−) **0.90**	(+) **0.85**	(−) 0.34
436-5	Upward	0.9–4.0	5	(−) **0.80**	(+) 0.04	(+) 0.12	(−) 0.03	(−) 0.06	(−) **0.55**	(+) **0.68**	(+) 0.19
483-1b	Upward	1.8–3.0	4	(−) **0.85**	0.00	(+) 0.14	(−) 0.40	0.00	(−) **0.40**	(−) 0.05	(+) 0.14
503-1	Upward	0.9–4.1	14	(−) **0.86**	(+) 0.19	(+) 0.20	(−) 0.28	(+) 0.04	(−) **0.41**	(+) 0.17	0.00
503-3	Downward	0.8–2.1	8	(−) **0.75**	(−) 0.06	(+) **0.43**	(+) 0.09	−	−	−	−
503-4	Upward	1.2–7.3	6	(−) **0.80**	(+) 0.21	(+) **0.60**	(−) 0.36	0.00	(−) **0.62**	(+) 0.31	(+) 0.02
512-1	Downward	2.1–5.1	4	−	−	−	−	(+) 0.02	(−) **0.75**	(+) 0.27	(−) 0.01
512-2	Downward	1.3–7.6	10	(−) **0.85**	(+) 0.01	(+) 0.09	0.00	0.00	(−) **0.89**	(+) **0.81**	(+) 0.01
522-2	Upward	1.4–3.3	10	(−) 0.15	(+) 0.13	(+) 0.09	(+) 0.08	(−) 0.19	(−) 0.06	(+) **0.60**	(+) 0.14
433-1	Downward	1.1–2.8	7	(−) **0.83**	(−) 0.03	(+) **0.55**	(+) 0.02	(−) 0.01	(−) **0.95**	(+) **0.49**	0.00

Type indicates the temporal direction of the FM sweep. f_0 range shows the range of fundamental frequency values for the syllable. N, the number of individual syllables included in the analysis. The r^2 values are listed together with a + or – to indicate the direction of the relationship; r^2 values above 0.40 are in bold. TE, tongue elevation; LV, larynx–vertebra distance; BG, beak gape.
*The second harmonic ($2f_0$) range is given for syllable 436-2.

characteristics of this syllable in comparison with similar syllables by other birds reveals a possible explanation. This upward-sweeping syllable shows the expected tongue elevation at the start, which has an inverse relationship with f_0, but unusually it also shows a somewhat strong positive relationship between LV and f_0 below 2 kHz, such that as the frequency of the song increases, the OEC volume also increases. Once the syllable reaches ~2 kHz and continues to rise, the OEC then contracts as expected. At first glance, this seems contradictory to the prediction made by the OEC model, as the OEC is expected to be largest at low f_0. However, an almost identical syllable to this is also sung by two other birds: syllable 436-5 (Fig. 3C) and syllable 503-1 (Fig. 6A). Both of these syllables differ from 407-4 in that the OEC is fully or almost fully expanded at low f_0. Although the acoustic properties of these syllables are similar, they each differ in respect to the duration of the recovery pause between each syllable in the sequence (represented by the shaded gray area in Figs 3 and 5). Syllable 407-4 has a pause of only ~62 ms, while syllables 503-1 and 436-5 have pauses of ~105 and ~135 ms, respectively. With such a short recovery pause, bird 407 may be unable to fully expand the OEC in preparation

Table 2. Rate of change of TE against f_0 below 2 kHz and LV against f_0 above 2 kHz

Syllable	Rate of change (mm kHz^{-1})	
	TE <2 kHz	LV >2 kHz
407-4	−1.90	−1.40
436-2	−2.84	−4.66
436-4	−7.08	−9.83
436-5	−5.30	−2.80
483-1b	−6.40	−2.20
503-1	−3.82	−1.84
503-3	−3.10	−
503-4	−4.70	−1.20
512-1	−	−3.50
512-2	−5.60	−2.10
522-2	−2.10	−1.00
433-1	−6.80	−6.60

Values were taken from the slopes of regression equations. Mean (±s.d.) rate of change for TE below 2 kHz was −4.51±1.79 mm kHz^{-1}, and mean rate of change for LV above 2 kHz was −3.38±2.59 mm kHz^{-1}.
TE, tongue elevation; LV, larynx–vertebra distance.

for the start of the next syllable. The cost of this can be seen in a comparison of the f_0:$2f_0$ ratio under 2 kHz for each syllable; for 407-4, the ratio is 1.31:1, while for 436-5 and 503-1, the ratios are 1.78:1 and 1.93:1, respectively. This indicates that bird 407 is not filtering out $2f_0$ as effectively at these low frequencies, although there may be a tradeoff such that other advantages are gained by a very fast repetition rate. This example illustrates one possible situation that runs counter to predictions due to other constraining factors.

Syllable 407-4 (Fig. 5A) also provides evidence that tongue elevation seems to be operating independently of OEC expansion. At the start of the syllable with the OEC mostly contracted and just beginning to expand, the tongue elevates to a peak height and then lowers. As the tongue lowers, the OEC continues to expand. This pattern is in contrast to that seen for other syllables reported here in which the tongue elevates when the OEC is fully or almost fully expanded. Syllables 512-1 and 512-2 (Fig. 5B) offer additional support for the independence of these two mechanisms. The downward-sweeping syllable 512-1 ends at a relatively high f_0 of 2.2 kHz and shows no tongue elevation at any point. However, the second downward-sweeping syllable, 512-2, terminates at 1.4 kHz and shows tongue elevation beginning from 2.0 kHz. Importantly, the OEC expands to almost the same volume for each (LV=~19.0 mm for 512-1 and ~19.5 mm for 512-2 at peak). If tongue elevation is a by-product of OEC expansion, it is not clear why 512-1 does not show any elevation. Of the 12 syllables analyzed for the present study, 512-1 is the only one that does not drop below 2 kHz and is also the only one that does not show any tongue elevation at all.

The complex syllable in Fig. 5C consists of two notes, and tongue elevation is observed between each note rather than at the start or end of the syllable. Note a occurs first at a high f_0, but then there is a pause in vocalization, the tongue elevates and the syllable resumes with note b rising up from a low f_0 of 1.9 kHz. Although tongue elevation begins during the pause between notes and thus could potentially be associated with either note, it is at a higher elevation for the start of note b than for the end of note a.

Tongue elevation is plotted against f_0 for six syllables from four birds in Fig. 6 to further illustrate the relationship. Syllable 503-3 (Fig. 6C) is unique among the syllables presented here in that, despite a brief initial rising FM, the f_0 of this otherwise downward-sweeping syllable is almost entirely below 2.0 kHz and shows
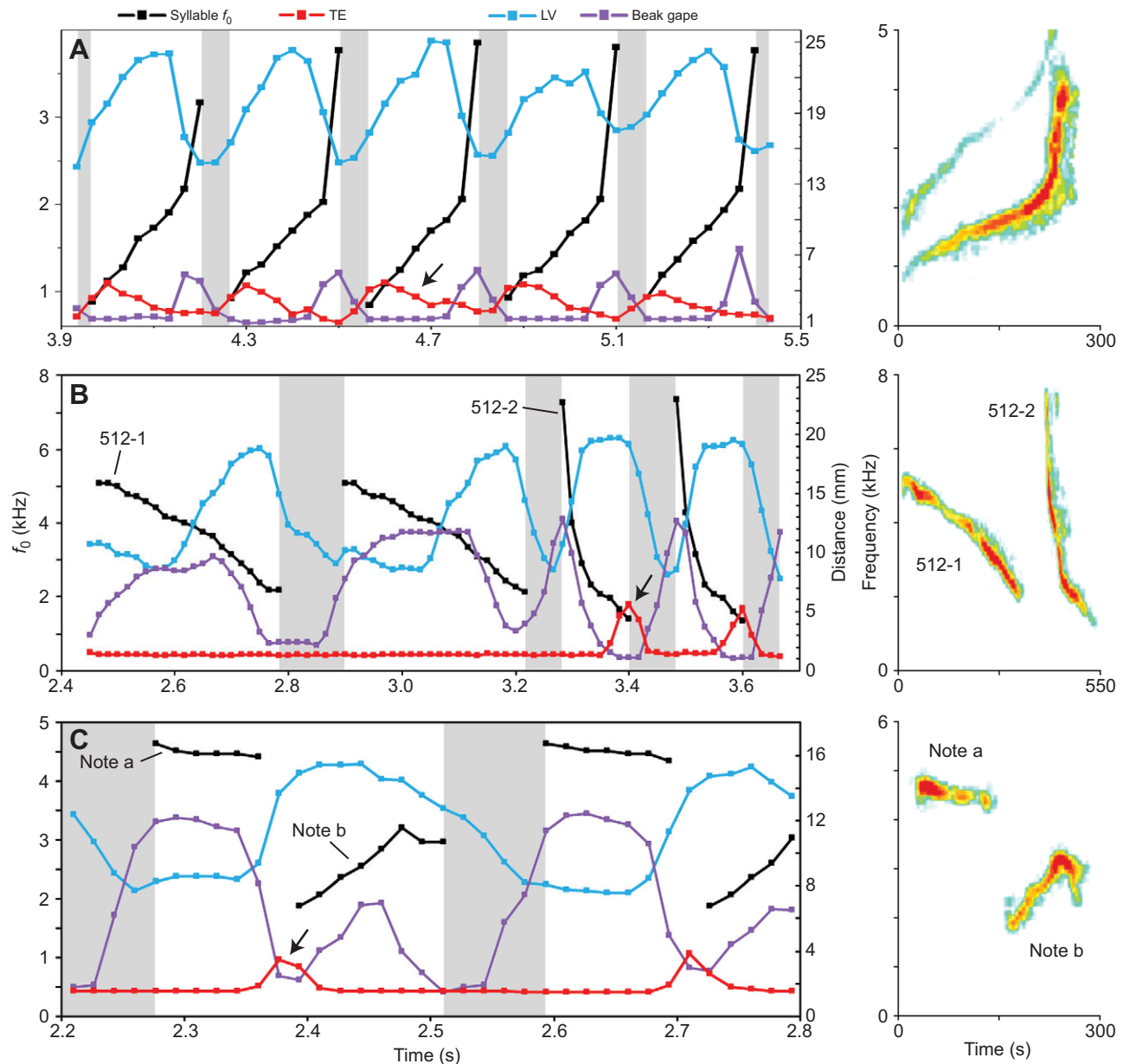
Fig. 5. Supra-syringeal motor patterns showing TE in syllables of three additional birds. A, syllable 407-4; B, syllables 512-1 and 512-2; C, syllable 483-1. Downward-sweeping syllables are shown occurring together in a series. Syllable 512-1 terminates at 2.2 kHz and does not elicit TE. Syllable 483-1 (C) is a complex syllable composed of notes a and b. The gray shading indicates the recovery pause.

tongue elevation for its entire duration. We hypothesize that the reduced slope of tongue elevation between about 0.8 and 1.0 kHz may be due to the tongue reaching the roof of the mouth when its elevation is about 7 mm.

Syllable 522-2 (Fig. 6F) is the only syllable of the 12 analyzed for this study that does not show a relationship between tongue elevation and f_0. Tongue elevation is at its maximum value when f_0 is below 2 kHz, but the maximum elevation is only about 3 mm and tongue position is quite variable, indicating a lack of stereotypy for the lingual motor pattern of this syllable. In addition, there is very little change in LV during this syllable; LV unusually remains large at ∼16 mm throughout the duration. Beak gape above 2 kHz is the only mechanism investigated here that shows a relationship with f_0 (Table 1). This syllable is unusual in that it has a long duration (∼500 ms) yet only has a small change in frequency from start to

end (from 1.5 kHz to 3.2 kHz). As there is variable tongue movement, the f_0:$2f_0$ ratio under 2 kHz was calculated for each sample to determine whether those that had less tongue elevation showed different resonance properties (i.e. more energy in $2f_0$). However, this ratio was consistently ∼1.7:1 regardless of the tongue position. A closer analysis of this syllable indicated that the bird does not seem to be using the tongue in the same way as observed in all other syllables. Instead, the tongue seems to float in a semi-resting position and even sometimes increases slightly in elevation as the syllable progresses (against prediction). This might be due to the unusual nature of the syllable as described above, or might be an individual characteristic of this bird (no other syllables were recorded for it). As to why there were no differences in amplitude ratios for the variable tongue elevation, it may be due to the tongue almost never elevating beyond ∼2 or 3 mm. Such a small elevation
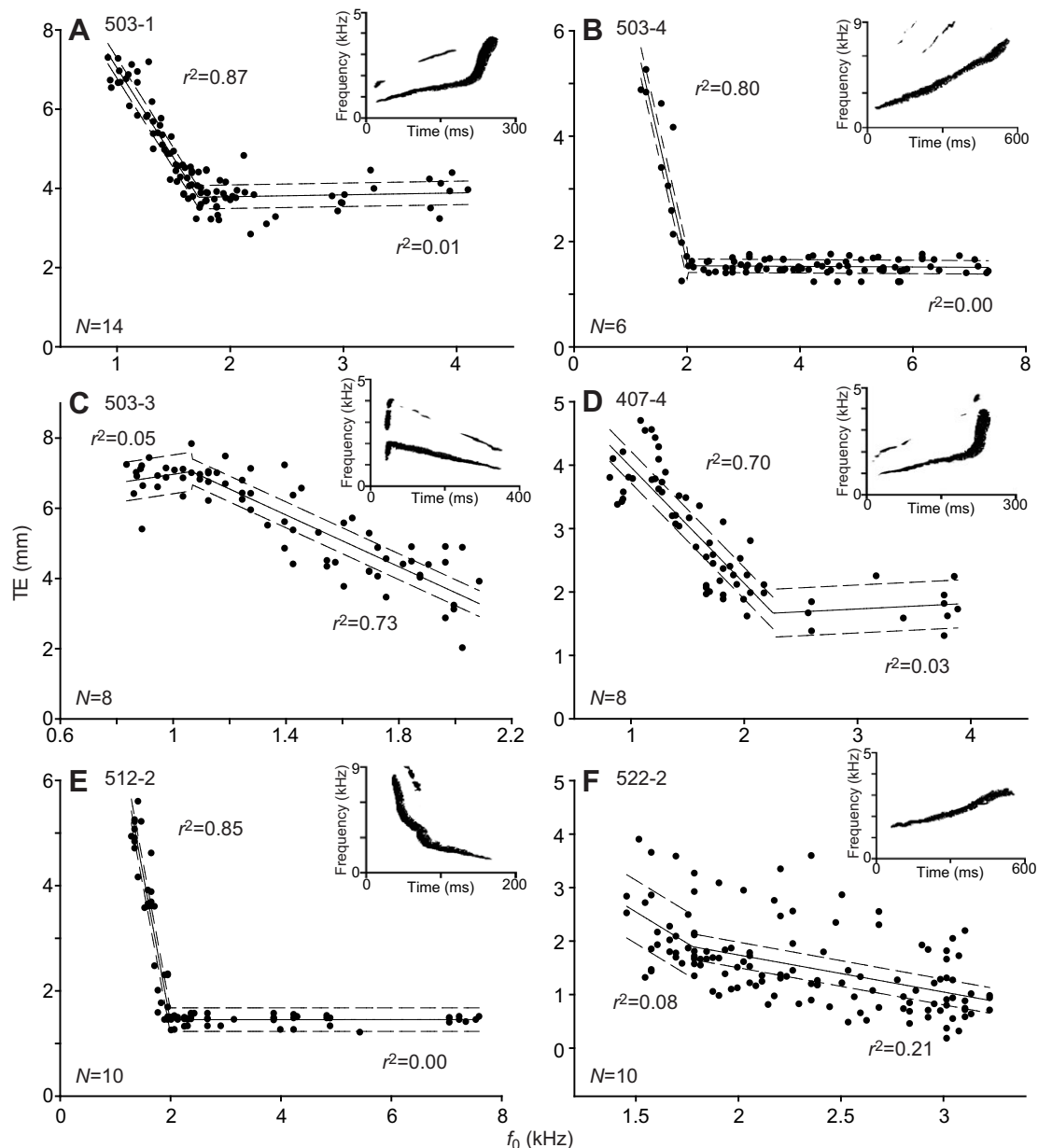
Fig. 6 . Timing and amplitude of TE as a function of f_0. (A–F) Six syllables from four different birds. D and E show syllables 407-4 and 512-2, which were also reported on in Fig. 5A and B, respectively. r^2 values for each segment are reported within the figure, and breakpoints between segments are reported here: A, 1719 Hz; B, 2010 Hz; C, 1065 Hz; D, 2257 Hz; E, 1993 Hz; F, 1778 Hz. N=number of syllables; dashed regression lines show the 95% confidence level.

might only cause an insignificant change in amplitude ratios, which are difficult to detect due to the constant, undirected tongue movement.

Breakpoints for tongue elevation were calculated for all syllables (excluding syllable 503-3, which occurs almost entirely under 2 kHz and has constant tongue elevation, and syllable 512-1, which occurs entirely above 2 kHz and has no tongue elevation). The mean (±s.d.) frequency below which tongue elevation occurred was 1908 Hz (±148 Hz; N=10).

Fletcher et al. (2006) developed a computational model in which OEC volume and a combination of beak gape and tongue position act as the controlling factors influencing the resonance properties of the songbird vocal tract. In order to understand the acoustics of this model, the vocal tract can be considered an analog of a Helmholtz resonator, which consists of a cavity (analogous to the OEC) and a

tube-like neck open to the air (analogous to the opening from the OEC to the beak). The resonance frequency of the cavity and its connected tube is inversely proportional to the square root of the cavity volume and proportional to the square root of the cross-sectional area of the open tube. By decreasing this opening, which in birds would be analogous to decreasing the cross-sectional area of the passage into the beak, the resonance frequency will decrease. One means of accomplishing this would be to elevate the tongue, which agrees with observations reported here.

The cross-sectional area of the passage between the OEC and the beak was estimated to be about 20 mm^2 in anesthetized and euthanized cardinals. Based on observations that the tongue only elevates when the beak is closed and the OEC is generally at its maximum volume, resonance frequency predictions were calculated with these values held constant and the cross-sectional area varied
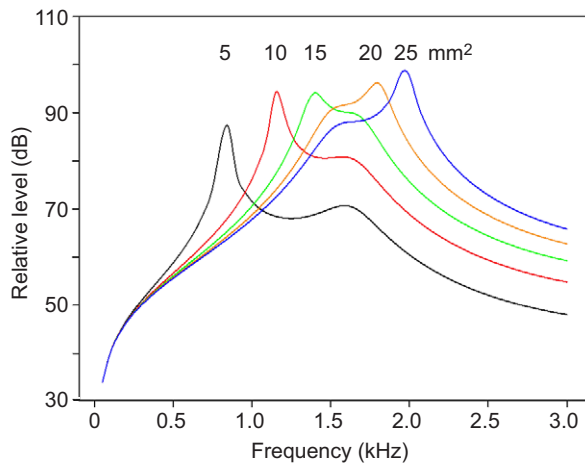
Fig. 7. Reducing the cross-sectional open area of passage from the OEC to the beak by raising the tongue shifts resonance peak to a lower frequency. The numbers above the peaks represent cross-sectional area in 5 mm^2 intervals from 5 to 25 mm^2. The cross-sectional area when fully open was estimated to be about 20 mm^2 in anesthetized and euthanized cardinals.

from 5 to 25 mm^2 in 5 mm^2 intervals using the model by Fletcher et al. (2006; Fig. 7). The resulting resonance peaks show that reducing the area of the passage from the OEC causes resonance frequencies to drop from about 2 kHz to as low as 0.8 kHz, which is very close to the minimum frequency range of cardinal vocalizations. These data support the hypothesis that the elevation of the tongue enables the bird to alter its vocal tract resonance to frequencies extending below 1 kHz.

DISCUSSION

The results reported here show that northern cardinal song is frequently accompanied by elevation of the tongue between the back of the beak and the OEC. Such tongue elevation occurs when the dominant frequency of the song is below ~2 kHz and has an inverse relationship with frequency. This is in agreement with a computational model of the OEC, which shows that vocal tract resonance can be tracked below 2 kHz by varying the cross-sectional area of the passage from the OEC into the beak.

The passerine tongue is supported anteriorly by the paraglossale and posteriorly by the basihyale (Bock and Morony, 1978). The basihyale in turn is connected to two hyoid horns, or cornuae, which together with the urohyale make up the hyoid apparatus (Homberger, 1986). The hyoid apparatus, to which the larynx is attached, moves freely from the skull; extrinsic muscles move the basihyale dorsoventrally and craniocaudally while the cornuae simultaneously move laterally, resulting in the expansion and contraction of the OEC (Homberger, 1986; Riede et al., 2006). Because the tongue is structurally linked to the hyoid apparatus, changes in OEC volume will therefore result in movement of the tongue, at least posteriorly where the larynx is located. However, as shown in Fig. 5B, nearly identical degrees of laryngeal displacement differ significantly in the movement of the anterior tip of the tongue, indicating that the bird may have control over this mechanism independent of the expansion of the OEC. This is also supported by the poor relationship between tongue elevation and LV below 2 kHz as shown in Fig. 4 and Table 1.

Many avian species, including cardinals and parrots, use the tip of the tongue to manipulate seeds and position them between the jaws for husking (Bock and Morony, 1978; Homberger, 1986). In parrots, the anterior tip of the paraglossale is rotated dorsally to push the seed

against the mandible. The paraglossale is moved by intrinsic lingual muscles which, although not disconnected from the extrinsic muscles that move the basihyale, do allow the tip of the tongue to move with a certain degree of independence. X-ray recordings of a cardinal husking a seed reveal rapid dorsoventral movements of the tip of the tongue but only slight movements of the larynx and cornuae (J.R.R., personal observation). It is likely that the tip of the tongue can be elevated to block the OEC opening during song by the same mechanism, which would account for the lack of a relationship between tongue elevation and LV reported here.

Although tongue elevation is typically observed at the beginning or end of syllables, it does not seem to be the case that it is simply connected to the onset or termination of vocalization. Instead, the data reported here indicate that tongue elevation is closely linked to the frequency of the song. Of the 12 syllables analyzed, one occurred with f_0 almost entirely under 2 kHz and showed tongue elevation for the duration of the syllable (syllable 503-3, Fig. 6C). Additionally, one syllable occurred entirely above 2 kHz and showed no tongue elevation at all (syllable 512-1, Fig. 5B). All FM sweep syllables that passed across 2 kHz showed tongue elevation occurring below about 2 kHz, regardless of the duration or f_0 range of the syllable. These observations strongly support the link between frequency and tongue elevation. However, the rate-of-change data reported in Table 2 do not provide sufficient evidence to say whether the bird is finely tracking f_0 with tongue elevation or whether the tongue is simply elevating to close the OEC opening regardless of the exact f_0. There may be complex interactions between OEC expansion and tongue elevation unique to certain syllable types (e.g. upward or downward sweeping) that make cross-syllable comparisons difficult.

In light of the observations reported here, a more complete picture of the songbird vocal tract filter can be constructed by building on the model of Fletcher et al. (2006). Sound generated at the syrinx radiates from the beak after passing through the filter components of the vocal tract, which include the trachea, glottis, OEC and beak. Lengthening of the trachea and constriction of the glottis may cause changes to the resonance frequency, but it is unknown whether songbirds employ either of these mechanisms. Daley and Goller (2004) measured tracheal length changes in vocalizing zebra finches but found that changes were of small magnitude and apparently not related to song frequency. Tracheal length changes were also measured in a singing cardinal by embedding segments of radio-opaque silver wire in the wall of the trachea and using X-ray cineradiography to analyze the changes in distance between the segments (R.A.S., unpublished data). The results showed an increase in tracheal length of ~15% at low frequencies, but this lengthening correlated strongly with OEC volume and may be passively driven by the expansion of the OEC. Riede et al. (2006) also reported that tracheal resonances, observed in the harmonics of FM sweeps, change very little during song in cardinals.

After passing through the glottis, sound enters the OEC, which acts as a major resonance filter by expanding to a large volume at low frequencies and contracting at high frequencies. Although observations of cardinals indicate that OEC volume is inversely proportional to f_0 above ~2 kHz, below 2 kHz it appears to be held constant at a large volume. In agreement with the data reported here and as originally suggested by Fletcher et al. (2006), the filtering properties of the OEC are further adjusted in cardinals at resonance frequencies below ~2 kHz by elevation of the tongue, which constricts the opening into the beak.

Fletcher's model predicts that the effectiveness of the beak as a filter is greatly reduced unless it is almost closed (Fletcher et al.,

2006). However, beak gape in cardinals seems to partially track f_0 even when the beak is widely open (e.g. Fig. 5). Opening of the beak is almost only observed above 2 kHz, such that the tongue elevates when the beak is closed. The acoustic filtering properties of beak gape and how they interact with other mechanisms of the vocal filter are still not well understood, and more research is required to explore this further.

We have shown that tongue elevation in cardinals is associated with resonance frequencies of the vocal tract that extend below ∼2 kHz. The behavioral advantage of this, if any, is unknown. The ability to produce very low frequencies at a relatively high amplitude may, for example, serve as a performance limit in vocal production that tends to exaggerate the perception of a singing male's size that females prefer and/or competing males tend to avoid. Although female cardinals sing, there are important differences between the sexes and we have no data on tongue elevation in female cardinals.

Supplementary information
Supplementary information available online at
http://jeb.biologists.org/lookup/suppl/doi:10.1242/jeb.126532/-/DC1

References
Beckers, G. J. L., Nelson, B. S. and Suthers, R. A. (2004). Vocal-tract filtering by lingual articulation in a parrot. *Curr. Biol.* **14**, 1592-1597.
Bock, W. J. and Morony, J. (1978). The preglossale of passer (Aves: Passeriformes) - a skeletal neomorph. *J. Morphol.* **155**, 99-109.
Daley, M. and Goller, F. (2004). Tracheal length changes during zebra finch song and their possible role in upper vocal tract filtering. *J. Neurobiol.* **59**, 319-330.
Doupe, A. J. and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567-631.
Elemans, C. P. H. (2014). The singer and the song: the neuromechanics of avian sound production. *Curr. Opin. Neurobiol.* **28**, 172-178.
Fletcher, N. H., Riede, T. and Suthers, R. A. (2006). Model for vocalization by a bird with distensible vocal cavity and open beak. *J. Acoust. Soc. Am.* **119**, 1005.
Goller, F. and Riede, T. (2013). Integrative physiology of fundamental frequency control in birds. *J. Physiol. Paris* **107**, 230-242.
Homberger, D. G. (1986). The lingual apparatus of the African grey parrot, Psittacus erithacus Linné (Aves: Psittacidae): description and theoretical mechanical analysis. In *Ornithological Monographs no. 39*, iii-xi, pp.1-233. Washington, DC: American Ornithologists' Union.
Hultsch, H. and Todt, D. (2004). Learning to sing. In *Nature's Music. The Science of Birdsong* (ed. P. Marler and H. Slabbekoorn), pp. 80-107. San Diego: Elsevier.
King, A. S. (1989). Functional anatomy of the syrinx. In *Form and Function in Birds*, Vol. 4 (ed. A. S. King and J. McLelland), pp. 105-192. London: Academic Press.
Klatt, D. H. and Stefanski, R. A. (1974). How does a mynah bird imitate human speech? *J. Acoust. Soc. Amer.* **55**, 822.
Larsen, O. N. and Goller, F. (2002). Direct observation of syringeal muscle function in songbirds and a parrot. *J. Exp. Biol.* **205**, 25-35.
Nottebohm, F. (1976). Phonation in the orange-winged Amazon Parrot, *Amazona amazonica*. *J. Comp. Physiol. A* **108**, 157-170.
Ohms, V. R., Snelderwaard, P. C., ten Cate, C. and Beckers, G. J. L. (2010). Vocal tract articulation in zebra finches. *PLoS ONE* **5**, e11923.
Ohms, V. R., Beckers, G. J. L., ten Cate, C. and Suthers, R. A. (2012). Vocal tract articulation revisited: the case of the monk parakeet. *J. Exp. Biol.* **215**, 85-92.
Patterson, D. K. and Pepperberg, I. M. (1994). A comparative study of human and parrot phonation: acoustic and articulatory correlates of vowels. *J. Acoust. Soc. Am.* **96**, 634.
Pepperberg, I. M. (2010). Vocal learning in Grey parrots: a brief review of perception, production, and cross-species comparisons. *Brain Lang.* **115**, 81-91.
Riede, T. and Goller, F. (2010). Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. *Brain Lang.* **115**, 69-80.
Riede, T. and Suthers, R. A. (2009). Vocal tract motor patterns and resonance during constant frequency song: the white-throated sparrow. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* **195**, 183-192.
Riede, T., Suthers, R. A., Fletcher, N. H. and Blevins, W. (2006). Songbirds tune their vocal tract to the fundamental frequency of their song. *Proc. Natl. Acad. Sci. USA* **103**, 5543-5548.
Riede, T., Schilling, N. and Goller, F. (2013). The acoustic effect of vocal tract adjustments in zebra finches. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* **199**, 57-69.
Suthers, R. A. (1990). Contributions to birdsong from the left and right sides of the intact syrinx. *Nature* **347**, 473-477.
Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. and Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167-1176.
Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall.