Inflation of the esophagus and vocal tract filtering in ring doves

Tobias Riede[1],*, Gabriël J. L. Beckers[1],†, William Blevins[2] and Roderick A. Suthers[1]

[1]Medical Sciences, Indiana University, Bloomington, IN 47405, USA and [2]Department of Veterinary Clinical Sciences, Diagnostic Imaging Section, Purdue University, West Lafayette, IN 47907, USA

*Author for correspondence at present address: Institut für Theoretische Biologie, Humboldt-Universität zu Berlin, Invalidenstrasse 43, 10115 Berlin, Germany (e-mail: tobiasriede@web.de)

†Present address: Behavioural Biology, Institute of Biology, Leiden University, PO Box 9516, 2300 RA Leiden, The Netherlands

Summary

Ring doves vocalize with their beaks and nostrils closed, exhaling into inflatable chambers in the head and neck region. The source sound produced at the syrinx contains a fundamental frequency with prominent second and third harmonic overtones, but these harmonics are filtered out of the emitted signal. We show by cineradiography that the upper esophagus, oral and nasal cavities collect the expired air during vocalization and that the inflated esophagus becomes part of the suprasyringeal vocal tract. The level of the second and third harmonics, relative to the fundamental frequency (f_0), is reduced in the esophagus and emitted vocalization compared with in the trachea, although these harmonics are still considerably higher in the esophagus than in the emitted signal. When the esophagus is prevented from fully inflating, there is a pronounced increase in the level of higher harmonics in the emitted vocalization. Our data suggest that the trachea and esophagus act in series as acoustically separate compartments attenuating harmonics by different mechanisms. We hypothesize that the trachea behaves as a tube closed at the syringeal end and with a variable, restricted opening at the glottal end that lowers the tracheal first resonance to match the f_0 of the coo. The inflated esophagus may function as a Helmholtz resonator in which the elastic walls form the vibrating mass. Such a resonator could support the f_0 over a range of inflated volumes.

Key words: tracheal resonance, Helmholtz resonator, birdsong, harmonic suppression, source-filter theory, *Streptopelia risoria*.

Introduction

The spectral content of vocalizations used in acoustic communication is often an important information-bearing parameter of the signal that is controlled by the sender. A classic example from human speech is the source-filter theory for perceiving and classifying vowel sounds according to the frequency of their formants (Peterson and Barney, 1952), which are determined by resonance characteristics ('filtering') of the vocal tract. There is evidence that vocal tract filtering also plays an important role in vocal communication of other animals, although little is known about the mechanisms by which they are controlled (but see Hoese et al., 2000; Beckers et al., 2004).

Many birds rely heavily on vocal communication, making them an excellent group in which to examine the mechanisms of vocal tract filtering. Young zebra finches (*Taeniopygia guttata*) learn certain aspects of their song's harmonic pattern from their tutor (Cynx and Shapiro, 1986; Cynx et al., 1990; Williams et al., 1989). Both zebra finches and budgerigars (*Melopsittacus undulatus*) outperform humans in their ability to detect decrements in the amplitude of a single component in a harmonic complex (Lohr and Dooling, 1998). Together with canaries (*Serinus canaria*), these species are able to discriminate between synthetic harmonic complexes based on phase differences in harmonic components that require a temporal resolution two or three times better than that of humans (Dooling et al., 2002). Red-winged blackbirds (*Agelaius phoeniceus*) and pigeons (*Columba* sp.) are able to discriminate human vowel sounds (Hienz et al., 1981). Some songbirds suppress the harmonic components in their songs with the aid of vocal tract resonances (Nowicki, 1987). The nature of this vocal tract filter is uncertain and may include changes in beak gape that are correlated with the fundamental frequency (f_0; Goller et al., 2004; Hoese et al., 2000; Suthers and Goller, 1997; Westneat et al., 1993). Young song sparrows tutored with song that is abnormally rich in harmonics usually suppress the harmonics in their copy of these songs (Nowicki et al., 1992).

The songs of ring doves (*Streptopelia risoria*) consist of stereotypic coos (Fig. 1) uttered in bouts of variable duration (Miller and Miller, 1958; Nottebohm and Nottebohm, 1971). A striking feature of the adult male coo is the concentration of its acoustic energy into a low f_0 (Ballintijn and ten Cate, 1997a,b; Gaunt et al., 1982). Harmonics are more prominent in the songs of juvenile males (Ballintijn and ten Cate, 1997a)

Fig. 1. Changes in the amount of esophageal inflation during a sequence of two coos. The enlargement of the esophagus (top graph) is produced by airflow through the syrinx in the expiratory direction while the beak and nares are closed. Complete beak closure, as seen in the x-ray video, is indicated by arrow 1. The amount of inflation was estimated by measuring the area of the esophagus (shaded area on the drawing of the dove) in the two-dimensional x-ray image. Maximum inflation occurs at the end of the coo (arrows 2 and 3). The esophagus does not completely collapse between bouts of cooing. Coos are shown spectrographically (middle panel) and as an oscillogram (bottom panel). i, inspiratory note; $e1$, first note of the coo; p, pause between first and second note; $e2$, second note of the coo.
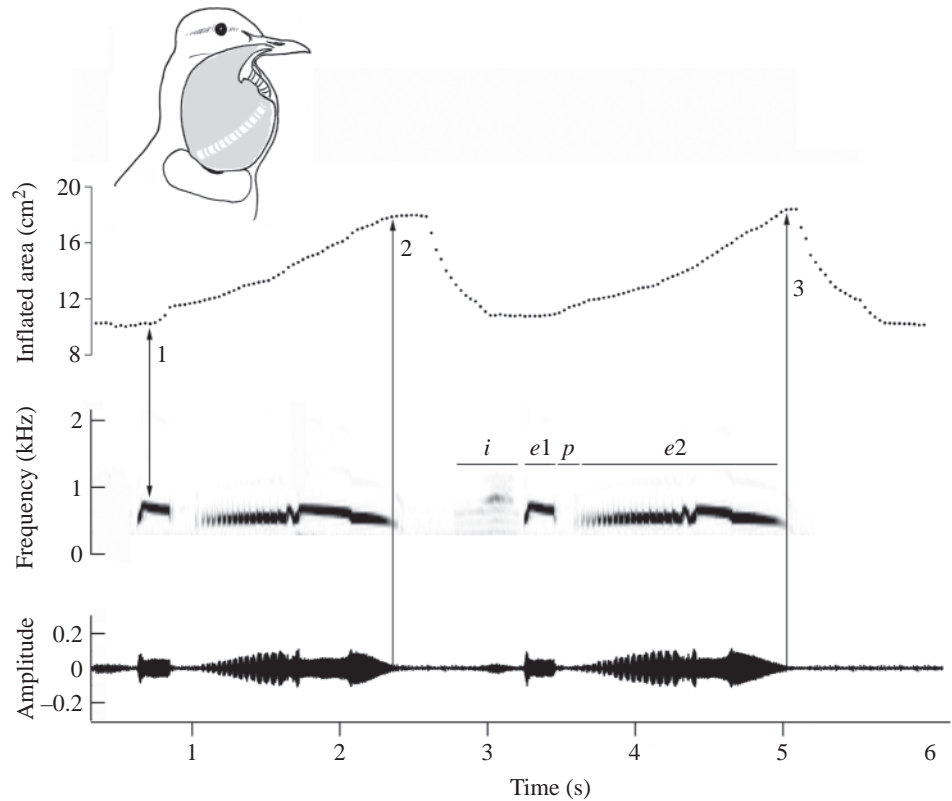
and females (Ballintijn and ten Cate, 1997b), which may use the resulting spectral cues to identify and respond differently to male *vs* female calls (Cheng, 1992; Cheng et al., 1998). Coos are presumably generated, as they are in pigeons (*Columba livia*), by vibration of a pair of lateral tympaniform membranes in the syrinx at the base of the trachea (Goller and Larsen, 1997; Larsen and Goller, 1999).

Recordings of sound close to the ring dove's syrinx indicate that a harmonic source signal is severely filtered by the vocal tract (Beckers et al., 2003), resulting in a vocalization with most of its energy at the fundamental frequency. The nature of this filter is unknown. Beak movements are not involved since ring doves, like most Streptopelidae, coo with their beaks and nares closed (Gaunt et al., 1982). Neither is it clear how tracheal resonance could contribute to the filter mechanism since the predicted quarter wavelength first resonance of the trachea, modeled as a stopped tube, is at a frequency nearly twice that of the coo's f_0 (Beckers et al., 2003). During a coo, subsyringeal respiratory pressure is maintained by the expiratory muscles (Gaunt et al., 1982), and the air stream through the syrinx is guided into a closed suprasyringeal cavity from which sound must radiate through body tissue into the environment. Cooing is accompanied by a lowering of the head and a prominent expansion of the neck, which has been assumed to involve inflation of the esophagus and the crop (Gaunt et al., 1982). It is not known if this inflation is simply a byproduct of air flowing into a sealed suprasyringeal space or if it is part of a vocal tract filter responsible for suppressing source-generated harmonics.

Inflatable structures in the cervical region, the esophagus or esophageal sacs have been described in many birds (for review, see McLelland, 1989). Inflation of the esophagus is facilitated by numerous longitudinal folds, the plica esophageales, on its luminal surface that increase the distensibility of the wall (McLelland, 1989). Male sage grouse (*Centrocercus urophasianus*) court females at the lek with displays that include inflation of a thin-walled, balloon-like pouch that forms an enlargement of the normal esophageal tube (Clarke et al., 1942; Honess and Allred, 1942) and may produce the directional patterns of the acoustic display (Dantzker et al., 1999). Inflatable structures associated with vocalization have also evolved in several other vertebrate taxa. In primates, various guenons (Cercopithecinae) develop paired or singular structures, the so-called air sacs, which develop directly from the laryngeal or pharyngeal cavity (Gautier, 1971). Perforation of the air sacs reduced the amplitude of the vocalization and enriched the spectral pattern so that more harmonics were visible in the spectrogram (Gautier, 1971).

The anuran vocal sac derives from the oral cavity and forms a large part of the supralaryngeal space. A heliox experiment in four anuran species revealed no consistent or predictable change in the frequency distribution of sound energy (Rand and Dudley, 1993), leading to the conclusion that the air sac is not a cavity resonator (Rand and Dudley, 1993). Instead of functioning as an acoustic filter, the highly elastic anuran vocal sac may increase the speed and decrease the energetic cost of reinflating the lungs following vocalization (Jaramillo et al., 1997).

It appears that inflatable structures have evolved various acoustic, energetic and communicative roles in different taxa. Doves provide a particularly interesting and accessible species in which to investigate these structures.

In the present study, we use imaging techniques to show that vocal tract inflation involves the esophagus but not the crop or cervical air sacs. Through measurement and experimental manipulation of intraesophageal and intratracheal pressures, we experimentally investigate the possible role of these structures as part of the vocal tract filter in doves.

Materials and methods

Subjects

We used 15 adult male ring doves (*Streptopelia risoria* L. 1758) that were obtained from a local breeder.

Cineradiography

X-ray imaging was performed with a Cardiac Digital Mobile Imaging System (Series 9800) at the Department of Veterinary Clinical Sciences, Diagnostic Imaging Section, Purdue University, West Lafayette, IN, USA on three male ring doves (I.D. 13, 14 and 15) as they vocalized spontaneously while sitting in a cage positioned between the x-ray source (a battery-buffered high-frequency generator; 15 kW; 60 kHz; pulses up to 150 mA at 30 pulses s^{-1} with 10 ms pulse width) and the digitally recorded phosphorscreen (playback speed up to 30 frames s^{-1} during real-time fluoroscopy). The digital video output signal from the fluoroscope was recorded on a S-VHS tape recorder together with the sound signal (Sennheiser ME 88 microphone on a K6 power module and a MP13, Rolls, Salt Lake City, UT, USA preamplifier), in order to synchronize vocal tract movement during sound production with the vocalization. The motion of the vocal tract, esophagus and crop was reconstructed from successive video images (Vegas Video, Sonic Foundry, Madison, WI, USA), allowing for a 33 ms interval between successive frames, and correlated with vocalizations on the video sound track.

Surgical procedures and data recording

All experiments, except cineradiography, were carried out at Indiana University, Bloomington, IN, USA. Prior to surgery, the bird was anesthetized with isoflurane (Abbott Laboratories, Chicago, IL, USA). Two kinds of surgery were performed. In some birds, a short tube was inserted through the wall of the esophagus and the overlying skin (esophagostomy). In other birds, a tube was attached to an opening in the wall of the trachea (tracheal tube).

Esophagostomy

In six doves, the right lateral cervical region was prepared by shaving the feathers. A stainless steel tube was introduced into the esophagus *via* the beak and pressed laterally so it could be palpated through the skin. An incision down into the esophagus was performed, and the end of the tube was pulled out through the wall of the esophagus and skin. The tube was fixed to the skin by circular sutures and capped with a removable plastic cap that allowed the inside of the tube to be cleaned. Two different tube systems were used.

In doves 1, 2 and 11, the esophageal tube (6 mm in length, 4 mm in diameter) was equipped with a cap so it could be closed to allow normal inflation of the esophagus during song. In some recordings, the cap was removed, opening the esophagus to the outside and preventing full inflation.

In doves 3, 7 and 12, the esophageal tube (11 mm in length, 4 mm in diameter) was equipped with a cap connected to a piezoresistive pressure transducer (FPM-02PG; Fujikura, Lexington, MA, USA) *via* a flexible silastic tube (Dow Corning, Midland, MI, USA; internal diameter = 1.02 mm, wall thickness = 0.57 mm). The bird wore an elastic belt around its thorax with a Velcro™ tab on the back to which the pressure transducer was attached. The silastic tube from the throat to the backpack was routed over the feathers, and care was taken to avoid pressure artifacts due to body movements. The dynamic range of the pressure transducer is ±140 cmH$_2$0 (±13.79 kPa) re ambient pressure. Its mechanical response is 2 ms. The frequency response of the pressure transducer and the tube is within 3 dB from DC to the dove's second harmonic. Most of the variation in response is due to resonances in tubing.

Tracheal tubes

In three doves (doves 4, 5 and 8), we measured intratracheal pressure. The left lateral cervical region was prepared by shaving the feathers. After skin incision, the trachea was exposed. A metal tube (8 mm in length, 1 mm diameter) was introduced into the middle third of the trachea by removing a piece of one tracheal ring to produce an opening equal to the diameter of the tube. The metal tube was held in position by tissue adhesive, which fixed it to the trachea and prevented air leaks, and by circular sutures where it exited the overlying skin. A flexible silastic tube connected the metal tube to the pressure transducer on the bird's backpack.

In dove 4, we simultaneously measured intraesophageal and intratracheal pressure by a combined implantation of an esophageal and a tracheal tube, each connected to its own pressure transducer in the backpack.

Spontaneous singing was recorded before and after surgery using a condensor microphone (Sennheiser MKH 40) placed 1 m in front of the cage. Vocalizations before surgery were automatically recorded and saved as uncompressed files on a computer (Avisoft-Recorder; www.avisoft.de). After surgery, the emitted vocalizations, esophageal pressures and tracheal pressures were recorded digitally on separate channels of a rotary storage recorder (Sypris Data Systems, Huntsville, AL, USA; Metrum model RSR 512).

Data analysis

Coos of doves 1, 2 and 11 were analyzed for total duration, duration of *e*1 and *e*2 notes, pause duration between *e*1 and *e*2 notes (*p*), mean and maximum fundamental frequency (*f*$_0$) in *e*2 note (Fig. 1) and mean and maximum sound level. All

measurements were performed using sound analysis software (PRAAT, version 4.1 for Windows; www.praat.org). Temporal parameters (total duration, duration of $e1$ and $e2$ notes, pause duration between $e1$ and $e2$ notes) were measured by hand in the time domain. Linear predictive coding (LPC; Markel and Gray, 1976) and associated peak-extraction algorithms were used to track f_0 values. Fundamental frequency values were computed at 0.01 s intervals, with a 0.049 s cosine window and six coefficients producing approximately 2 Hz resolution. Before computing final f_0 outcomes, each track was visually inspected by overlaying it on a corresponding narrowband spectrogram (0.040 s Hanning window). Fundamental frequency measurements were analyzed as maximum or mean values of the $e2$ note. The reported mean and maximum sound level values in dB are not relative to a common standard and can only be compared between treatments within individuals.

The microphone signal (in doves 1, 2 and 11) and the microphone, tracheal and esophageal signals (in doves 3, 4, 5, 7, 8 and 12) were analyzed. We selected a 120 ms portion within the $e2$ segment (Fig. 1) with little f_0 variability using the spectrogram (30 ms Hanning window, time step 15 ms). An average power spectrum was calculated for that portion. The amplitude values for f_0, the second harmonic ($2f_0$) and third harmonic ($3f_0$) were derived from the power spectrum matrix. Fundamental frequency was automatically detected by choosing the maximum sound level in the expected frequency range (450–750 Hz); correctness was visually confirmed. Second and third harmonics were either derived by choosing the maximum amplitude in the expected frequency range (depending on actual f_0 value) or, if no higher harmonics were obvious in the spectrogram, by averaging over a 50 Hz range within the expected frequency range. In the latter case, the level of the overtone is below the noise floor. Since this phenomenon occurred only in the emitted calls and not systematically within individuals, we decided to include the measurements of such calls. The spectra of esophageal and tracheal sounds were corrected for the resonance characteristics of the tube system (see 'transducer calibration'). We calculated the level of $2f_0$ and $3f_0$ relative to that of the f_0.

The first 20 coos (up to five from a single bout) were analyzed, except for dove 5 (five coos) and dove 7 (11 coos) with less than 20 vocalizations. Statistical analysis was carried out in SPSS for Windows, version 10.1.

Transducer calibration

The harmonic distortion due to a nonlinear response of the piezoresistive transducers is small and can be ignored for our purposes (Beckers et al., 2003). Resonance characteristics of the silastic tubing, however, change the emphasis of certain frequency ranges through linear filtering, the effect of which is not negligible. We therefore tested the frequency response of the system (pressure transducer plus silastic tube plus metal tube implant) by recording speaker-generated pink noise. The generated sound was recorded 50 cm in front of the speaker with both the transducer system and a reference microphone (Sennheiser ME 80). All amplitude measurements were corrected by subtracting the transfer function of the transducer system from the respective esophageal or tracheal signal. Variance in the frequency response of the pressure transducer system introduced deviations of -0.3 ± 3.7 dB
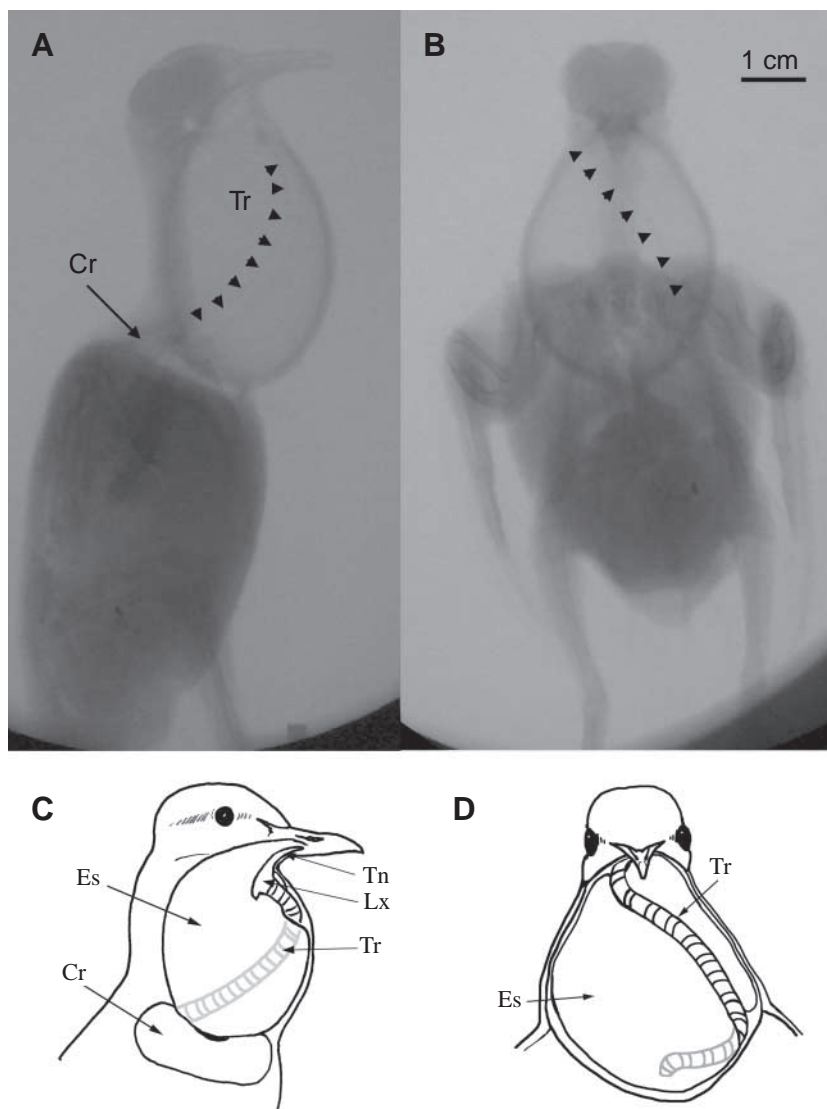

Fig. 2. (A,B) Cineradiograph of a cooing ring dove during maximum inflation of the esophagus; (A) lateral view; (B) frontal view. The course of the trachea around the esophagus is indicated by arrowheads. The crop represents a separate cavity from the inflated upper esophagus. (C,D) Schematic drawing of the (C) lateral and (D) frontal view of a dove during full inflation of the esophagus. Cr, crop; Es, esophagus; Lx, larynx; Tn, tongue; Tr, trachea. Drawings by Sue Anne Zollinger.

(mean ± 1 s.d.) and –1.7±3.1 dB in the levels of $2f_0$ and $3f_0$, respectively, for the trachea signal and +1.8±6.8 dB and +1.5±4.8 dB in the levels of $2f_0$ and $3f_0$, respectively, for the esophageal signal.

Results

Cineradiography

We videotaped 15 and 12 perch coos from doves 13 and 14, respectively, and eight nest coos from dove 15. The following description of events associated with vocalization applies to all individuals.

During silent respiration, when doves are not vocalizing, the walls of the esophagus are either collapsed or partly inflated. During periods of vocalization, coos are typically repeated several times to form a bout. Each bout is preceded by a back-and-forth movement of the head while the beak is closed. This is accompanied by an increase of the neck diameter and a partial inflation of the esophagus before the first coo.

Each coo begins with a short note ($e1$ in Fig. 1), having a mean duration of ~200–250 ms and occupying ~7–9 video frames, each of which represents 33 ms. Although Gaunt et al. (1982) report that the beak is closed during the entire coo, the beak of our birds did not appear to be completely closed during approximately the first third to half of $e1$ (3.4±1.6 video frames; N=15, 11 and 6 coos from three doves). The beak closes by the middle of $e1$ and remains closed during the remainder of the coo, including the pause, p, and the $e2$ note (Fig. 1).

Inflation of the esophagus, assessed by its area in the cineradiographs, begins to increase during $e1$ after closure of the beak and continues to gradually increase during the coo, reaching maximum expansion at the end of $e2$ (Fig. 1). Esophageal inflation is accompanied by a lateral displacement of the trachea (Fig. 2). In order to verify the identity of the

esophagus and to rule out other structures such as air sacs, we did positive contrasting in birds that had swallowed barium sulfate. This radiopaque agent is often used to demonstrate the structure of the inner surface of the intestines. Although most of the agent was collected in the crop, cineradiography shortly after oral barium sulfate application demonstrated traces of the agent in the partly inflated esophagus. Furthermore, we x-rayed a feeding bird with a partly inflated esophagus and observed food items inside the inflated cavity.

When maximally inflated, the esophagus has a cranial to caudal dimension of 5–6 cm and a maximum diameter of 3.2–4.1 cm, compared with an uninflated length of 4–4.5 cm and a collapsed lumen of negligible diameter (0 cm^2 area). During inflation, the cranial end of the esophageal lumen becomes a continuous air chamber with the buccopharyngeal cavity. There is no valve-like constriction between these structures, which fuse to form one large cavity. The beak opens at the end of each $e2$ note, which is accompanied by a rapid partial collapse of the esophagus. The esophagus remains partially inflated between the coos within a bout. The beak opens at the end of $e2$ in the last coo of a bout, and the esophagus undergoes an initial rapid partial collapse. The esophagus does not necessarily collapse completely, since we observed in the x-rays that birds can remain with a partially inflated esophagus for at least 5 min. Startling the bird during this period can cause a rapid collapse of the esophagus.

We assume that retraction in the esophageal and skin wall may be reduced after a coo (between coos in a bout and after a bout) so that, for a period of time, the relaxed esophagus has an enlarged lumen even during normal respiration when intraesophageal pressure is presumably around ambient. Skin is a viscoelastic material with nonlinear biomechanical properties. Recovering after stress is time dependent. On release, the skin retracts very quickly at first,

Table 1. *Levels of overtones in different compartments*

Dove I.D.	$2f_0$			$3f_0$		
	Trachea	Emitted	Att.	Trachea	Emitted	Att.
4	–16.7±2.2 (20)	–38.1±4.6 (20)	21.4*	–12.2±2.6 (20)	–38.2±4.6 (20)	26.0*
5	–13.3±3.3 (12)	–44.7±1.2 (5)	31.3*	–29.4±4.1 (12)	–49.7±1.3 (5)	20.3*
8	–29.9±1.9 (20)	–38.2±4.5 (20)	8.3*	–27.2±2.4 (20)	–36.0±7.3 (20)	8.8*
		t=–3.04, P<0.05, N=3			t=–3.65, P<0.05, N=3	
Dove I.D.	Esophagus	Emitted		Esophagus	Emitted	
3	–33.5±6.8 (20)	–43.7±2.4 (20)	10.2*	–35.3±7.3 (10)	–48.2±4.0 (20)	12.9*
4	–36.3±6.9 (20)	–38.1±4.6 (20)	1.8	–43.0±7.2 (20)	–38.2±4.6 (20)	–4.8
7	–30.6±6.5 (11)	–42.0±2.5 (11)	11.4*	–28.0±5.1 (11)	–46.5±2.7 (11)	18.5*
12	–24.7±1.8 (20)	–33.6±2.9 (20)	8.9*	–36.2±2.9 (20)	–36.7±3.4 (20)	0.5
		t=–3.75, P<0.05, N=4			t=–1.25, P=0.14, N=4	

Levels of $2f_0$ and $3f_0$ (mean ± s.d.; in dB re f_0; measured from a 120 ms portion within the $e2$ segment with little f_0 variability) relative to the level of the fundamental frequency (f_0) in trachea, esophagus and emitted vocalization. Number of calls considered is given in parentheses. 'Att.' refers to the amount of attenuation comparing the respective compartment (esophagus or trachea) and the emitted signal. For each individual, attenuation was tested on significance (two-sample t-test); *P<0.01; no asterisk P>0.05. For each compartment, the levels of $2f_0$ and $3f_0$ were tested against the emitted sound, lumping all individuals (paired t-test). Test statistics are given below each compartment.

followed by a slow creeping recovery. The longer the force is applied, the greater the recovery time after release (Barel et al., 1998).

The cranial aspect of the cervical esophagus forms a separate cavity from its distinct caudal expansion, the crop (Fig. 2C,D). During the vocalization, the expanding esophagus pushes the crop caudally. The lumen between the esophagus and crop is constricted, and although the crop has been observed to contain some air it was never inflated during vocalizations.

Properties of vocal signals in the trachea and esophagus

To investigate how source harmonics are filtered out of the radiated vocalization, we measured the amplitude of $2f_0$ and $3f_0$ in the trachea, esophagus and emitted vocalization, relative to that of f_0, during a 120 ms segment of $e2$ (Table 1). Fig. 3 shows examples of one coo recorded in the trachea (Fig. 3A) and one recorded in the esophagus (Fig. 3B). The data for all doves are summarized in Fig. 4 and Table 1. In all three doves tested, the level of $2f_0$ and $3f_0$ in the trachea, relative to the f_0, was significantly higher than in the emitted vocalization (Fig. 4; Table 1). In these birds, the mean relative level of $2f_0$ was 8.3–21.4 dB lower in the emitted signal than in the trachea. The mean relative level of $3f_0$ was 8.8–26.0 dB lower in the emitted signal compared within the trachea.

In three of the four doves for which we have esophageal measurements, the relative mean level of the emitted $2f_0$ was significantly lower (8.9–11.4 dB; Table 1) than in the esophagus, but there was no significant difference between these levels in dove 4. In two birds, doves 3 and 7, $3f_0$ was also significantly lower (12.9 and 18.5 dB, respectively) in the vocalization compared with the esophagus. The relative level of $3f_0$ was not significantly different in the esophagus compared with the emitted vocalization in doves 4 or 12 (Fig. 4; Table 1).

Effect of degree of esophagus inflation on harmonic level

The esophagus inflates during each coo. Its cross-sectional area approximately doubles from the beginning of $e1$ to the end of $e2$ (Fig. 1) as suprasyringeal pressure rises. This is followed by partial deflation at the end of the coo. Harmonics can be prominent during $e1$ and present during most of $e2$. The relative levels of $2f_0$ and $3f_0$ measured over a 120 ms segment in the middle of $e1$ and near the end of $e2$ in five birds differ at most by 3 dB (Fig. 5; Table 2). These data suggest that the substantial changes in the volume of the inflated esophagus, as well as tracheal and esophageal pressure, during the course of a coo have only a marginal effect on the spectrum and, if esophageal inflation is part of
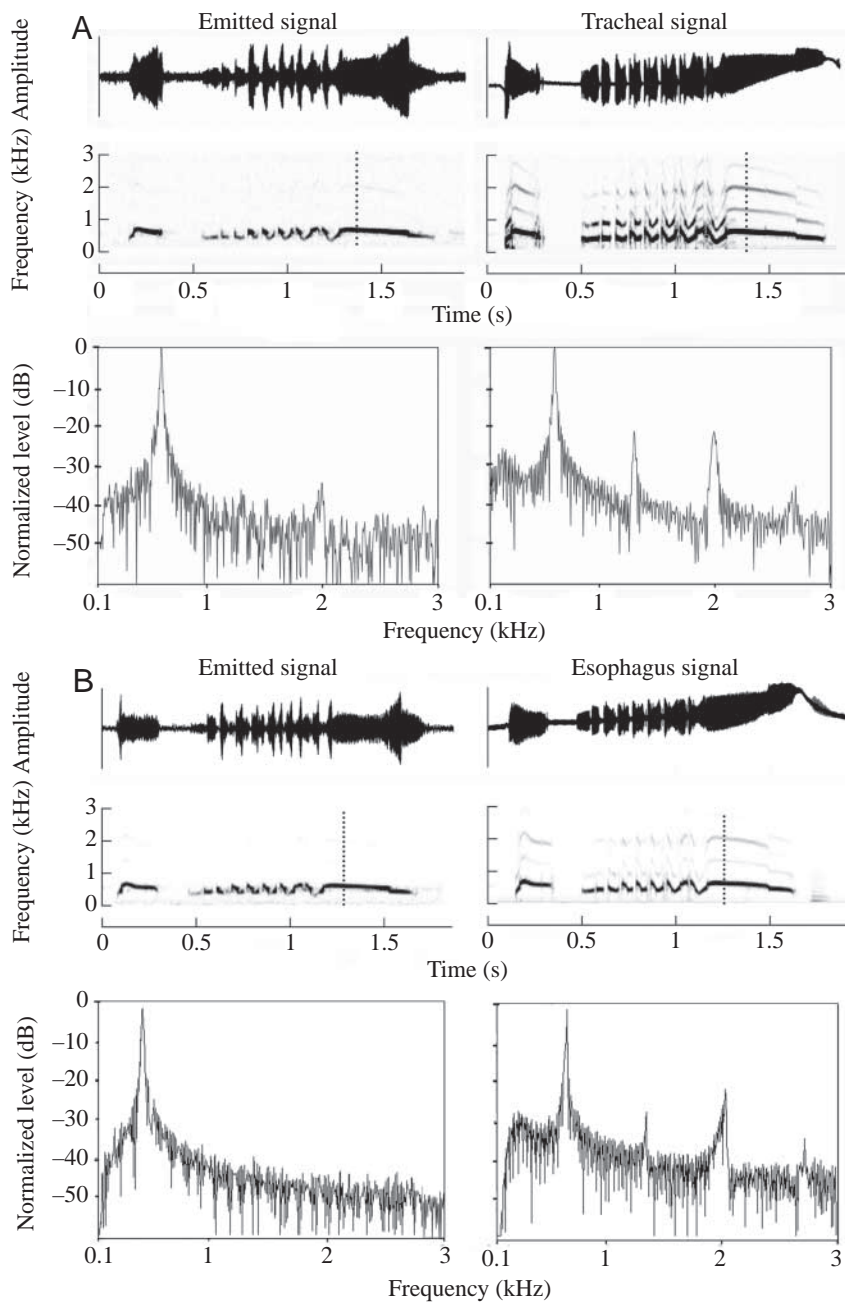

Fig. 3. Vocalizations and sounds recorded in different compartments of the vocal tract. (A) Emitted vocalization and tracheal sound from the same coo. (B) Emitted vocalization and esophageal sound from the same coo, uttered by a different dove. Oscillogram (top panels), spectrogram (middle panels) and power spectrum (bottom panels) of signals recorded in different compartments of the vocal tract. The DC offset in the $e2$ note of the tracheal and the esophageal signals is due to the pressure increase in the respective compartments. This offset can vary from coo to coo, although it appears similar in the examples shown. Coos in A and B come from different individuals.
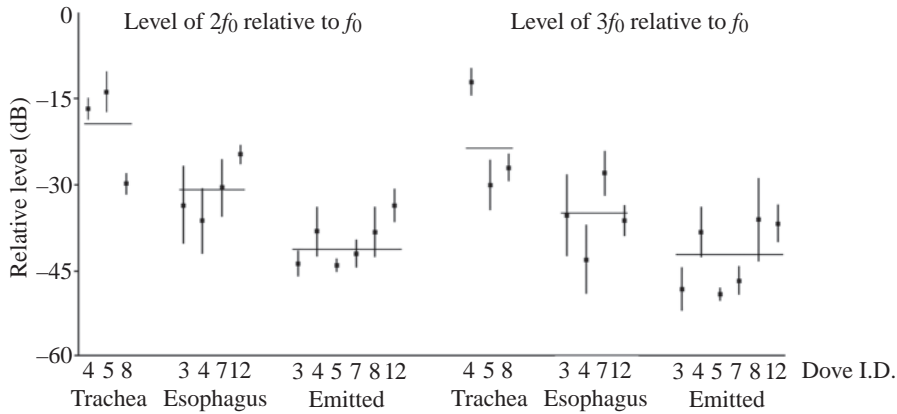
Fig. 4. Relative level of harmonics in the emitted vocalization and in the tracheal and esophageal sound. Each data point indicates the mean ± S.D. for individual birds. Horizontal lines indicate the mean sound level for the anatomical compartment, calculated from the mean individual values.

the filter mechanism, these changes in its volume are not large enough to defeat the filter.

Acoustic effects of preventing esophageal inflation

We further investigated the possible role of esophageal inflation as part of a vocal tract filter by inserting a tube through the esophageal wall and adjacent skin to open the esophagus to ambient pressure and prevent full inflation. After this treatment, we observed no visible external expansion during coo vocalizations, which indicates that the esophagus is not or is only slightly inflated. The back-and-forth movement of the head, which normally accompanies inflation of the esophagus immediately preceding a coo bout, was exaggerated when the tube was open.

Coos recorded when the tube was plugged (esophagus inflated) were compared with those produced when the tube was open (esophagus collapsed). The most prominent acoustic difference was a 13–17 dB increase in the level of $2f_0$ relative to f_0 and a 9–13 dB increase in the relative level of $3f_0$ when the esophagus was collapsed (Table 3; Fig. 6). This indicates that the inflation of the esophagus may be a necessary part of the filter. There were no significant differences in mean and maximum f_0, temporal parameters or mean and maximum sound levels of coos between treatments (Table 3).

Discussion

The results of our experiments provide four new findings that contribute to an understanding of the nature of the vocal tract filter in ring doves. (1) During vocalization, the inflated upper esophagus, but not the crop, forms a single continuous large chamber with the oral cavity. (2) Sound pressure measurements in the otherwise intact trachea and esophagus show that there is a substantial decrease in the relative levels of $2f_0$ and $3f_0$ in the inflated esophagus compared with those in the trachea and that overall the relative levels of these harmonics in the esophagus exceed those in the radiated vocalization. Only one bird (dove 4) formed an exception to the latter, possibly because
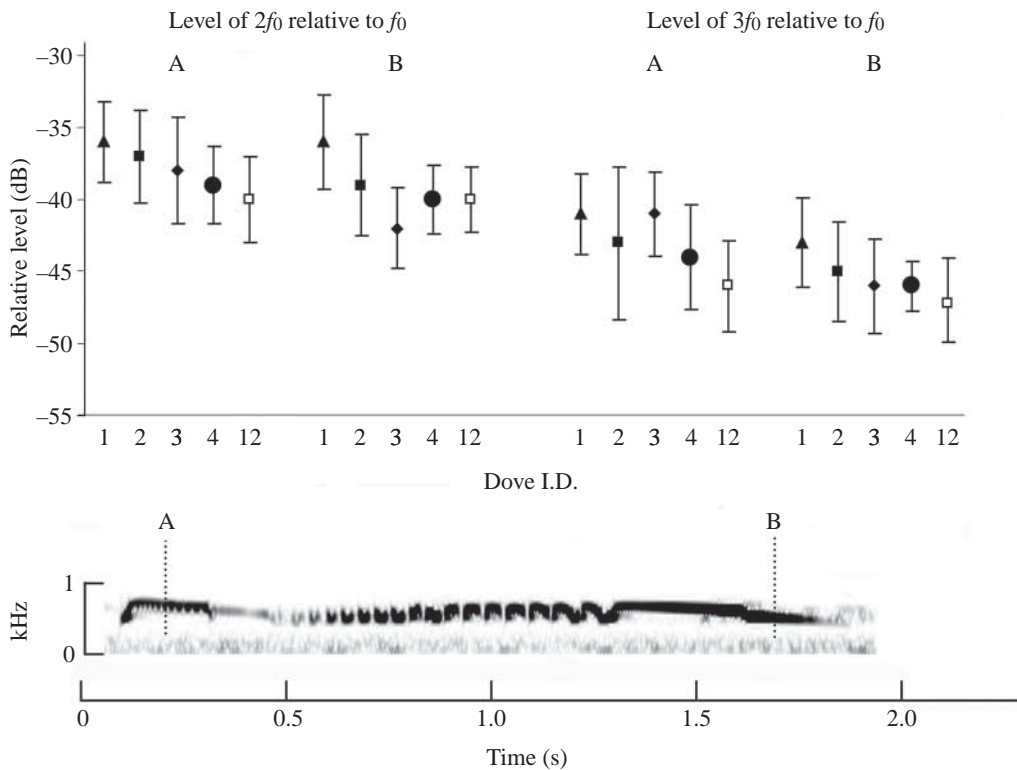

Fig. 5. Relative level of harmonics (upper panel) in the middle of the $e1$ note (measured around point A) and at the end of the $e2$ note (measured around point B). The lower panel shows the spectrogram of a coo with the points of measurements indicated by vertical dotted lines.

Table 2. *Levels of overtones at the beginning and the end of the coo*

Dove I.D.	Beginning of the call		End of the call	
	$2f_0$	$3f_0$	$2f_0$	$3f_0$
1	36±2.8	41±2.8	36±3.3	43±3.1
2	37±3.2	43±5.3	39±3.5	45±5.4
3	38±3.7	41±2.9	42±2.8	46±3.3
4	39±2.7	44±3.6	40±2.4	46±1.7
12	40±3.0	46±3.2	40±2.3	47±2.9

Levels of $2f_0$ and $3f_0$ (in dB re f_0; mean ± S.D.) relative to the level of the fundamental frequency (f_0) at the beginning and at the end of a coo call. In each of the five doves, 20 calls were considered.

it had simultaneously inserted tracheal and esophageal tubes, which may have affected the vocal tract mobility and acoustics. (3) The filter function of the partially inflated esophagus at the beginning of the coo is almost as strong as that of the fully inflated esophagus at the end of the coo. (4) Preventing inflation of the esophagus by inserting an open tube through

the esophageal wall and skin alters the suprasyringeal vocal tract filter function, as indicated by the relatively high levels of $2f_0$ and $3f_0$ in the radiated vocalization compared with those observed in the trachea. An acoustical model of the dove's vocal tract (Fletcher et al., in press) predicts that the inflated esophagus acts as an amplifier as well as a low pass filter. The fact that coo sound level did not increase with the appearance of harmonics may be due to loss of the amplifying function.

The spectrum of a periodically oscillating source is expected to show a negative slope. The source signal from the human vocal fold, for example, decreases by ~6–12 dB per octave (Titze, 1994). The vocal tract resonances selectively modulate the source signal, changing its spectrum. In the ring dove, we found a successive attenuation of the overtones from the trachea *via* esophagus to the emitted signal. These findings suggest that the filter mechanisms responsible for suppressing source-generated harmonics in ring dove coos occur in two stages. The first stage occurs when tracheal sound enters the (partially) inflated esophagus. Some further filtering out of harmonics occurs during transmission of esophageal sound to the outside environment. We hypothesize that the trachea and esophagus

Table 3. *Coo acoustics with and without esophagus inflation*

Coo acoustic	Dove I.D.	Closed tube	Open tube	
Mean call level (db)	1	62.2±0.7	62.0±1.3	$t=1$
	2	62±1.3	60±1.1	
	11	66±1.0	66±0.9	
Max. call level (db)	1	72±2.1	69±2.1	$t=1$
	2	72±2.9	69±2.2	
	11	73±0.5	73±0.4	
$e1$ note duration (ms)	1	0.23±0.03	0.22±0.04	$t=0.7$
	2	0.20±0.01	0.21±0.01	
	11	0.25±0.02	0.23±0.02	
Pause (p) duration (ms)	1	0.15±0.02	0.13±0.01	$t=0$
	2	0.13±0.02	0.12±0.02	
	11	0.18±0.02	0.21±0.03	
$e2$ note duration (ms)	1	1.4±0.09	1.38±0.07	$t=-0.48$
	2	1.20±0.06	1.25±0.1	
	11	1.1±0.1	1.1±0.1	
Total duration (s)	1	1.8±0.07	1.74±0.08	$t=-0.48$
	2	1.5±0.05	1.58±0.09	
	11	1.56±0.1	1.6±0.1	
Mean f_0 in $e2$ (Hz)	1	599±24	583±18	$t=-0.19$
	2	473±9	493±16	
	11	535±10	537±7.8	
Maximum f_0 in $e2$ (Hz)	1	687±13	660±20	$t=0.19$
	2	512±7	541±28	
	11	577±13	565±7.8	
Level $2f_0$ (dB) re f_0	1	−38±5	−25±4	$t=12.9**$
	2	−33±4	−16±10	
	11	−30±4	−15±8	
Level $3f_0$ (dB) re f_0	1	−42±4	−29±5	$t=6.9*$
	2	−42±2	−33±5	
	11	−45±3	−30±6	

Means ± S.D. for 10 acoustic parameters measured in the emitted signal from three doves while esophagus tube is closed (20 calls per dove) and while the esophagus tube was open (20 calls per dove). Comparisons have been carried out with pairwise t-test ($N=3$); *$P<0.05$; **$P<0.01$. Note that sound levels (dB) are not relative to a common standard and can be compared within an individual only.
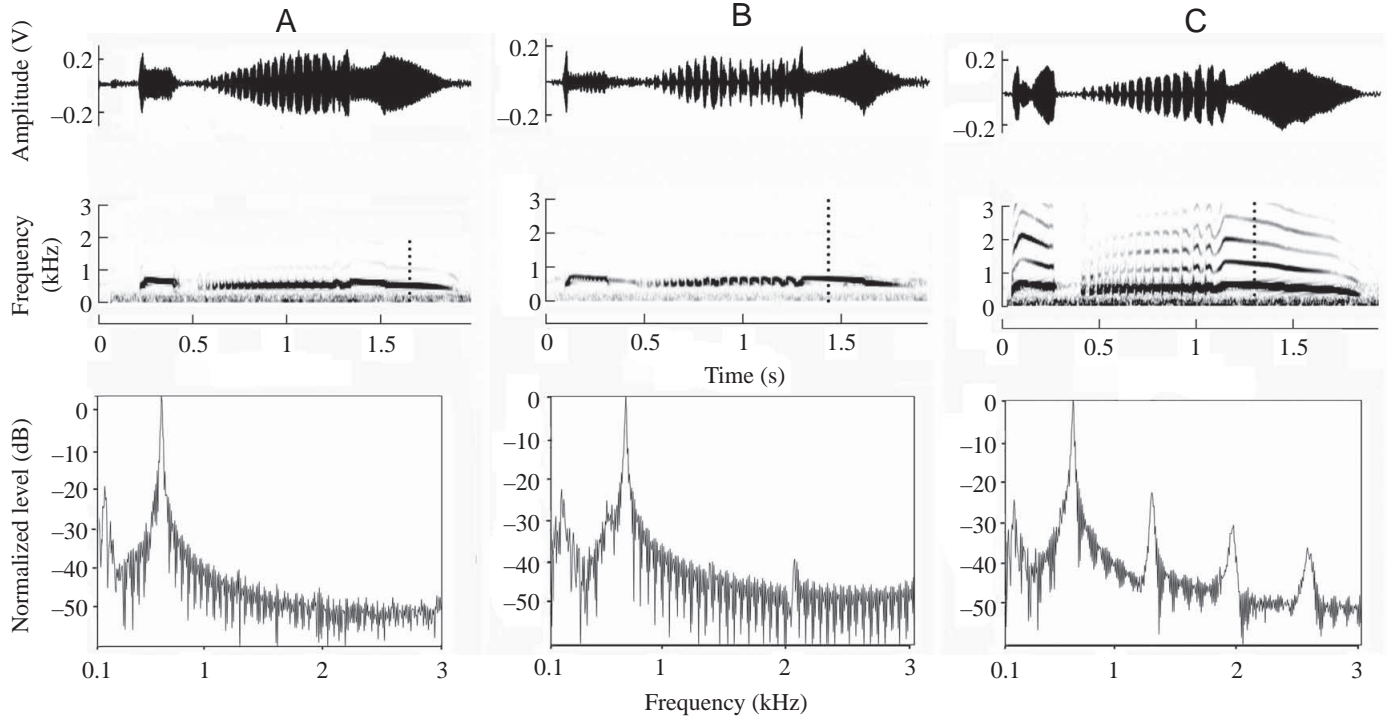
Fig. 6. Oscillograms and spectrograms of coo vocalizations (A) before esophageal tube was implanted, (B) after tube implant surgery and with closed tube and (C) after tube implant surgery and with open tube. The power spectra (bottom panels) are derived from a 100 ms segment during the second half of the *e*2 note, centered on the time indicated by the dotted line in the spectrogram.

behave as two acoustically separate compartments with different filter characteristics: first, the trachea as a uniform tube-like multi-band-pass filter with its lowest resonance tuned to the fundamental frequency of the coo by a laryngeal aperture adjustment and, second, the inflated esophagus with its wall and overlying skin acting as a Helmholtz resonator that is relatively independent of the amount of inflation.

The tracheal filter

The discrepancy between tracheal length and the quarter wavelength of the coo's f_o does not eliminate the possibility that tracheal resonance might support the f_o. The dove's trachea can be considered as a uniform tube (Ballintijn et al., 1995) closed at one end (the syrinx), as for instance modeled by Fletcher and Tarnopolsky (1999). The resonant frequencies (F) of a stopped tube are described by:

$$F_i = \frac{(2i-1)c}{4L} \ , \tag{1}$$

where $i=1$ to the number of formants of interest, c is the speed of sound (355 m s^{-1} in humidified air at ~40°C), L is the length of the tube (in m) and F_i is the frequency (in Hz) of the ith formant.

In ring doves, the distance between the tracheal bifurcation (the location of the syrinx) and the larynx is ~7.5 cm. Equation 1 predicts that the first resonant peak of a uniform stopped tube of this length will be 1180 Hz, almost twice the average f_o of a coo, which is ~600 Hz. The resonant frequency

might be lowered somewhat if the trachea is stretched as it is pushed laterally during inflation of the esophagus. Although a fresh postmortem trachea can be elongated ~30% by stretching, a 1 cm increase in length (to 8.5 cm) during phonation seems more realistic for a living bird. This latter length increase would shift the first resonance down to ~1045 Hz, which is still far from the 600 Hz average f_0 of the coo.

If, however, the bird constricts its laryngeal aperture during phonation, perhaps supplemented by modest elongation of the trachea, it might lower tracheal resonance by increasing the end loading of the trachea (Flanagan, 1972). The larynx acts as a protective valve at the cranial end of the trachea, and its aperture varies during feeding and respiration (Zweers, 1982; Zweers and Berkhoudt, 1988; Zweers et al., 1981).

Reducing the opening at the laryngeal end of a trachea that is closed at the syringeal end decreases its fundamental resonant frequency (F1) according to equation 2 (after equations 3.71 and 3.72 in Flanagan, 1972):

$$F_i = \left(1 + \frac{8A_t a}{3\pi A_m L}\right)^{-1} \left(\frac{(2i-1)c}{4L}\right) , \tag{2}$$

where $i=1$ to the number of formants of interest, A_t is the cross-sectional area of the trachea (radius = 2 mm), A_m is the cross-sectional area of the larynx (of the radiating area) (the thickness of the larynx constriction is assumed to be negligible); a is the radius of the cross-sectional area of the

larynx ('larynx opening'; maximum 2 mm); c is the speed of sound (355 m s^{-1} in humified air at ~40°C) and L is the length of the trachea (7.5 cm).

If we add a correction, x, for the thickness of the larynx, i.e. treat the larynx as a separate 4 mm-long tube added to the trachea, and replace the cross-sectional areas (A_t and A_m) with their respective radii (a and r_t), the equation becomes:

$$F_i = \left(1+\frac{8r_t^2(a+x)}{3\pi a^2 L}\right)^{-1}\left(\frac{(2i-1)c}{4L}\right). \tag{3}$$

Calculations on the effect of radiation load, i.e. the width of the aperture on the open end of the stopped tube model, indicate that if the laryngeal aperture during cooing is constricted to ~25% (Fig. 7) of the tracheal diameter it could reduce the first resonance peak of a 7.5 cm-long trachea to ~600 Hz, i.e. the f_0 of the dove's coo (Fig. 8). This calculation is only a rough estimate and includes simplifying assumptions that, for example, neglect the difference between a round *vs* a slit-like opening of the larynx. The accuracy of our estimates and the possible role of the larynx in tuning the trachea to the coo's f_0 need to be verified experimentally.

Support for a strong tracheal resonance might additionally come from the impedance change due to the sudden increase in vocal tract radius at the transition from the glottis to the inflated esophagus. If the diameters of adjacent tubes, or segments of the vocal tract, exceed a critical 6-fold ratio, they may decouple and produce separate autonomous resonances (Titze, 1994). As a result, a separate tracheal resonance peak may appear in the frequency spectrum.

There is some support in the literature for a laryngeal influence on tracheal resonance in birds. Rüppell reported that removing the larynx of a crane's trachea increased the first (?) resonance (he used the term 'Resonanzerscheinung') by an amount equivalent to a 12 cm increase in length (fig. 28 in Rüppell, 1933). Nowicki and Podos (Nowicki, 1987; Podos et al., 1995) also suggested that the larynx may play a role in tuning the vocal tract resonance for songbirds.

Changes in the laryngeal aperture might also explain the resonance-like peak around 600 Hz in the inspiratory 'wah' sounds that are produced during inhalation immediately following the end of the coo (Beckers et al., 2003). This sound
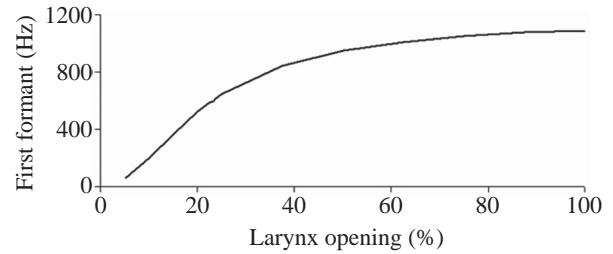
Fig. 8. Relationship between laryngeal opening and the first resonance (F1) of the trachea, calculated using equation 3 for a uniform tube closed at one end with a variable opening at the other end.

has multiple harmonics above an f_0 of 150–250 Hz. The f_0 is present at a high level inside the trachea but is almost completely absent in the emitted vocalization, which has a formant around 600 Hz (range 500–800 Hz; see fig. 3 in Beckers et al., 2003; Figs 1, 9 in present study). It is particularly interesting that the first formant (F1) of the wah sound can increase from ~500 to 850 Hz during the course of its emission (Fig. 9). It is tempting to speculate that this increase in the first resonance is caused by the opening of the larynx that accompanies inspiration (Zweers et al., 1981). Since the beak is open during inspiration, we assume that the filtering effect of a partly inflated esophagus is minimal, although the beak itself could also be part of the resonator.

A stopped, uniform tube supports the odd-numbered harmonics of its first resonance. If vocal tract resonance is unimportant, $3f_0$ should be 6–12 dB lower than $2f_0$, as predicted for the source spectrum. A first resonance at the dove's f_0 should be accompanied by a second resonance that increases the amplitude of $3f_0$ relative to $2f_0$. There is some evidence that this occurs since both the absolute amplitude of $3f_0$ (see table 2, 'normal air' signal in Ballintijn and ten Cate, 1998) and the relative amplitude of $3f_0$ (see fig. 2, 'emitted signal' in Beckers et al., 2003; Table 3 in present study) are larger than, or very similar to, those of $2f_0$.

If the air-collecting passages above the larynx are opened to the external environment, either by the bird opening its beak or by a tube implanted in the esophagus, the bandwidth of the emitted signal increases. In theory, this spectral change might be due to one or more of the following effects. (1) Sound emitted from the trachea can pass out of the bird through the beak or open esophageal tube and no longer has to pass through the esophageal wall and overlying skin. There are, at present, no measurements to support or refute this putative filter mechanism. (2) If the inflated esophageal chamber and its wall have a resonant frequency close to the bird's f_0, deflation will eliminate this effect. This hypothesis is consistent with the effect of an open esophageal tube and of the open beak on the inspiratory wah sound. However, the filter characteristics of the inflated esophagus remain to be determined. (3) When the suprasyringeal vocal tract is open to the outside, its pressure will remain close to ambient. The inability to pressurize the trachea might affect the source spectrum by altering the rate of syringeal airflow or the
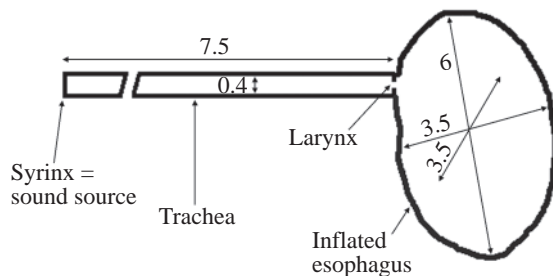

Fig. 7. Schematic drawing of the ring dove's suprasyringeal vocal tract during full inflation of the esophagus. Dimensions are in cm. Tracheal diameter varies somewhat along its length, and its cross-section is not perfectly circular.
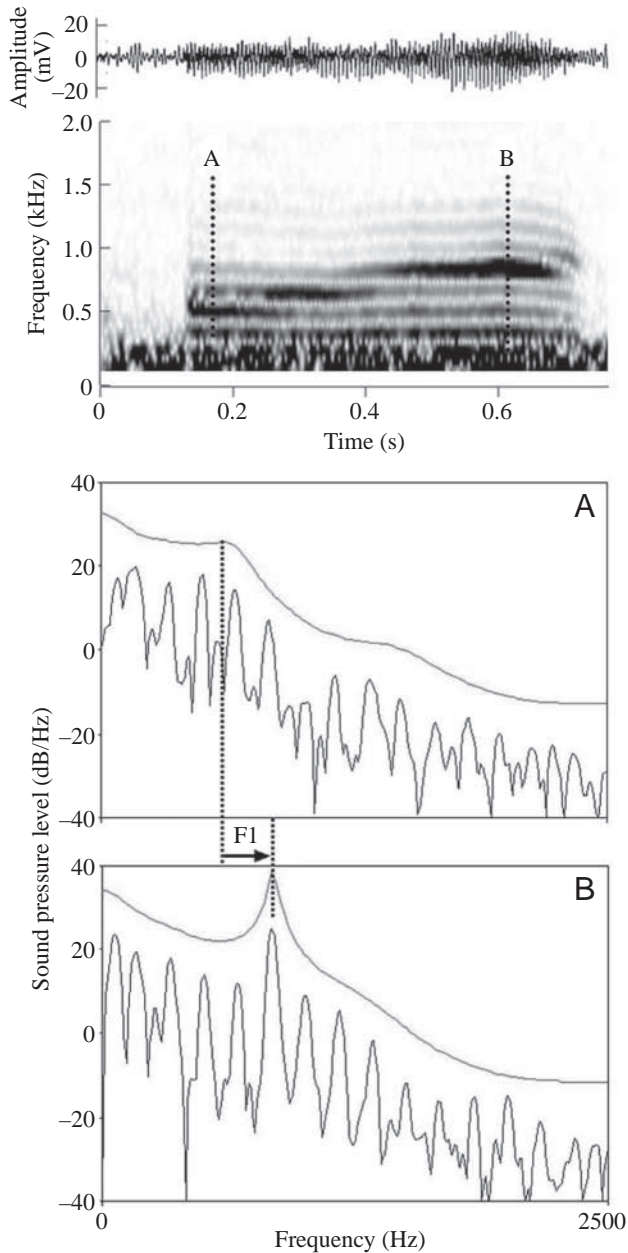
Fig. 9. Time series, spectrogram and spectra of an inspiratory sound. The spectra were derived from 50 ms segments centered around the dotted lines at the beginning and end of the call. The spectra A and B show the fast Fourier transform (FFT)-based spectrum (lower curve) and the linear predictive coding-based envelope of power spectrum (see Owren and Bernacki, 1998) indicating formant frequencies. While the fundamental frequency is increasing from 165 Hz to 170 Hz, the first resonance (F1) increases from 500 Hz to 850 Hz.

pressure gradient across the syrinx. However, it seems unlikely that changes in suprasyringeal pressure disrupt the normal filter mechanism because we observed similar filtering effects in the *e*1 note when the pressure was near ambient and in the *e*2 note when suprasyringeal pressure was highest (Fig. 5; Table 2).

The inflated esophageal filter

The fact that the harmonic content of the coo is little affected by changes in the inflation, and hence volume, of the esophagus argues against an esophageal filter based on cavity resonance. An alternative model to that of an air cavity surrounded by a thin layer of skin is a Helmholtz resonator of the kind in which the walls of the chamber, rather than the air in the exit pipe, form the vibrating mass. The analysis of these two kinds of resonator is otherwise the same (Fletcher, 1992). Like any simple resonator, the resonant frequency of the inflated esophagus depends upon a mass-like term and a spring-like term. The important mass in this model is the flexible wall, since this is much greater than the mass of the enclosed air or the mass of the co-moving air outside (Fletcher et al., in press).

Doves are an important model species in many studies of animal communication and it is evident that their distinctive low-frequency vocal signals are modulated by complex filter mechanisms. Although we are not yet able to provide a definitive detailed description of the vocal tract filter in doves, our experiments provide new morphological, physiological and acoustic information that takes us a step closer to understanding this remarkable vocal system.

References

Ballintijn, M. R. and ten Cate, C. (1997a). Sex differences in the vocalizations and syrinx of the collared dove (*Streptopelia decaocto*). *Auk* **22**, 22-39.

Ballintijn, M. R. and ten Cate, C. (1997b). Vocal development and its differentiation in a non-songbird: the collared dove (*Streptopelia decaocto*). *Behaviour* **134**, 595-621.

Ballintijn, M. R. and ten Cate, C. (1998). Sound production in the collared dove: a test of the 'whistle' hypothesis. *J. Exp. Biol.* **201**, 1637-1649.

Ballintijn, M. R., ten Cate, C., Nuijens, F. W. and Berkhoudt, H. (1995). The syrinx of the collared dove (*Streptopelia decaocto*): structure, inter-individual variation and development. *Neth. J. Zool.* **45**, 455-479.

Barel, A. O., Lambrecht, R. and Clarys, P. (1998). Mechanical function of the skin: state of the art. In *Skin Bioengineering, Techniques and Applications in Dermatology and Cosmetology*, vol. 26, *Current Problems in Dermatology* (ed. P. Elsner, A. O. Barel, E. Beradesea, B. Gabard and J. Serup), pp. 69-83. Basel: Karger.

Beckers, G. J. L., Suthers, R. A. and ten Cate, C. (2003). Pure-tone birdsong by resonance filtering of harmonic overtones. *Proc. Natl. Acad. Sci.. USA* **100**, 7372-7376.

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.

Cheng, M. F. (1992). For whom does the female dove coo? A case for the role of vocal self-stimulation. *Anim. Behav.* **43**, 1035-1044.

Cheng, M. F., Peng, J. P. and Johnson, P. (1998). Hypothalamic neurons preferentially respond to female nest coo stimulation: demonstration of direct acoustic stimulation of luteinizing hormone release. *J. Neurosci.* **18**, 5477-5489.

Clarke, L. F., Rahn, H. and Martin, M. D. (1942). Seasonal and sexual dimorphic variations in the so-called "air sac" region of the sage grouse. *Bull. Wyo. Game Fish Dep.* **2**, 13-27.

Cynx, J. and Shapiro, M. (1986). Perception of the missing fundamental by a species of songbird (*Sturnus vulgaris*). *J. Comp. Psychol.* **100**, 356-360.

Cynx, J., Williams, H. and Nottebohm, F. (1990). Timbre discrimination in zebra finch (*Taeniopygia guttata*) song syllables. *J. Comp. Physchol.* **104**, 303-308.

Dantzker, M. S., Grant, D. B. and Bradbury, J. W. (1999). Directional acoustic radiation in the strut display of male sage grouse *Centrocercus urophasianus*. *J. Exp. Biol.* **202**, 2893-2909.

Dooling, R. J., Leek, M. R., Gleich, O. and Dent, M. L. (2002). Auditory temporal resolution in birds: discrimination of harmonic complexes. *J. Acoust. Soc. Am.* **112**, 748-759.

Flanagan, J. (1972). *Speech Analysis, Synthesis and Perception*. New York: Springer-Verlag.

Fletcher, N. H. (1992). *Acoustic Systems in Biology*. New York: Oxford University Press.

Fletcher, N. H. and Tarnopolsky, A. (1999). Acoustics of the avian vocal tract. *J. Acoust. Soc. Am.* **105**, 35-49.

Fletcher, N. H., Riede, T., Beckers, G. J. L. and Suthers, R. A. (in press). Vocal tract filtering and the 'coo' of doves. *J. Acoust. Soc. Am.*

Gaunt, A. S., Gaunt, S. L. L. and Casey, R. M. (1982). Syringeal mechanics reassessed: evidence from *Streptopelia*. *Auk* **99**, 474-494.

Gautier, J. P. (1971). Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopithecinae; liason avec les cris d'espacement. *Biol. Gabon.* **7**, 229-267.

Goller, F. and Larsen, O. N. (1997). *In situ* biomechanics of the syrinx and sound generation in pigeons. *J. Exp. Biol.* **200**, 2165-2176.

Goller, F., Mallinckrodt, M. J. and Torti, S. D. (2004). Beak gape dynamics during song in the zebra finch. *J. Neurobiol.* **59**, 289-303.

Hienz, R. D., Sachs, M. B. and Sinnott, J. M. (1981). Discrimination of steady-state vowels by blackbirds and pigeons. *J. Acoust. Soc. Am.* **70**, 699-706.

Hoese, W. J., Podos, J., Boetticher, N. C. and Nowicki, S. (2000). Vocal tract function in birdsong production: experimental manipulation of beak movements. *J. Exp. Biol.* **203**, 1845-1855.

Honess, R. F. and Allred, W. J. (1942). Structure and function of the neck muscles in inflation and deflation of the esophagus in the sage grouse. *Bull. Wyo. Game Fish. Dep.* **2**, 5-12.

Jaramillo, C., Rand, A. S., Ibanez, R. and Dudley, R. (1997). Elastic structures in the vocalization apparatus of the Tungara frog (*Physalaemus pustulosus*) Leptodactylidae. *J. Morph.* **233**, 287-295.

Larsen, O. N. and Goller, F. (1999). Role of syringeal vibrations in bird vocalizations. *Proc. Roy. Soc. Lond.* **266**, 1609-1615.

Lohr, B. and Dooling, R. J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *J. Comp. Psychol.* **112**, 36-47.

Markel, J. D. and Gray, A. H. (1976). *Linear Prediction of Speech*. New York: Springer-Verlag.

McLelland, J. (1989). Digestive system. In *Form and Function in Birds*, vol. 1 (ed. A. S. King and J. McLelland), pp. 69-181. New York: Academic Press.

Miller, W. J. and Miller, L. S. (1958). Synopsis of behaviour traits of the ring-neck dove. *Anim. Behav.* **6**, 3-6.

Nottebohm, F. and Nottebohm, M. E. (1971). Vocalizations and breeding behaviour of surgically deafened ring doves (*Streptopelia risoria*). *Anim. Behav.* **19**, 313-327.

Nowicki, S. (1987). Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature* **325**, 53-55.

Nowicki, S., Marler, P., Maynard, A. and Peters, S. (1992). Is the tonal quality of birdsong learned – evidence from song sparrows. *Ethology* **90**, 225-235.

Owren, M. J. and Bernacki, R. H. (1998). Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals. In *Animals Acoustic Communication* (ed. S. L. Hopp, M. J. Owren and C. S. Evans), pp. 129-162. Berlin: Springer-Verlag.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in a study of vowels. *J. Acoust. Soc. Am.* **24**, 175-184.

Podos, J., Sherer, J. K., Peters, S. and Nowicki, S. (1995). Ontogeny of vocal-tract movements during song production in song sparrows. *Anim. Behav.* **50**, 1287-1296.

Rand, A. S. and Dudley, R. (1993). Frogs in helium: the anuran vocal sac is not a cavity resonator. *Physiol. Zool.* **66**, 793-806.

Rüppell, V. W. (1933). Physiologie und Akustik der Vogelstimme. *J. Ornithol.* **81**, 433-542.

Suthers, R. A. and Goller, F. (1997). Motor correlates of vocal diversity in songbirds. In *Current Ornithology*, vol. 14 (ed. V. Nolan, Jr, E. Ketterson and C. F. Thompson), pp. 235-288. New York: Plenum Press.

Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall.

Westneat, M. W., Long, J., John H., Hoese, W. and Nowicki, S. (1993). Kinematics of birdsong: functional correlation of cranial movements and acoustic features in sparrows. *J. Exp. Biol.* **182**, 147-171.

Williams, H., Cynx, J. and Nottebohm, F. (1989). Timbre control in zebra finch (*Taeniopygia guttata*) song syllables. *J. Comp. Psychol.* **103**, 366-380.

Zweers, G. A. (1982). The feeding system of the pigeon (*Columba livia* L.). *Adv. Anat. Embryol. Cell Biol.* **73**, 1-108.

Zweers, G. A. and Berkhoudt, H. (1988). Larynx and pharynx of crows (*Corvus corone* L. and *Corvus monedula* L., Passeriformes: Corvidae). *Neth. J. Zool.* **37**, 365-393.

Zweers, G. A., van Pelt, H. C. and Beckers, A. (1981). Morphology and mechanics of the larynx of the pigeon (*Columba livia L.*): a drill-chuck system (Aves). *Zoomorphology* **99**, 37-69.