Journal of Physiology - Paris xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Physiology - Paris

journal homepage: www.elsevier.com/locate/jphysparis

Original Research Paper

An integrated model for motor control of song in *Serinus canaria*

Rodrigo Gogui Alonso *, Ana Amador, Gabriel B. Mindlin

Physics Department, FCEyN, Universidad de Buenos Aires, and IFIBA Conicet, Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 27 May 2016
Received in revised form 25 November 2016
Accepted 1 December 2016
Available online xxxx

Keywords:
Birdsong production
Dynamical systems
Nonlinear phenomena
Biomechanical models
Neural control

ABSTRACT

Birdsong is a learned motor behavior controlled by an interconnected structure of neural nuclei. This pathway is bilaterally organized, with anatomically indistinguishable structures in each brain hemisphere. In this work, we present a computational model whose variables are the average activities of different neural nuclei of the song system of oscine birds. Two of the variables are linked to the air sac pressure and the tension of the labia during canary song production. We show that these time dependent gestures are capable of driving a model of the vocal organ to synthesize realistic canary like songs.

1. Introduction

Birdsong production is a complex behavior that emerges when a highly specialized peripheral vocal organ, the syrinx, is driven by a set of well-coordinated physiological instructions. In particular, the syringeal configuration and the respiratory activity are tuned so that the proper airflow necessary for phonation is established (e.g. Mindlin and Laje, 2005; Schmidt and Wild, 2014; Suthers and Zollinger, 2004). These physiological instructions are generated by a neural circuitry which is reasonably well characterized (e.g. Nottebohm et al., 1976; Wild, 1997; Wild et al., 2000). The song production circuit for oscine birds such as canaries (*Serinus Canaria*) includes the dorsomedial nucleus (DM), which projects to the respiratory pacemaker circuit (nucleus retroambigualis, RAm and nucleus parambigualis, PAm) and the tracheosyringeal motor nucleus (*nXII*). These structures are also found in non-oscines (Wild et al., 1997). What probably allows oscine birds to generate far more complex and rich vocalizations than non-oscines is an additional set of neural nuclei. In oscines, DM connects through n. uvaeformis (Uva) in the thalamus to the telencephalic nucleus HVC (used as proper name), which, in turn, connects to the robust arcopallial nucleus (RA). RA projects to the respiratory pacemaker and *nXII*, closing an indirect pathway that has not yet been found in non-oscine birds. Fig. 1 summarizes this description. All these nuclei are necessary for song production given that lesions in any of them dramatically affect song. It is worth mentioning that other neural nuclei could be included in this description (e.g. see Fig. 1 in Ashmore et al., 2005) but in this work we consider the minimal neural architecture that exhibits looped connectivity and that is able to generate the appropriate behavior, i.e. song. However, the amount of knowledge on the anatomy and function of each of these nuclei differs substantially because measuring their activity during song production varies in difficulty (Sturdy et al., 2003).

The physiological instructions required for birdsong production have been thoroughly studied (e.g. Gardner et al., 2001; Goller and Larsen, 1997; Goller and Suthers, 1995; Goller and Suthers, 1996a, b; Mindlin and Laje, 2005; Suthers, 1990). A sufficiently high pressure pulse through the labia (the avian analog to vocal folds) in a prephonatory position can generate auto-sustained oscillations that modulate airflow to generate sound. Recently, further evidence was put forward in favor of a myoelastic-aerodynamic mechanism for avian phonation, the same mechanism used to produce human speech (Elemans et al., 2015; Titze, 1988). Expiratory and inspiratory muscles are highly coordinated to generate song. In canary song, a large repertoire and acoustic diversity can be achieved using only four types of expiratory pressure pulses, which are highly conserved across individuals and colonies (Alliende et al., 2010). An example of a typical canary song that makes use of all four types of pressure patterns is shown in Fig. 2.

The level of complexity of the neural pathway could easily dissuade anyone from trying to model the dynamics of the song system. Nevertheless, in the case of canaries, it has been reported that the pressure patterns used during song can be reproduced as the solutions of a low dimensional nonlinear dynamical system (see Fig. 3) (Alonso et al., 2009; Amador and Mindlin, 2014; Trevisan

* Corresponding author at: Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina.
E-mail address: ralonso@df.uba.ar (R.G. Alonso).

http://dx.doi.org/10.1016/j.jphysparis.2016.12.003
0928-4257/

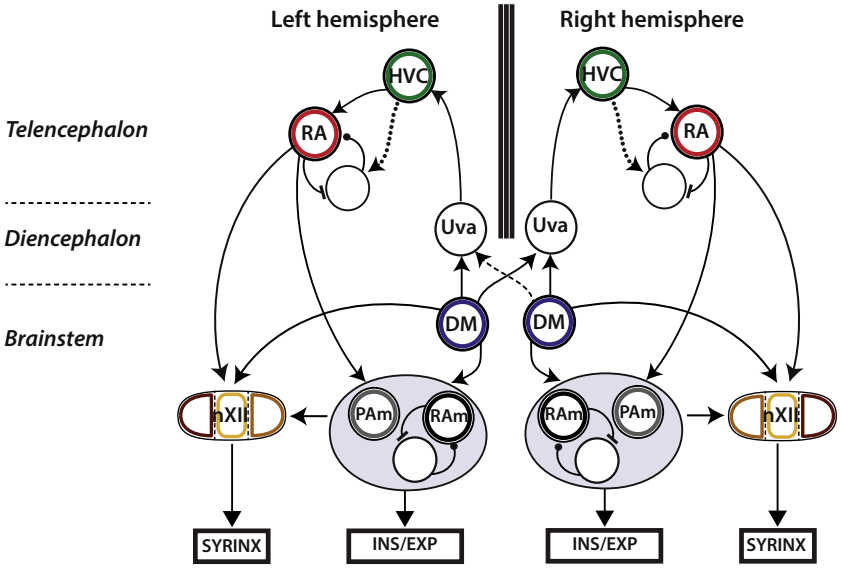

Fig. 1. Diagram of the architecture of the song system. The circles represent the nuclei that participate in our description. The arrows represent the connections between different nuclei. The abbreviations correspond to those in the text. Note the bilateral and circular nature of the song system's architecture, a central feature of the model presented in this manuscript.

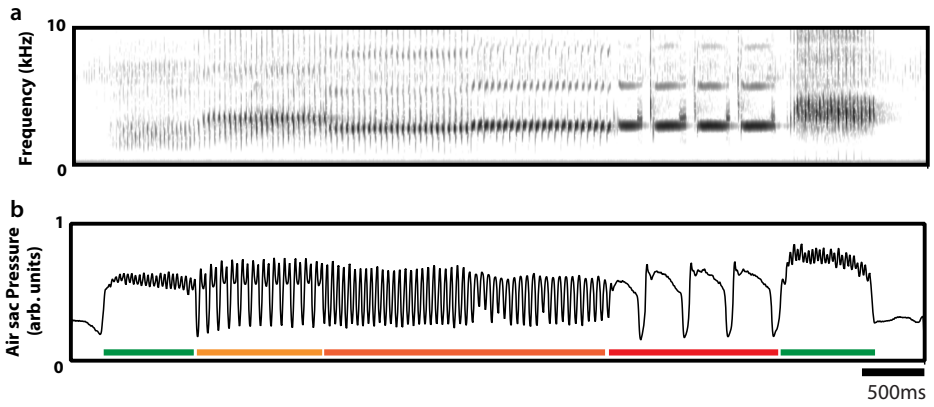

Fig. 2. Experimental measurement of typical canary song and air sac pressure. (a) Experimental sonogram of a complete canary song. (b) Air sac pressure patterns produced during song. Colored lines below the pressure patterns represent different types of pressure gestures. We use the same color coding as in Fig. 3.

et al., 2006). In this figure, we show an experimental time series of air sac pressure and its synthetic proxy obtained as the solution of a low dimensional nonlinear system. The model was built as the neural oscillator shown in Fig. 3b. Note that, in order to generate the different patterns, the parameters have to be varied as shown in the bottom panel of Fig. 3b. As is usually the case in nonlinear systems, what determines the qualitative nature of the solutions is not a singular choice of the parameters but open sets of the parameter space in which topological features are preserved. For details see Alonso et al. (2015).

Assuming that the expiratory pressure pulses during phonation reflect the average activity of the expiratory related areas of the respiratory pacemaker (RAm), the following questions can be posed: 1. Is it possible to formulate a nonlinear model for the activity of the expiratory related area compatible with its anatomy that reproduces the observed respiratory patterns when driven by time dependent functions? 2. Can those input functions be generated by the neural architecture described above? These questions led to a computational model (Alonso et al., 2015) capable of reproducing the air sac pressure patterns measured during song production (Alliende et al., 2010) as well as explaining the reported deformation of the pressure patterns during cooling (Goldin et al., 2013).

A distributed model allows us to reproduce the activity in one specific nucleus at the expense of having to constrain the activities of the other nuclei in the song pathway. Therefore, if try to include in our description the activity of the syringeal motor nucleus *nXII*, severe restrictions will arise from the fact that it coexists with a given respiratory pattern. Each respiratory pattern is the result of a whole set of nuclei presenting specific activities. Furthermore, some of the nuclei that participate in the generation of the respiratory pattern are inputs to *nXII* as well. It is then natural to ask whether this integrated, highly constrained architecture can generate the appropriate *nXII* activity patterns in order to generate canary song. In this work we test the hypothesis that, by adjusting only the relative weights of the inputs to *nXII*, the distributed activity patterns associated to a pressure pulse will drive *nXII* to generate the appropriate solutions, i.e. sound with a plausible temporal modulation of the fundamental frequency.

2. Materials and methods

2.1. Conceptual description of the integrative model of the song system

The empirical model used to generate the canary pressure patterns is built in a modular way. The first module in the model is a

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

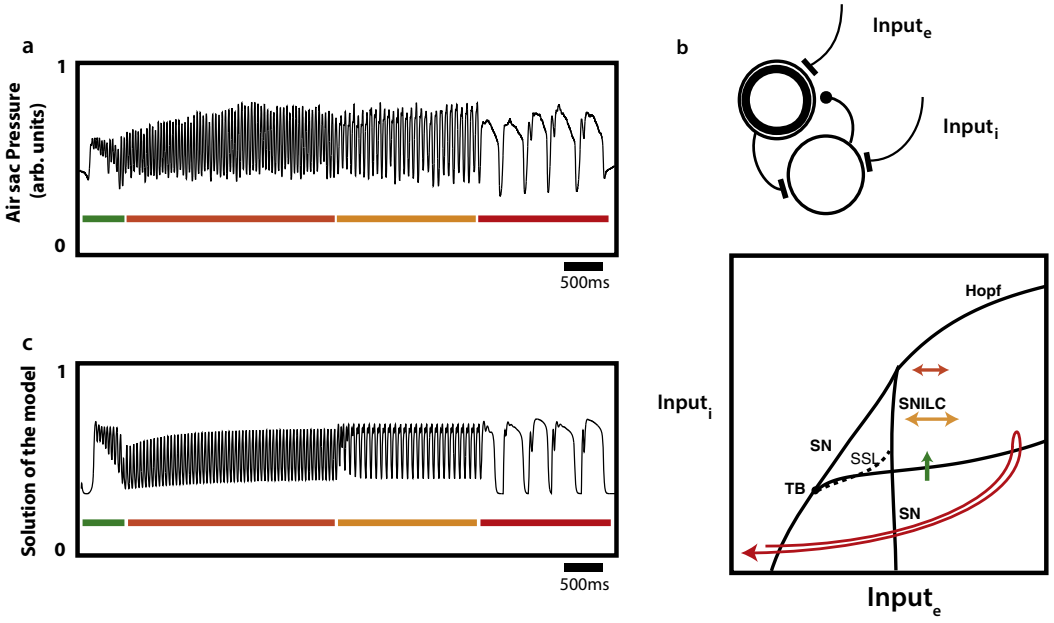

Fig. 3. The dynamical origin of the respiratory patterns used in canary song. (a) Experimental time series of air sac pressure during song production. Colored lines below the pressure patterns represent different types of pressure gestures, as in Fig. 2b. (b) (Top panel) Basic dynamical model capable of reproducing the different types of pressure gestures. It consists of a forced neural oscillator, i.e. a set of interconnected excitatory and inhibitory neurons. (b) (Bottom panel) Bifurcation diagram of the model illustrated in the top panel in terms of the inhibitory and excitatory inputs. Black lines represent bifurcations that separate regions of the parameter space with qualitative different solutions. (c) Solutions of integrating the dynamical system displayed in (b). Colored arrows in (b) indicate the different explorations of the input space required to reproduce the corresponding pressure gesture types in (c). The colors of the arrows match the colors below the corresponding patterns.

neural motif for the expiratory related area (putatively RAm), with excitatory and inhibitory populations (see Fig. 3b). We require that the activity of the excitatory population fits the measured pressure patterns (see Fig. 3a). Details of the temporal forcing that is necessary in order to obtain the appropriate pressure patterns (bottom of Fig. 3c) can be found in Amador and Mindlin (2014).

This basic motif is embedded within a larger set of nuclei in the circular model, and more than one nucleus affects its dynamics (see Fig. 1). The expiratory related motif receives inputs from an "initiating area" in the brainstem (hereafter referred to as IA; putatively DM). The model assumes that the IA also projects through Uva to the telencephalon. In this description, the first stage of the telencephalic part of the song system is HVC, followed by RA. The activity of HVC is modeled as a constant basal component that could be different for different syllables plus bursts elicited by the inputs from the thalamus. HVC activity then serves as input to RA. The activity of the excitatory population in RA is fed back to the expiratory related area, closing a "circular" description of the song system. This model was described in Alonso et al. (2015), and its solutions compared with the typical pressure patterns used by canaries during song production. The inspiratory related area (IR) (putatively Pam), is active during silent periods, and receives inputs from RA and the expiratory related area. The nucleus *nXII* is modeled as an excitatory population which receives inputs from RA, the expiratory related area, the inspiratory related area and the IA.

2.2. Mathematical implementation of the integrative model of the song system

The variables in our model are the set-average activities of the neurons in a set of interconnected areas (Hoppensteadt and Izhikevich, 1997). The dynamics for these variables are ruled by one of the simplest time continuous neural network models, the additive network:

$$\frac{dx_i}{dt} = -x_i + S\left(\rho_i + \sum_j \alpha_{ij} x_j\right)$$

$$S(x) = \frac{1}{1+e^{-x}}$$

where x_i describes the averaged activity of the neurons in the ith area, ρ_i the external input to the ith area, and the coefficients a_{ij} describe the connectivity between the regions.

Fig. 4b illustrates the way in which our neural model generates the instructions needed to synthesize the sound shown in Fig. 4a. We start our description with a pulse of activity in the initiating area.

Mathematically, the activity in the IA is represented by a square function a few milliseconds long. The blue circle in Fig. 4b represents the IA and its activity is one of the two inputs that arrive in the expiratory related area. We represent the passage through the thalamus by delaying the activation with respect to the burst in the initiating area. The green circle represents the nucleus HVC. We do not write a dynamical system model for HVC. Instead, we propose that its activity consists of a constant level of average activity (putatively originated within the nucleus) and a burst consequence of the input from the brainstem. The latter is mathematically expressed as a square function about 20–40 ms wide, delayed with respect to the burst in brainstem. For two classes of syllables (pulsatile and P1 solutions), the amplitude of the final pressure fluctuations could be better fitted if an additional fluctuation in HVC activity was included (green dashed lines in Figs. 6 and 7). These elements were added and used as inputs of the RA nucleus.

The activities that were actually simulated (i.e., that are computed as the solutions of a dynamical system) were the excitatory and inhibitory populations in RA (e_{ra},i_{ra}), the excitatory and inhibitory populations in the expiratory related area (e_{er},i_{er}), the excitatory population in the inspiratory related area (e_{ir}) and the

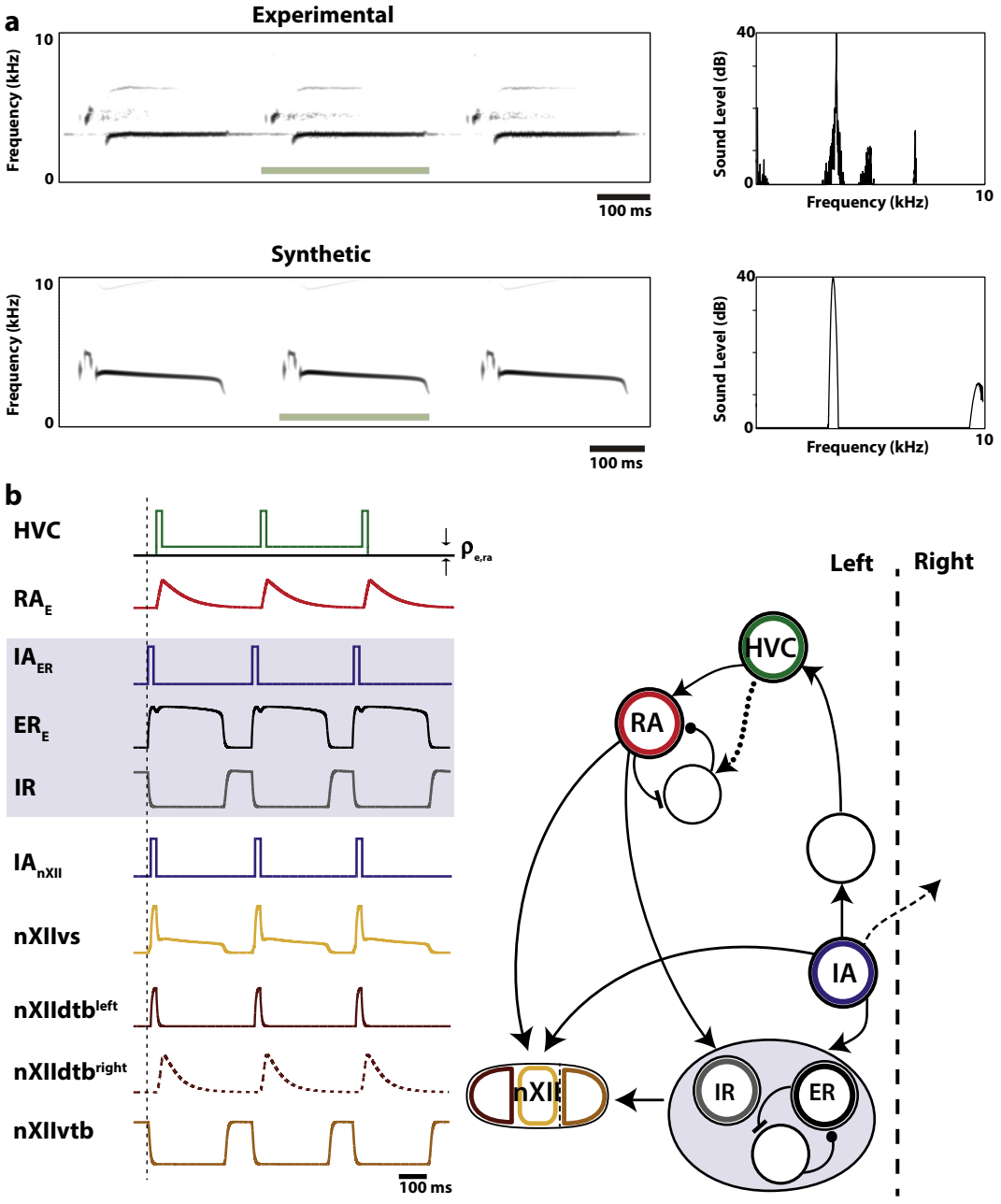

Fig. 4. P0 canary syllables reproduced with a low dimensional model. (a) (Left) The sonogram of a set of syllables produced with P0 respiratory patterns, and the result of integrating our neuro-biomechanical model (i.e., the neural model driving the biomechanical model). (Right) Representation of spectral properties (sound level vs. sound frequency) of synthetic and experimental syllables marked with the green line. Sound Analysis Pro Similarity Score: 96.23% Similarity, 94.13% Accuracy, Mean Values - Symmetric comparison. In (b) we display the neural architecture assumed by our model. The colors of the nuclei correspond to the colors of the time series of their average activity. The arrows represent the connections between different nuclei. Dashed arrows represent contralateral connections. In our model, HVC consists of a constant term (associated with a basal uniform activity ρ_{era}), plus the input from the brainstem, which is illustrated in the figure. The simulations start from square pulses in the initiating area (IA). The rest of the time traces correspond to the numerical simulations of the model. Both hemispheres participate in the construction of the P0 solution. According to the model, lesions in HVC would destroy the second part of the syllable (the long whistle). The labels of the time traces are defined in the text.

excitatory populations in *nXII* ($e_{nXII\,vs}, e_{nXIIdtb}, e_{nXII\,vtb}$), the nucleus which contains motor neurons that innervate the syringeal muscles. The dynamical system ruling their dynamics reads as:

$$\begin{cases} \frac{de_{ra}}{dt} = 20(-e_{ra} + S(\rho_{e,ra} + \alpha_{era,Fd} F_{delayed} + \alpha_{era,era} e_{ra} + \alpha_{era,ira} i_{ra})) \\ \frac{di_{ra}}{dt} = 20(-i_{ra} + S(\rho_{i,ra} + \alpha_{ira,Fd2} F_{delayed2} + \alpha_{ira,era} e_{ra} + \alpha_{ira,ira} i_{ra})) \end{cases}$$

$$\begin{cases} \frac{de_{er}}{dt} = 250(-e_{er} + S(\rho_{e,er} + \alpha_{eer,ra} e_{ra} + \alpha_{eer,F} F + \alpha_{eer,eer} e_{er} + \alpha_{eer,ier} i_{er})) \\ \frac{di_{er}}{dt} = 250(-i_{er} + S(\rho_{i,er} + \alpha_{ier,ra} e_{ra} + \alpha_{ier,eer} e_{er} + \alpha_{ier,ier} i_{er})) \\ \frac{de_{ir}}{dt} = 250(-e_{ir} + S(\rho_{e,ir} + \alpha_{eir,ra} e_{ra} + \alpha_{eir,eer} e_{er})) \end{cases}$$

$$\begin{cases} \frac{de_{nXII\,vs}}{dt} = 250(-e_{nXII\,vs} + S(\rho_{e,nXII} + \alpha_{nXII\,vs,FnXII} F_{nXII} + \alpha_{nXII\,vs,era} e_{ra} \\ \qquad + \alpha_{nXII\,vs,eer} e_{er} + \alpha_{nXII\,vs,eir} e_{ir})) \\ \frac{de^{l,r}_{nXIIdtb}}{dt} = 250(-e^{l,r}_{nXIIdtb} + S(\rho_{e,nXII} + \alpha^{l,r}_{nXIIdtb,FnXII} F_{nXII} + \alpha^{l,r}_{nXIIdtb,era} e_{ra})) \\ \frac{de_{nXII\,vtb}}{dt} = 250(-e_{nXII\,vtb} + S(\rho_{e,nXII} + \alpha_{enXII\,vtb,eir} e_{ir})) \end{cases}$$

where $\alpha_{i,j}$ represents the influence of the *j*th population on the *i*th dynamics. The parameters ρ account for constant inputs to the modeled populations: $\rho_{p,ra}$ represents the constant component input to the *p*th population of RA (*p* = *e*, *i*, either excitatory or inhibitory), arriving from HVC. The *Fs* parameters represent the inputs

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

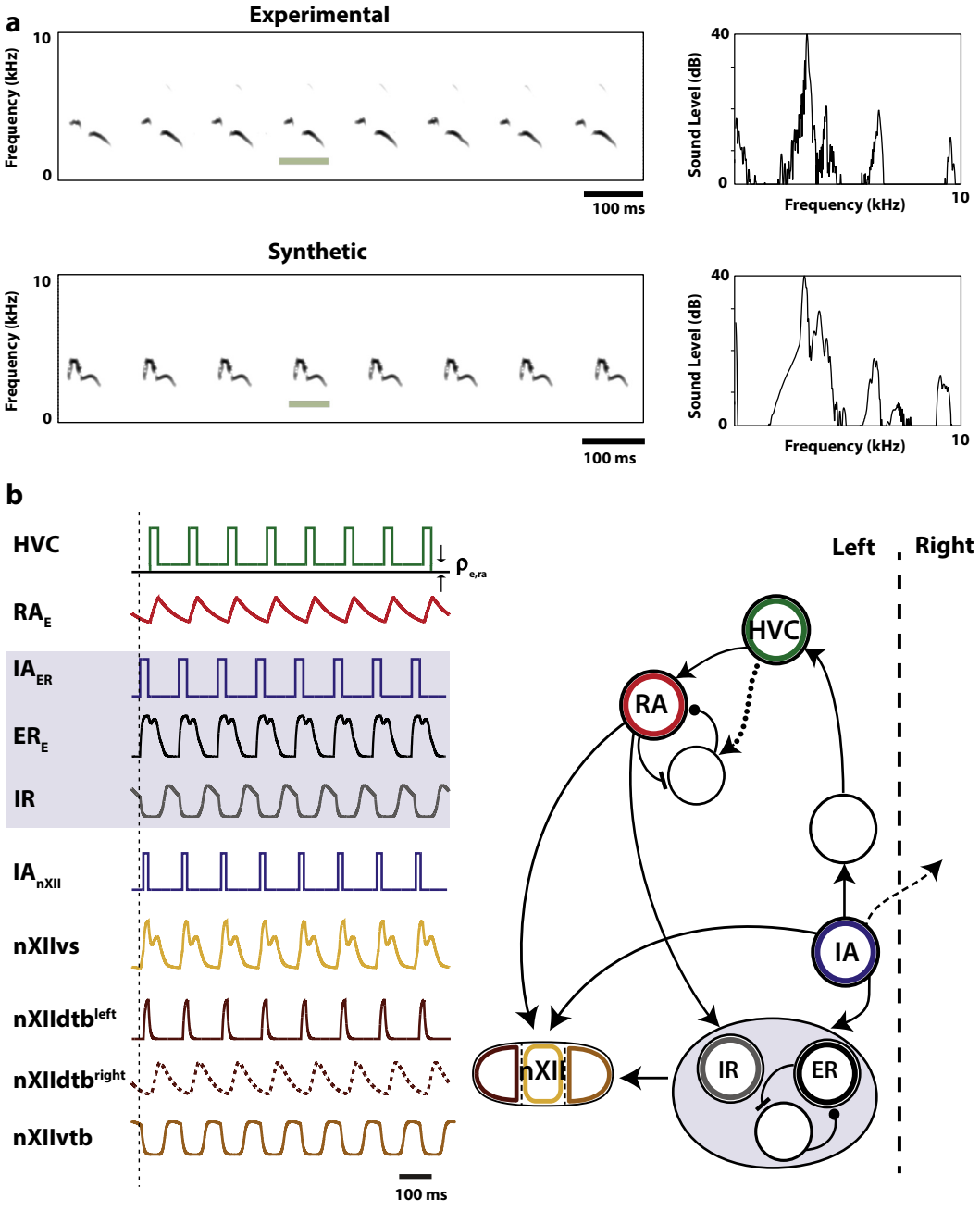

Fig. 5. The model also reproduces the P2 solutions. (a) (Left) As in Fig. 4, the sonogram of a set of syllables produced with P2 respiratory patterns, and the result of integrating our neuro-biomechanical model. (Right) Representation of spectral properties (sound level vs. sound frequency) of synthetic and experimental syllables marked with the green line. (Sound Analysis Pro Similarity Score: 99.78% Similarity, 80.02% Accuracy, Mean Values - Symmetric comparison). In (b) we show the neural architecture assumed by our model. HVC consists of a constant term (associated with a basal uniform activity ρ_{era}), plus the input from the brainstem, which is shown in the figure. As in the P0 solutions, both brain hemispheres participate in the construction of P2 solutions. According to the model, lesions in HVC would destroy the second part of the syllable.

associated with the activity originated in the IA. In particular, *F* and F_{nXII} stand for the direct input from the IA to the expiratory related area and *nXII* respectively, while $F_{delayed}$ is the indirect component of the input to RA that arrives through HVC after being originated in the IA. In our simulations, both $F_{delayed}$ and $F_{delayed2}$ stand for bursts of activity in the area that represents HVC. *F* is modeled as a square function. F_{nXII} and $F_{delayed}$ are square functions as well, but are delayed with respect to *F*. We introduce these delays to reflect the time needed for the activity originated in the IA to travel to the target nucleus; they are 10 ms for F_{nXII} and 30 ms for $F_{delayed}$. Note that $F_{delayed}$ has to travel through the Uva and HVC to reach RA, while F_{nXII} activates a direct connection between the IA and *nXII* at the level of the brainstem (see Fig. 1). The amplitude of the pulses was set to 10 (arb. units) for all the simulated patterns. The output of this system is the activity of *nXIIvs*, the activity of *nXIIdtb*, the activity of *nXIIvtb* and the activity of the expiratory related area (ER). The activity of *nXIIvs* affects the configuration of the avian vocal organ through the activation of intrinsic muscles that control the fundamental frequency of the vocalization. The activity of *nXIIdtb* and *nXIIvtb* affects the configuration of the avian vocal organ through the activation of intrinsic muscles that control the gating of the syrinx, and the expiratory related area generates the expiratory pressure pulses required for phonation (see Section 2.3 for a complete description of the syrinx). The anatomy of this pathway is

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

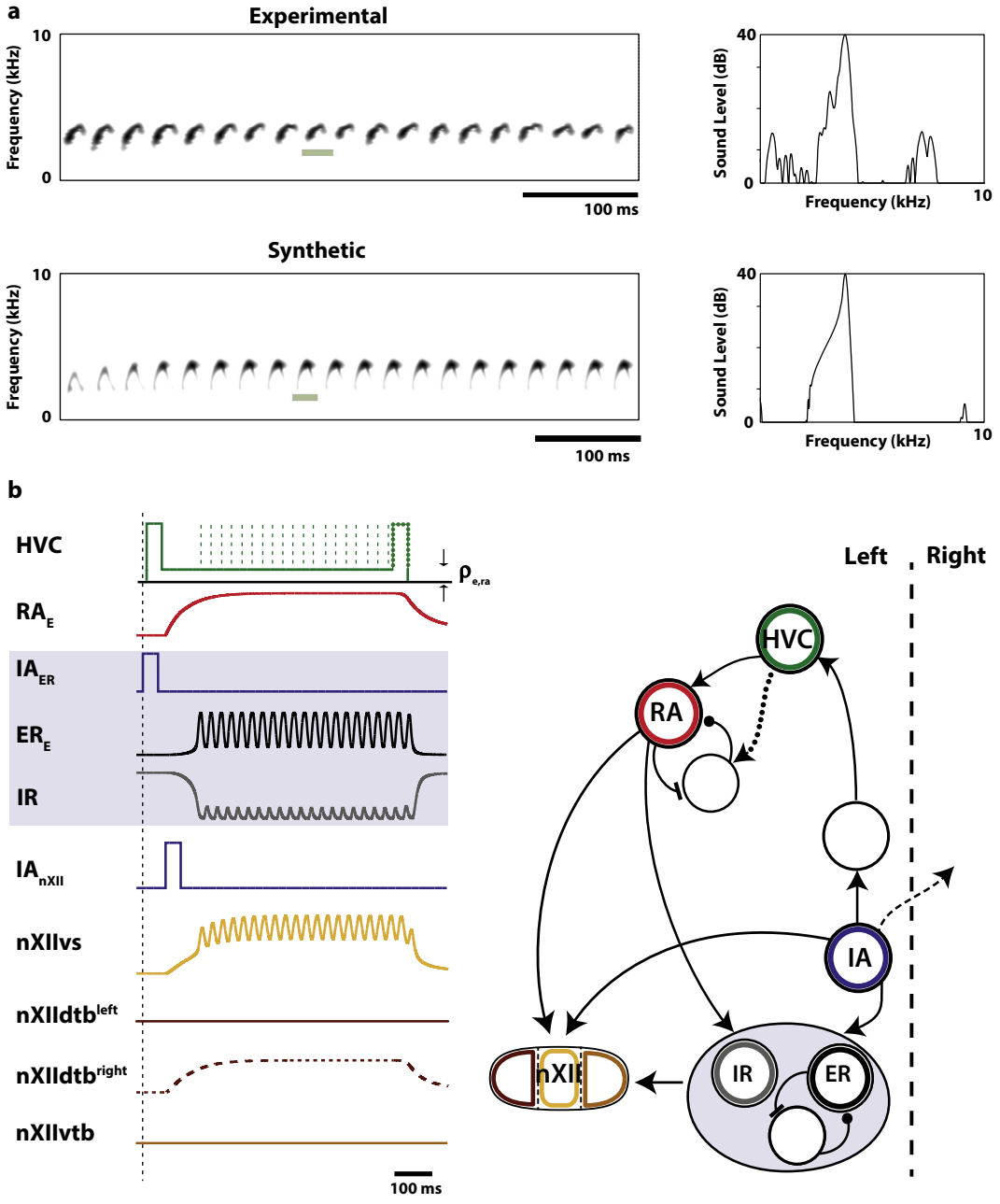

Fig. 6. Simulations for the pulsatile solutions. The pulsatile solutions are syllables uttered at a high syllabic rate. They are generated with only one side of the syrinx. (a) (Left) Sonogram of a representative example and its simulation, with our integrated model. (Right) Representation of spectral properties (sound level vs. sound frequency) of synthetic and experimental syllables marked with the green line. (Sound Analysis Pro Similarity Score: 99.68% Similarity, 88.66% Accuracy, Mean Values - Symmetric comparison). The neural architecture and the activities for the different nuclei are displayed in (b). The bilateral structure of the song system takes part in the simulations through the gating: the model reproduces the gating that allows unilateral phonation. In HVC the constant term that is needed in the simulations (reported as a constant input to RA in the equations) is represented by ρ_{era}, and we show periodic activity that might or might not be present with dashed lines. The model reproduces the solutions in both cases. When this periodic activity is present, it allows us to adjust the amplitude of the solutions, as well as the phase difference between pressure and tension. Green dotted lines represent inhibitory HVC activity, which corresponds to a direct excitation of the inhibitory population of RA.

symmetric, as there are left and right nuclei in the bird's brain. For the solutions reported here, left and right activities could be either synchronized for all the nuclei or not, with the exception of the activities of left and right *nXII* in charge of gating. For this reason, they are explicitly labeled either right or left ($e^{l,r}_{nXIIdtb}$).

2.3. Mathematical model for the syrinx

The syrinx is a phonating device located at the junction between the trachea and the bronchi. Its configuration can be described by different time dependent quantities representing the activity of different muscles. Studies on one particular species (brown thrashers, *Toxostoma rufum*) indicate that the ventral tracheobronchialis muscle (vTB) controls the active separation of the tissues that oscillate during phonation, while the dorsal tracheobronchialis muscle (dTB) controls the active closing of the tissues. A large pair of muscles called syringealis ventralis (vS) is assumed to control the fundamental frequency of the vocalizations by changing the syringeal configuration in such a way that the labia get stretched when the muscles are active.

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

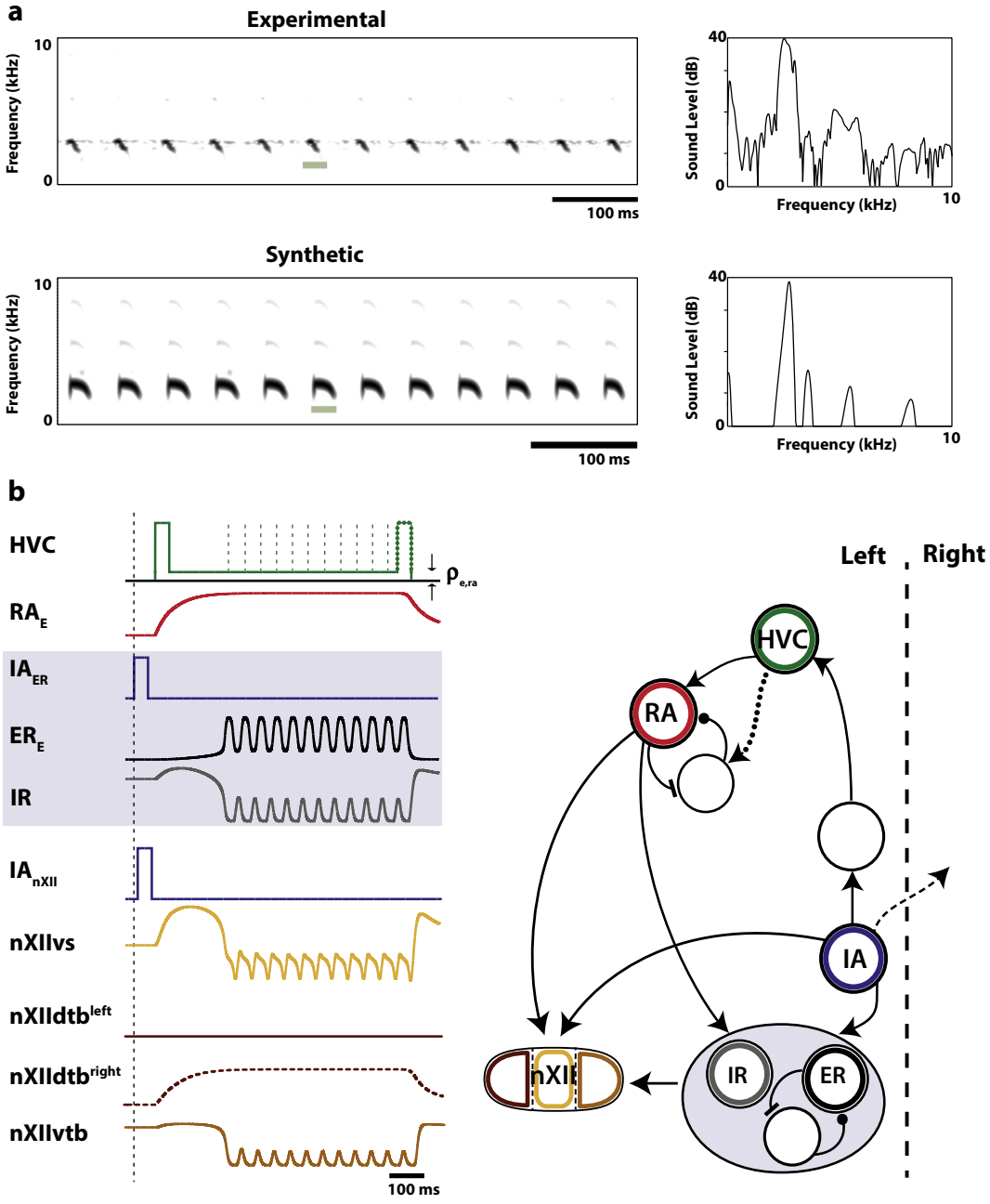

Fig. 7. Simulating the P1 solutions. As in the case of the pulsatile solutions, P1 syllables are generated with one side of the syrinx only. (a) (Left) Sonogram of a representative set of syllables and its simulation. (Right) Representation of spectral properties (sound level vs. sound frequency) of synthetic and experimental syllables marked with the green line. (Sound Analysis Pro Similarity Score: 95.45% Similarity, 70.96% Accuracy, Mean Values - Symmetric comparison). The neural architecture and the activities according to our model are shown in (b). HVC activity is, as in the previous cases, modeled as a constant activity level (ρ_{era}), plus bursts elicited by the arrival of a pulse from the brainstem, and the dashed lines represent additional periodic activity that might or might not be present. The model reproduced the periodic fluctuations of pressure in both cases. Green dotted lines represent inhibitory HVC activity, which corresponds to direct excitation of the inhibitory population of RA. This is necessary to stop the P1 gesture. The other time series correspond to the simulations of the model, for the appropriate parameters (see text).

The central nervous system (CNS), specifically area *nXIIvs*, is responsible for a slow change in the syringeal configuration (slow compared to the labial movement) that affects the tension of the oscillating labia located within the syrinx. This time dependent parameter modulates the dynamics of the fast variables, which are the positions of the labia (delicate tissues that actually modulate the airflow in order to produce the sound). The dynamics of these fast variables can be studied by writing Newton's equations for the tissues, which read as:

$$\frac{dx}{dt} = y$$

$$\frac{dy}{dt} = \gamma P(t)y - \gamma x^2 y - \gamma^2 T(t)(x + cx^3) + \gamma^2 G^{l,r}$$

where $P(t)$ stands for the pressure of the air sacs, responsible for the energy transfer from the airflow to the labia, $T(t)$ stands for the labial tension, and $G^{l,r}(t)$ stands for the gating: a force term independent of labial position or labial velocity which can interrupt the phonation by actively pushing the labia away from the phonating configuration. The time dependent parameters are the output of the neuronal model. In this way:

$$P(t) = e_{er}$$

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

$$T(t) = e_{nXIIvs}$$

$$G^{l,r}(t) = e^{l,r}_{nXIIdtb} - e_{nXIIvtb}$$

The syrinx is a bi-partite structure. The left and the right sides of the syrinx each have one pair of opposing labia that can be set in motion and controlled independently. In the case of the canary, the high frequency vocalizations are generated with the right side (fundamental frequencies higher than 3.5 kHz, approximately), and the low frequency vocalizations are generated with the left side (Suthers et al., 2004). The equations above describe the labial dynamics of one side. It is assumed that both labia on each side will have opposed, synchronized dynamics. Therefore, for each side of the syrinx, only one variable is used for describing the modulations of the airflow through that side of the syrinx. In our simulations, both sound sources are active in principle: which side of the syrinx is actually generating sound depends on the values of the gating. In this way, we not only intend to test the hypothesis that *nXIIvs*, with all its constraints, is capable of generating the correct frequency modulations, but that, *nXIIdtb* left and right and *nXIIvtb*, with all their constraints, are also capable of generating the necessary gating activity. By doing so, we are capable of generating both unilateral and bilateral vocalizations depending on the gating pattern that emerges from our equations.

3. Results

The integrated model for the song system presented in the previous section was successful in generating time traces that were good approximations of the respiratory patterns used during song by domestic canaries (*Serinus canaria*). Since it is an integrated model, the parameters that lead to a given solution for the variable $e_{er,}$ lead to specific time traces for the other nuclei in the system. Therefore, if we include the area *nXIIvs* in our integrated model, the dynamics of the variable accounting for its average activity is severely constrained. The reason is that *nXIIvs* receives inputs from RA, ER and IR, whose dynamics are determined by the integrated model. Moreover, *nXIIvs* is additionally constrained by a pulse from the IA (putatively DM), modeled as a 10 ms delayed pulse after the IA starts the syllable generation.

3.1. The P0 and P2 solutions

We tested the hypothesis that by adjusting the parameters of $\alpha_{nXIIvs,FnXII}$, $\alpha_{nXIIvs,era}$, $\alpha_{nXII,eer,}$ and $\alpha_{nXII,ier,}$ we could integrate the model so that e_{nXIIvs}, together with e_{er}, were capable of driving the syringeal model to synthesize syllables generated with the chosen respiratory patterns. For example, in the case of the long expiratory pressure pulses that we call P0 solutions (Alliende et al., 2010; Alonso et al., 2015), a typical syllable is an almost constant frequency sound preceded by a very brief sound whose frequency is very rapidly modulated (see Fig. 4a). We started the procedure by choosing the parameter values of the integrated model of the song system necessary to generate pressure patterns as in (Alonso et al., 2015), for each of the four qualitatively different pressure patterns identified in (Alliende et al., 2010). A representative syllable associated with each of the patterns was selected and then ($\alpha_{nXIIvs,FnXII}$, $\alpha_{nXIIvs,era}$, $\alpha_{nXIIvs,eer}$, $\alpha_{nXIIvs,ier}$) were varied between (0,0,0,0) and (10,10,10,−10). For each set of $\alpha_{nXIIvs,j}$, we integrated the model for the song system and used the resulting time traces for (e_{er}, $e_{nXIIvs,}$ $e_{nXIIdtb}$, $e_{nXIIvtb}$) to synthesize song. The synthetic and the experimentally recorded sounds were first qualitatively compared: a simulation was deemed to be successful if the value and temporal modulations of the fundamental frequencies of both sounds were similar. Then, we performed a quantitative comparison using Sound Analysis Pro software (Tchernichovski et al., 2000), and the results are presented in the caption of Fig. 4.

In Fig. 4a we show the sonogram of a sound file in which a set of three syllables was recorded. The syllabic rate is very low (about 2.85 Hz) and corresponds to respiratory patterns which we classify as P0 solutions (Alliende et al., 2010). Each syllable begins with a brief sound of high fundamental frequency, followed by a long whistle of low fundamental frequency. The high pitch sound is generated with the right side of the syrinx, while the rest of the sound is generated with the left side (Suthers et al., 2004).

In our model, the patterns of activity of the expiratory related area ultimately responsible for the respiratory gesture start with a pulse in the initiating area of the brainstem, which projects directly to the expiratory related nucleus. This first pulse is responsible for the initial fluctuation of the pressure (see Fig. 4b, ER_E, black trace). The initiating area also projects upwards through the thalamus (Uva) to the telencephalon. Therefore, the same pulse is responsible for activity in HVC and RA at later times. RA projects to the expiratory related area of the respiratory system. In this way, the respiratory pattern consists of a fast fluctuation and a slow modulation, the latter being the result of the telencephalic processing of the pulse generated in the IA.

The same pulse in the initiating area is responsible for the activity patterns in other nuclei of the song system such as the ones that affect the biophysical parameters controlling the syrinx. The variable e_{nXIIvs} is assumed in this model to take the same values on the left and right sides, and is associated to the tension of the labia ($T(t)$). The values of $e^{l,r}_{nXIIdtb}$ and $e_{nXIIvtb}$ correspond to the patterns of activity responsible for gating at the syrinx. Here, the left and right $e_{nXIIdtb}$ patterns must be different in order produce bilateral syllables.

We generated the time traces used to synthesize the sound displayed in Fig. 4a as follows. The activity e_{nXIIvs} is the solution of an additive model driven by a direct input from the initiating area IA as well as inputs from the excitatory population of neurons in RA and of excitatory neurons in the expiratory and inspiratory related area (Sturdy et al., 2003). The time traces in our simulations are displayed in Fig. 4b. The patterns of activity related to gating were generated in two ways: **1.** In the case of *nXIIdtb* the initial pulse in the initiating area activates the population $e^{l}_{nXIIdtb}$. The same pulse elicits activity on the contralateral side through the thalamus, and after the processing in RA, drives the activity pattern in the right hemisphere. In this way, the right side is free to produce the brief high frequency sound (while the left side is shut down by the adduction of the phonating membranes). In the same way, in the second part of the syllable the brief sound is silenced when the long gating gesture is activated on the right side ($e^{r}_{nXIIdtb}$), and the left side is free to phonate. **2.** In the case of *nXIIvtb,* the population $e_{nXIIvtb}$ receives inputs from the excitatory population of RA and the excitatory population of the inspiratory related area (IR), which leads to activation patters synchronized with inspiration. These activation patterns correspond to active abduction of the labia. The time traces obtained in our model are shown in Fig. 4b. Note that this detailed description corresponds to the solutions provided by the model after a pulse is set in the IA.

The generation of the appropriate respiratory pattern dramatically constrains what the populations of *nXII* can display as solutions. The activity of this nucleus, which consists of excitatory neurons, is severely constrained by all its afferents, and all we can fit is the relative weight of those driving functions. Even with these restrictions, here we show that the patterns capable of modulating the fundamental frequency and producing the correct gating are possible to achieve within the framework of the model. The list of parameters necessary to generate these solutions is the following:

$$(\rho_{e,ra}, \alpha_{era,Fd}, \alpha_{era,era}, \alpha_{era,ira}) = (-3.4, 5, 6, -3)$$

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

$(\rho_{i,ra}, \alpha_{ira,Fd2}, \alpha_{ira,era}, \alpha_{ira,ira}) = (-7, 0, 6, 3)$

$(\rho_{e,er}, \alpha_{eer,ra}, \alpha_{eer,F}, \alpha_{eer,eer}, \alpha_{eer,\mathrm{ier}}) = (-7.45, 10, 1, 10, -1.1)$

$(\rho_{i,er}, \alpha_{ier,\mathrm{ra}}, \alpha_{ier,\mathrm{eer}}, \alpha_{ier,\mathrm{ier}}) = (-11.5, 0, 10, 2)$

$(\rho_{e,ir}, \alpha_{eir,\mathrm{ra}}, \alpha_{eir,\mathrm{eer}}) = (0, 10, -10)$

$(\rho_{e,nXII}, \alpha_{nXIIvs,FnXII}, \alpha_{\mathrm{nXIIvs,era}}, \alpha_{\mathrm{nXIIvs,eer}}, \alpha_{\mathrm{nXIIvs,eir}}) = (-3, 1, 0.5, 0.5, 0)$

$(\rho_{e,nXII}, \alpha^{r}_{nXIIdtb,FnXII}, \alpha^{r}_{nXIIdtb,era}) = (-3, 0, 10)$

$(\rho_{e,nXII}, \alpha^{l}_{nXIIdtb,FnXII}, \alpha^{l}_{nXIIdtb,era}) = (-3, 10, 0)$

$(\rho_{e,nXII}, \alpha_{enXIIvtb,\mathrm{eir}}) = (-3, 10)$

Nonlinear systems display qualitatively different solutions within open regions of the parameter space. Therefore, our fitting is qualitative in nature. We report simply one value within an open region of the parameter space at which our model displays a solution with the features that we aim to reproduce (see Section 3.4 for a discussion on the robustness of our results).

What we call P2 solutions are generated at syllabic rates between 5 and 12 Hz (Alliende et al., 2010). Although their phonation does not necessarily involve bilateral patterns, according to our model, the pressure gestures require a direct drive from the initiating area as well as a driving processed by the telencephalon. Besides the syllabic rate, a P2 solution differs form a P0 one in that it does not contain a long whistle of constant frequency. An example of P2 solution is displayed in Fig. 5. The mechanism, however, is very similar to what we describe in the previous subsection. Therefore, our description is limited to the parameters used in our model to synthesize the solutions.

$(\rho_{e,ra}, \alpha_{era,Fd}, \alpha_{era,era}, \alpha_{era,ira}) = (-3.83, 1, 5, -10)$

$(\rho_{i,ra}, \alpha_{ira,Fd2}, \alpha_{ira,era}, \alpha_{ira,ira}) = (-7, 0, 10, 8)$

$(\rho_{e,er}, \alpha_{eer,ra}, \alpha_{eer,F}, \alpha_{eer,eer}, \alpha_{eer,\mathrm{ier}}) = (-7.45, 9, 1, 9.85, -5.25)$

$(\rho_{i,er}, \alpha_{ier,\mathrm{ra}}, \alpha_{ier,\mathrm{eer}}, \alpha_{ier,\mathrm{ier}}) = (-11.5, 9.45, 9, 1.6)$

$(\rho_{e,ir}, \alpha_{eir,\mathrm{ra}}, \alpha_{eir,\mathrm{eer}}) = (-2, 10, -10)$

$(\rho_{e,nXII}, \alpha_{nXIIvs,FnXII}, \alpha_{\mathrm{nXIIvs,era}}, \alpha_{\mathrm{nXIIvs,eer}}, \alpha_{\mathrm{nXIIvs,eir}}) = (-3, 1, 3, 3, 0)$

$(\rho_{e,nXII}, \alpha^{r}_{nXIIdtb,FnXII}, \alpha^{r}_{nXIIdtb,era}) = (-3, 0, 10)$

$(\rho_{e,nXII}, \alpha^{l}_{nXIIdtb,FnXII}, \alpha^{l}_{nXIIdtb,era}) = (-3, 10, 0)$

$(\rho_{e,nXII}, \alpha_{enXIIvtb,\mathrm{eir}}) = (-3, 10)$

As before, a quantitative comparison between the recorded syllables and their synthetic counterparts was performed using Sound Analysis Pro. The results are presented in Fig. 5.

3.2. Pulsatile and P1 solutions

Pulsatile solutions are produced at syllabic rates higher than 25 Hz, whereas P1 solutions are generated at syllabic rates between 13 Hz and 25 Hz. In both cases, only one side participates in the phonation. The main difference between these solutions is that in pulsatile solutions, the pressure fluctuations are mounted on a constant value (unique expiratory pulse), while P1 solutions present mini-breaths between sounds (one expiratory pulse per note). In both cases, the IA affects the respiratory pattern after processing through the telencephalon and the direct input is not necessary. Basically, the input through Uva to HVC leads to an activation pattern in RA, which, in turn, leads to oscillations in the expiratory related area. If additional oscillations are elicited by internal dynamics in HVC, the solutions obtained at the expiratory related area are similar (see Fig. 6b, green dashed lines).

It is interesting to note that the activity in RA involves a slow pattern of high activity for both cases. In our model, the population of neurons whose activity is e_{nXIIvs} receives inputs from RA, ER and IR. Alternatively, the population of neurons whose activity is $e^{l,r}_{nXIIdtb}$ receives inputs only from RA. In Fig. 6 we display an example of a pulsatile solution produced by the left side. Assuming a symmetrical configuration in which the pulse of activity from the IA generates a similar RA activity in both hemispheres, the highly activated RA population drives the right side into silence through the activation of the gating $e^{r}_{nXIIdtb}$. In other words, a similar activity in RA is used to generate a vocalization with an oscillatory regime in ER and, contralaterally, an active closing through *nXIIdtb*. The parameters used in this simulation are:

$(\rho_{e,ra}, \alpha_{era,Fd}, \alpha_{era,era}, \alpha_{era,ira}) = (-5.25, 5, 10, -10)$

$(\rho_{i,ra}, \alpha_{ira,Fd2}, \alpha_{ira,era}, \alpha_{ira,ira}) = (-12, 5, 10, 2)$

$(\rho_{e,er}, \alpha_{eer,ra}, \alpha_{eer,F}, \alpha_{eer,eer}, \alpha_{eer,\mathrm{ier}}) = (-7.5, 6, 0, 10, -6.2)$

$(\rho_{i,er}, \alpha_{ier,\mathrm{ra}}, \alpha_{ier,\mathrm{eer}}, \alpha_{ier,\mathrm{ier}}) = (-11.5, 6, 10, 2)$

$(\rho_{e,ir}, \alpha_{eir,\mathrm{ra}}, \alpha_{eir,\mathrm{eer}}) = (0, 0, -10)$

$(\rho_{e,nXII}, \alpha_{nXIIvs,FnXII}, \alpha_{\mathrm{nXIIvs,era}}, \alpha_{\mathrm{nXIIvs,eer}}, \alpha_{\mathrm{nXIIvs,eir}}) = (-3, 0, 1.5, 1.1, 0)$

$(\rho_{e,nXII}, \alpha^{r}_{nXIIdtb,FnXII}, \alpha^{r}_{nXIIdtb,era}) = (-3, 0, 1)$

$(\rho_{e,nXII}, \alpha^{l}_{nXIIdtb,FnXII}, \alpha^{l}_{nXIIdtb,era}) = (-3, 1, 0)$

$(\rho_{e,nXII}, \alpha_{enXIIvtb,\mathrm{eir}}) = (-3, 10)$

The P1 solutions are generated with an analogous mechanism. The parameters used to generate the example displayed in Fig. 7 read as:

$(\rho_{e,ra}, \alpha_{era,Fd}, \alpha_{era,era}, \alpha_{era,ira}) = (-3.5, 5, 10, -10)$

$(\rho_{i,ra}, \alpha_{ira,Fd2}, \alpha_{ira,era}, \alpha_{ira,ira}) = (-12, 5, 10, 2)$

$(\rho_{e,er}, \alpha_{eer,ra}, \alpha_{eer,F}, \alpha_{eer,eer}, \alpha_{eer,\mathrm{ier}}) = (-7.55, 4.5, 0, 10, -4.5)$

$(\rho_{i,er}, \alpha_{ier,\mathrm{ra}}, \alpha_{ier,\mathrm{eer}}, \alpha_{ier,\mathrm{ier}}) = (-11.5, 4.5, 10, 2)$

$(\rho_{e,ir}, \alpha_{eir,\mathrm{ra}}, \alpha_{eir,\mathrm{eer}}) = (0, 1, -10)$

$(\rho_{e,nXII}, \alpha_{nXIIvs,FnXII}, \alpha_{\mathrm{nXIIvs,era}}, \alpha_{\mathrm{nXIIvs,eer}}, \alpha_{\mathrm{nXIIvs,eir}}) = (-3, 0, 1, 1.7, 6)$

$(\rho_{e,nXII}, \alpha^{r}_{nXIIdtb,FnXII}, \alpha^{r}_{nXIIdtb,era}) = (-3, 0, 1)$

$(\rho_{e,nXII}, \alpha^{l}_{nXIIdtb,FnXII}, \alpha^{l}_{nXIIdtb,era}) = (-3, 1, 0)$

$(\rho_{e,nXII}, \alpha_{enXIIvtb,\mathrm{eir}}) = (-3, 10)$

In this work, we show one example of a downsweep and another one of an upsweep to illustrate that this is not a challenge for this model.

Once we fitted the mathematical implementation of the neural model, we used the time dependent parameters corresponding to e_{er}, e_{nXIIvs}, $e^{l,r}_{nXIIds}$ and $e_{nXIIvtb}$ to feed a biophysical model of the syrinx

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

Table 1
Summary of the scaling factors needed for the neural instructions to drive the biophysical model of the syrinx.

	P0	P2	P1	PULSATILE
P(t)	$e_{er} * 2$	$e_{er} * 2$	$e_{er} * 2 - 0.85$	$e_{er} - 0.25$
$T^l(t)$	$e_{nXIIvs} * 29$	$e_{nXIIvs} + 1.5$	$e_{nXIIvs} * 1.7 + 1.5$	$e_{nXIIvs} * 30 - 9.7$
$G^l(t)$	$(e^l_{nXIIdtb} * 40) - (e_{nXIIvtb} * 40)$	$(e^l_{nXIIdtb} * 15 - 2) - (e_{nXIIvtb} * 25 - 3)$	$(e^l_{nXIIdtb} * 1) - (e_{nXIIvtb} * 5 - 0.1)$	$e^l_{nXIIdtb} * 1$
$T^r(t)$	$e_{nXIIvs} * 28.5$	$e_{nXIIvs} * 3 + 0.7$	$e_{nXIIvs} * 1 + 3$	$e_{nXIIvs} * 30$
$G^r(t)$	$e^r_{nXIIdtb} * 20 + 7$	$(e^r_{nXIIdtb} * 30 + 2.7) - (e_{nXIIvtb} * 5)$	$e^r_{nXIIdtb} * 10 + 3$	$e^r_{nXIIds} * 30$

in order to generate synthetic syllables. By nature of our model, in which the nonlinearity is provided by a sigmoidal function, solutions are normalized to unity. Therefore, the neural activity must be scaled so as to translate it into the physiological instructions that drive the syrinx. For each of the different types of syllables mentioned above the tension, gating and pressure time traces $(T^{l,r}(t), G^{l,r}(t), P(t))$ were scaled during the phonating intervals in order to generate syllables with similar frequency modulations to those of the experimental examples. Note that tension gestures differ between the left and right sides of the syrinx so that the correct frequency modulations for bilateral syllables can be correctly fitted. The parameter γ was set to 9000 for all the syllables. Table 1 summarizes the scaling factors:

For both the pulsatile and P1 syllable types, we compared the original sounds with their synthetic counterparts using Sound Analysis Pro, and the results are shown in the respective figure captions.

3.3. Bilateral instructions

The mechanism underlying the generation of the pressure patterns allows us to classify them in two groups. The first group includes the P0 and P2 solutions given that, in those cases, the expiratory related area receives both a direct input from the brainstem and a second input after the initial activation is processed by the telencephalon. The other group encompasses the pulsatile and P1 solutions needing only the second input described above.

For the solutions of the first group, part of the motor gestures needed to synthesize a syllable has to undergo a bottom up journey to the telencephalon through the thalamic nucleus Uva. In the cases in which both sides of the syrinx participate sequentially in the generation of the syllable, non-synchronized bilateral patterns of activity are necessary, at least for the generation of the instructions controlling gating. One way to achieve sequential non-symmetric activation is by allowing the contralateral but not the ipsilateral bottom up journey of the neural instructions generated in the IA. The pattern of activity in charge of controlling the labial tension might or might not present both contralateral and ipsilateral connections in the bottom up journey of the neural instructions from the IA. In Fig. 8, we show how our model generates the bilateral patterns of gating for the case of the P2 solutions.

The pulsatile and P1 solutions use only one side. Notice that we left the possibility open for internal dynamics in HVC in the form of periodic activity that reinforces the amplitudes of the oscillations at the ER area. In this case, the pattern of HVC activity needs not be synchronized in the two hemispheres. However, for these solutions the model requires the bottom up signal to silence the contralateral side. Since this signal is transmitted both ipsilaterally and contralaterally, we predict activity at the transitions between different syllable types in both hemispheres (see Figs. 6b and 7b).

3.4. Robustness of the solutions representing motor gestures

A typical feature of nonlinear systems is that qualitatively similar solutions exist for open sets of parameters. The boundaries between these regions are curves in parameter space called

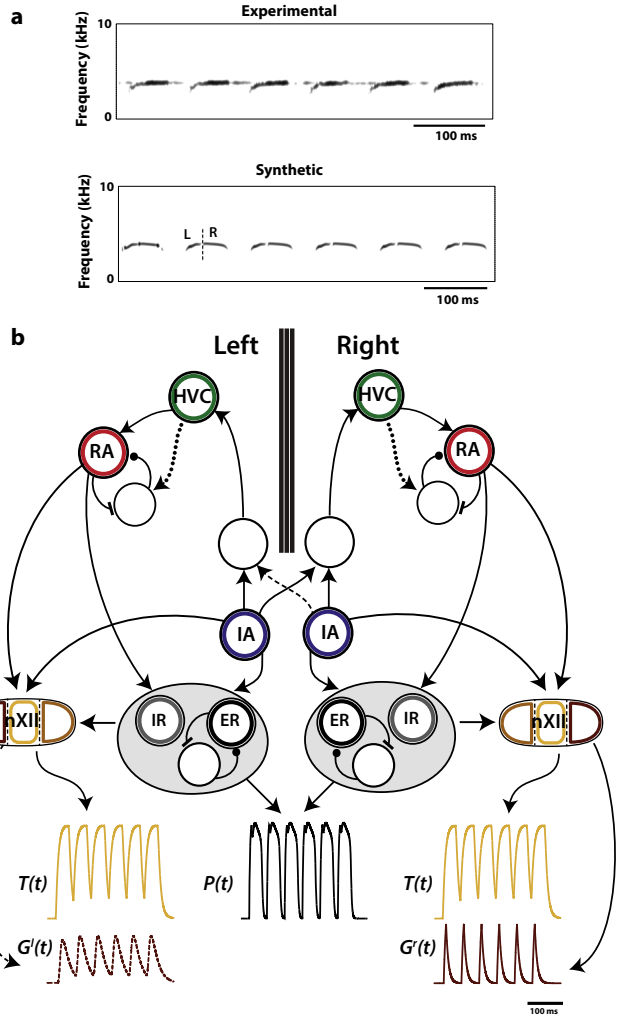

Fig. 8. The bilateral structure of the song system and its functionality. The P0 and P2 syllables are generated with both sides of the syrinx. We display the neural architecture and the time traces used to simulate a P2 solution. The sonograms of a recorded syllable and its simulation are shown in (a). The time traces for tension and gating are displayed below the corresponding nuclei. Note that the time trace that generates the modulation of the tension on one side, as it is connected contralaterally to a region of *nXII* in control of gating, silences the other side.

bifurcations. Taking advantage of this characteristic, our nonlinear model is able to gracefully reproduce the four typical pressure patterns found in canaries without the need to find a singular point in a high dimensional parameter space, which would be unintelligible in a biological sense. This was one of the keystones in the construction of the original circular model (Alonso et al., 2015).

In this work we show that, in spite of all the restrictions imposed by the architecture and the functional requirements needed to generate the pressure patterns, it is possible to obtain tension gestures capable of driving the synthesis of songs with the right frequency modulation. We also explore whether these

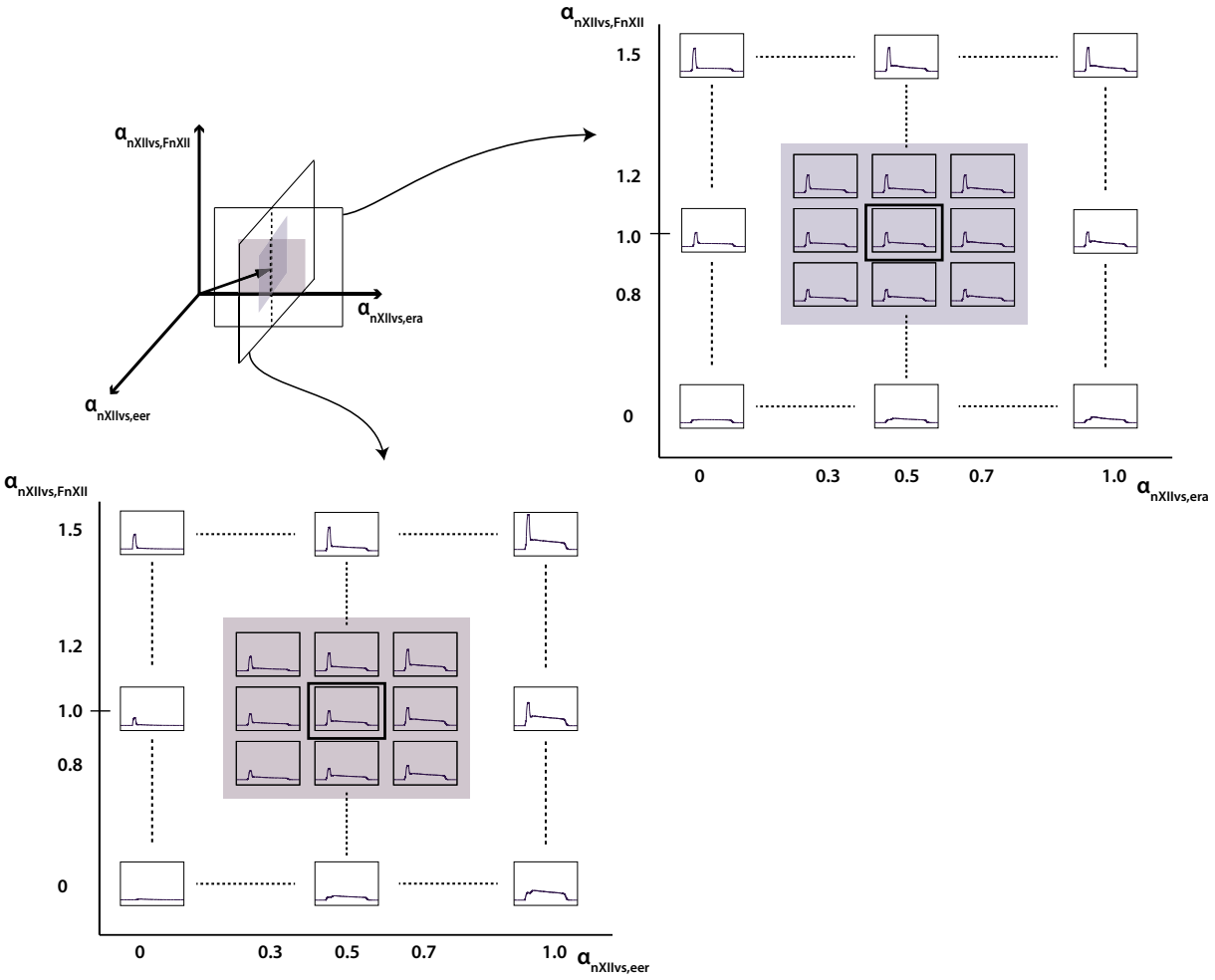

Fig. 9. Stability of the solutions of *nXIIvs*. This figure illustrates the robustness of the solutions representing motor gestures. In the top left panel, we display the 3-dimensional parameter space that corresponds to the inputs from RA ($\alpha_{nXIIvs,era}$), *FnXII* ($\alpha_{nXIIvs,FnXII}$) and ER($\alpha_{nXIIvs,eer}$) to the excitatory population of *nXII*vs, for the P0 syllable type displayed in Fig. 4. The top-right and bottom-left panels show an exploration of a wide range of the parameter space, centered on the fitted gesture for *nXIIvs* (see Fig. 4b). Note that the gestures obtained with the model vary smoothly within a wide neighborhood of the fitted values of $\alpha_{nXIIvs,era}$, $\alpha_{nXIIvs,FnXII}$ and $\alpha_{nXIIvs,eer}$ (shaded areas). This shows that small variations of the inputs to this nucleus give rise to *vs* gestures with the same qualitative properties (i.e. a sharp peak at the onset followed by a slow decaying modulation), whereas departing from this region would give rise to qualitative different gestures.

gestures are robust under parameter variations, testing this hypothesis numerically. In Fig. 9 we show how the tension used to generate a P0 syllable is deformed as the parameters that control the influence of the nuclei *nXIIvs* vary. Note that, even though the gesture changes, its qualitative shape are preserved. In particular, its key features, a brief pulse followed by a larger smoothly decaying one, remain unaltered.

4. Discussion

In this work we presented an integrated model for song production in *Serinus canaria.* The model is an operational one: a set of ordinary differential equations whose variables are the activities of different areas of the song system. Some of them are clearly associated with neural nuclei (HVC, RA and Uva), while others are defined with less precision: the "expiratory related area" (ER) can be interpreted as RAm, the inspiratory related area (IR) can be interpreted as PAm, while the "initiating area" in the brainstem (IA) can be interpreted as DM. This model is also minimal in the sense that only a reduced set of neural nuclei is used. In the song system there are other nuclei participating in loops that include HVC (e.g. DMP, MMAN, Nif) but these were not considered in order to obtain the minimal neural architecture required to generate canary song.

We have shown that this integrated model is capable of generating the respiratory patterns and sounds observed in *Serinus canaria* song. In particular, we tested the hypothesis that, even considering all the restrictions imposed by all the other areas, the activity of the brain region responsible for controlling the syrinx is capable of generating realistic instructions. We understand by "realistic instructions" sets of time dependent parameters that can drive a model of the avian vocal organ to synthesize realistic songs. Here we worked with song examples corresponding to the four classes of air sac pressure patterns of the *Serinus canaria* song, and found the value of the parameters in the model needed to reproduce them. Then, we explored the gestures obtained when the weights of the connections between the different areas of the song system and the *nXII* were varied, and for each set of parameters, we synthesized sounds. The result of these numerical simulations is that we were always able to find connection strengths such that the final synthetic sounds were the proper ones. In this way, we are able to generate the entire acoustic repertoire described for the *Serinus canaria* song.

One interesting aspect of the model presented here is that the same time traces needed to modulate the fundamental frequency of the vocalizations when injected as $T(t)$ in the syrinx model were capable of generating the appropriate gating pattern when injected as $G(t)$. In other words, similar time traces at T and G have

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

dramatically different effects in the acoustics. Interestingly enough, the outcome is coherent: fitting an observed frequency modulation generates a gating pattern compatible with its production. This result is particularly remarkable, since it was not obvious that the patterns that modulate the fundamental frequencies were even possible to achieve. In order to generate a given pressure pattern, the activities of all the areas are constrained, and the area responsible for the labial tension (part of *nXII*) is assumed to follow its drivers passively. In this model we show not only that the appropriate tension pattern can be generated for a given pressure pattern, but that the required gating pattern is obtained as well.

The presence of a continuous representation of time in HVC has been discussed in the literature. This would be the result of a sequence of brief bursts active sequentially during song production (Lynch et al., 2016; Hahnloser et al., 2002). In terms of average population activity, this would correspond to a continuous time trace forcing the activity of RA. In our model there are continuous components of HVC activity, with different values for different syllables. This is mathematically expressed in terms of the different values of constant inputs ρ arriving to RA, which are necessary for the generation of the different syllables ($\rho_{e,ra}$, $\rho_{i,ra}$). In this aspect, the model presented here incorporates the need of a continuous component of HVC activity. Additionally, as suggested in recent work (Amador et al., 2013), the feedback from the brainstem incorporates a component of heterogeneity which correlates with the song. In other words, there is motor information in HVC, and not just a representation of time.

Furthermore, the model requires inputs to HVC from the brainstem, which arrive through Uva, and provide a *syllable-type* dependent component. This leads to specific predictions. For example, peaks of HVC activity close to the beginning of the syllables for the P0 solutions, or peaks of HVC activity at the transitions between pulsatile and P1 solutions, or close to the beginning of the pulsatile solutions. Both P1 and pulsatile solutions can be generated either with additional periodic activity in HVC emerging from its internal dynamics or without it. Based on this, the model for HVC presented here includes a combination of a continuous nonspecific activity and sparse peaks of activity that occur at specific times related to the song (Amador et al., 2013).

Another aspect of birdsong production that is being currently discussed is whether the telencephalon acts in a top-down fashion or not. In a top-down paradigm, HVC controls the timing of the system downstream. Evidence supporting this view was presented in work that described the effect of cooling HVC in zebra finches (Long and Fee, 2008). Since the syllables stretched as HVC temperature was decreased, it was suggested that HVC was acting as a time controller running slower. Recent work with canaries showed an initial stretching of syllables when cooling HVC but an effect of "breaking" if the temperature was dropped below a critical point. See Goldin et al. (2013) for further details. The breaking of some syllables can be interpreted by considering an additional effect beyond stretching: the slowing down of the synaptic inputs into HVC. This suggests that a more integrated architecture is needed in order to reproduce how temporal features of the song are affected by temperature manipulations within a large range. Consistent with this view, it was shown that temperature manipulations in other parts of the song system affect the song timing in finches (Hamaguchi et al., 2016), which suggests the need for a distributed network to generate behavior.

The basic idea of this model is that a group of canary syllables (which we call P2 and P0 solutions) are generated when the nuclei in the brainstem (both respiratory related and *nXII*) receive two inputs: one directly from the brainstem (IA), and a second one which originates in the same area, but is processed through the song system in a bottom up journey that enriches the instructions in the telencephalon before returning to the brainstem. Since the bottom up part of the process allows contralateral activations, well-coordinated bilateral motor patterns emerge naturally in this framework. A second group of syllables (P1 and pulsatile solutions) is generated with the second part of the mechanism alone.

Premotor activity within HVC is synchronized between hemispheres during song production (Schmidt, 2003) despite the absence of commissural connections between these two nuclei or any other forebrain song control nuclei (Fig. 1). This observation suggests that the bilaterally projecting brainstem respiratory centers may provide a synchronizing input to the forebrain song system. Schmidt and colleagues (Ashmore et al., 2008) show that activation of Pam can drive neural activity bilaterally in HVC and in RA, and that this activation is abolished by lesioning Uva. Moreover, microstimulations delivered to Pam, but not *nXII*ts, during singing causes disruptions of song sequencing (Ashmore et al., 2005). Altogether, these results suggest that Pam can influence song pattern generation, possibly through its ascending thalamic projections, as it has been shown that Uva is critical for interhemispheric coordination and song production (either unilateral or bilateral damage to Uva immediately and severely impairs singing in adult birds) (Coleman and Vu, 2005; Williams and Vicario, 1993). Here we show a functional example of how these neural mechanisms can be put together to generate birdsong.

The model we present here was designed to generate pressure patterns with the morphology of the measured air sac pressure time traces reported in the literature. A basic neural structure (an excitatory population coupled with an inhibitory one) driven by simple time traces can generate the diversity of observed signals. Despite the generalized notion that, given enough free parameters, any solution can be obtained from a dynamical system, this is not true. In fact, dynamical systems can be classified by the topological structure of their solutions (Mindlin et al., 1990). In previous work, we showed that the rest of the song system could generate the simple driving parameters necessary to reproduce the pressure patterns (Alonso et al., 2015). This experimental set of time traces have specific shapes, and our model was built so that it can generate them, as well as reproduce the morphological changes observed under cooling (Goldin et al., 2013).

The model also makes a series of specific predictions for neural recordings. One of them is the existence of synchronized activity in HVC. In particular: **1.** The existence in both hemispheres of synchronized brief bursts at the transition between pulsatile solutions and solutions of type P1 (in addition to the eventual continuous components previously described). **2.** The existence in both hemispheres of synchronized brief bursts a few tens of milliseconds after the syllable onset for each syllable of P2 and P0 type. **3.** The existence of synchronized brief bursts in both hemispheres at the beginning of pulsatile syllables. **4.** The existence of brief bursts in Uva right before onsets (temporally very close to inspiratory activity), at least before each syllable of type P2, before the pulsatile set of syllables, and before each syllable of type P0. As we noted before, pulsatile syllables as well as syllables of type P1 can be reinforced by periodic activity in HVC (which would be necessary for stretching under cooling). Those oscillations in HVC need not be present in both hemispheres and if they are, they need not be synchronized. Another prediction of the model is a set of time traces (e_{nXIIvs}) that could be comparable to the envelopes of the activity of the muscle syringealis ventralis (vS). The rationale for associating our variable e_{nXIIvs} with the activity of vS is that despite recent reports of synergistic contributions of different motor gestures to achieve frequency modulation (Alonso et al., 2014), vS is concurrent with frequency control, especially in those species in which oscillations are born in Hopf bifurcations. Notice that syllable related non-homogeneity in HVC projecting neurons was reported initially in juveniles and remained significant in adults

Please cite this article in press as: Alonso, R.G., et al. An integrated model for motor control of song in *Serinus canaria*. J. Physiol. (2016), http://dx.doi.org/10.1016/j.jphysparis.2016.12.003

(Okubo et al., 2015). The bursts described here coexist with the continuous components previously described.

In this work we assume the same activity patterns of $e_{nXII\,vs}$ in both hemispheres. Yet there could be differences in right and left vS activities for complex syllables involving bilateral control. In that case, as in the gating gestures in our model, the activity would be the result of a direct input from the IA leading to a fast modulation of $e_{nXII\,vs}$ on one side, and a contralateral slower thalamic modulation of the $e_{nXII\,vs}$ on the other side. This would give rise to sequential patterns on both sides of the syrinx and lead to different vS activities.

In this work we show that, by adjusting only the relative weights of the inputs to *nXII*, the distributed activity patterns associated to a pressure pulse will drive *nXII* to generate the appropriate solutions in order to produce canary song. It is interesting to note that the gating patterns are intimately associated with the tension patterns: what controls the tension of one side, fed to the contralateral gating muscles, generates the minimal instructions that the syrinx requires to synthesize the appropriate song. Moreover, there is a close relationship between the pressure patterns and the tension patterns. In other words, this model depicts a scene in which essentially similar instructions, driving pressure, tension and gating muscles drive a biomechanical device to synthesize realistic canary songs.

Acknowledgements

We thank Cecilia T. Herbert for comments to this manuscript and English editing. Funding: This work was supported by CONICET, ANCyT, UBA, and NIH through R01-DC-012859 and R01-DC-006876.

References

Alliende, J.A., Mendez, J.M., Goller, F., Mindlin, G.B., 2010. Hormonal acceleration of song development illuminates motor control mechanism in canaries. Dev. Neurobiol. 70 (14), 943–960.

Alonso, L.M., Alliende, J.A., Goller, F., Mindlin, G.B., 2009. Low-dimensional dynamical model for the diversity of pressure patterns used in canary song. Phys. Rev. E 79 (4), 41929.

Alonso, R., Goller, F., Mindlin, G.B., 2014. Motor control of sound frequency in birdsong involves the interaction between air sac pressure and labial tension. Phys. Rev. E 89 (3), 032706.

Alonso, R.G., Trevisan, M.A., Amador, A., Goller, F., Mindlin, G.B., 2015. A circular model for song motor control in Serinus canaria. Front. Comput. Neurosci. 9.

Amador, A. et al., 2013. Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature 495 (7439), 59–64 <http://www.nature.com/doifinder/10.1038/nature11967> (accessed June 14, 2016).

Amador, A., Mindlin, G.B., 2014. Low dimensional dynamics in birdsong production. Euro. Phys. J. B 87 (12), 1–8.

Ashmore, R.C., Renk, J.A., Schmidt, M.F., 2008. Bottom-up activation of the vocal motor forebrain by the respiratory brainstem. J. Neurosci. 28 (10), 2613–2623.

Ashmore, R.C., Wild, J.M., Schmidt, M.F., 2005. Brainstem and forebrain contributions to the generation of learned motor behaviors for song. J. Neurosci. 25 (37), 8543–8554.

Coleman, M.J., Vu, E.T., 2005. Recovery of impaired songs following unilateral but not bilateral lesions of nucleus uvaeformis of adult zebra finches. J. Neurobiol. 63 (1), 70–89.

Elemans, C., Rasmussen, J.H., Herbst, C.T., During, D.N., Zollinger, S.A., Brumm, H., Srivastava, K., Svane, N., Ding, M., Larsen, O.N., Sober, S.M., Svec, J.G., 2015. Universal mechanisms of sound production and control in birds and mammals. Nat. Commun. 6.

Gardner, T., Cecchi, G., Magnasco, M., Laje, R., Mindlin, G.B., 2001. Simple motor gestures for birdsongs. Phys. Rev. Lett. 8720 (20), 208101.

Goldin, M.A., Alonso, L.M., Alliende, J.A., Goller, F., Mindlin, G.B., 2013. Temperature induced syllable breaking unveils nonlinearly interacting timescales in birdsong motor pathway. PLoS One 8 (6), e67814.

Goller, F., Larsen, O.N., 1997. A new mechanism of sound generation in songbirds. Proc. Natl. Acad. Sci. USA 94 (26), 14787–14791.

Goller, F., Suthers, R.A., 1995. Implications for lateralization of bird song from unilateral gating of bilateral motor patterns. Nature 373 (6509), 63–66.

Goller, F., Suthers, R.A., 1996a. Role of syringeal muscles in controlling the phonology of bird song. J. Neurophysiol. 76 (1), 287–300.

Goller, F., Suthers, R.A., 1996b. Role of syringeal muscles in gating airflow and sound production in singing brown thrashers. J. Neurophysiol. 75 (2), 867–876.

Hahnloser, R.H.R., Kozhevnikov, A.A., Fee, M.S., 2002. An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature 419 (6902), 65–70.

Hamaguchi, K., Tanaka, M., Mooney, R., 2016. A distributed recurrent network contributes to temporally precise vocalizations. Neuron 91 (3), 680–693.

Hoppensteadt, F.C., Izhikevich, E.M., 1997. Weakly Connected Neural Networks. Springer.

Lynch, G.F. et al., 2016. Rhythmic continuous-time coding in the songbird analog of vocal motor cortex. Neuron 90 (4), 877–892.

Long, M.A., Fee, M.S., 2008. Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature 456, 189–194.

Mindlin, G.B., Hou, X.-J., Solari, H.n.G., Gilmore, R., Tufillaro, N., 1990. Classification of strange attractors by integers. Phys. Rev. Lett. 64 (20), 2350.

Mindlin, G.B., Laje, R., 2005. The Physics of Birdsong. Springer Verlag, Berlin.

Nottebohm, F., Stokes, T.M., Leonard, C.M., 1976. Central control of song in the canary, Serinus canarius. J. Compar. Neurol. 165 (4), 457–486.

Okubo, T.S., Mackevicius, E.L., Payne, H.L., Lynch, G.F., Fee, M.S., 2015. Growth and splitting of neural sequences in songbird vocal development. Nature 528 (7582), 352–357.

Schmidt, M.F., 2003. Pattern of interhemispheric synchronization in HVc during singing correlates with key transitions in the song pattern. J. Neurophysiol. 90 (6), 3931–3949.

Schmidt, M.F., Wild, J.M., 2014. The respiratory-vocal system of songbirds: anatomy, physiology, and neural control. Prog. Brain Res. 212, 297–335.

Sturdy, C.B., Wild, J.M., Mooney, R., 2003. Respiratory and telencephalic modulation of vocal motor neurons in the zebra finch. J. Neurosci. 23 (3), 1072–1086.

Suthers, R.A., 1990. Contributions to birdsong from the left and right sides of the intact syrinx. Nature 347 (6292), 473–477.

Suthers, R.A., Vallet, E., Tanvez, A., Kreutzer, M., 2004. Bilateral song production in domestic canaries. J. Neurobiol. 60 (3), 381–393.

Suthers, R.A., Zollinger, S.A., 2004. Producing song: the vocal apparatus. Ann. N.Y. Acad. Sci. 1016 (1), 109–129.

Tchernichovski, O. et al., 2000. A procedure for an automated measurement of song similarity. Anim. Behav. 59. Available at: <http://www.idealibrary.com>.

Titze, I.R., 1988. The physics of small-amplitude oscillation of the vocal folds. J. Acoust. Soc. Am. 83 (4), 1536–1552.

Trevisan, M.A., Mindlin, G.B., Goller, F., 2006. Nonlinear model predicts diverse respiratory patterns of birdsong. Phys. Rev. Lett. 96 (5), 58103.

Wild, J.M., 1997. Neural pathways for the control of birdsong production. J. Neurobiol. 33, 653–670.

Wild, J.M., Li, D., Eagleton, C., 1997. Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory-vocal nuclei in the brainstem of pigeon (Columba livia) and zebra finch (Taeniopygia guttata). J. Comp. Neurol. 377, 392–413.

Wild, J.M., Williams, M., Suthers, R., 2000. Neural pathways for bilateral vocal control in songbirds. J. Comp. Neurol. 423, 413–426.

Williams, H., Vicario, D.S., 1993. Temporal patterning of song production: participation of nucleus uvaeformis of the thalamus. J. Neurobiol. 24 (7), 903–912.