

1 Leonardo Lancia and Bodo Winter

## 2 **The interaction between competition,** 3 **learning, and habituation dynamics in** 4 **speech perception**

5  
6  
7  
8 **Abstract:** Even though the outcome of the perception of phonological patterns is  
9 categorical, this process might still arise from continuous dynamics. Here, we  
10 propose a unified dynamical account of three types of behavior that are usually  
11 studied in isolation: short-term perceptual behavior, long-term perceptual habit-  
12 uation, and even longer-term perceptual learning. We develop a model and test  
13 its predictions in two speech identification tasks on an acoustic continuum be-  
14 tween the French words [sep] and [step]. When presenting stimuli sequentially  
15 from one end of the continuum to the other, we found that the presentation order  
16 systematically changed the position of the perceptual switch from one word to the  
17 other. We also found that response times were slower and more variable around  
18 this perceptual switch, regardless of its position on the acoustic continuum. And,  
19 throughout the experiment, participants became more sensitive to small acoustic  
20 differences between stimuli. Our model can account for these results and for a  
21 surprising finding, namely that the initial presentation order affected responses  
22 even late in the experiment. Overall, our results point to the importance of the  
23 relation between fast processes responsible for competition, and slow processes  
24 responsible for habituation and learning in explaining how listeners can perceive  
25 speech categorically in a way that is both flexible and robust.

26

27

28 **Leonardo Lancia:** Dept. of Linguistics, Max Planck Institute for Evolutionary Anthropology.

29 E-mail: leonardo\_lancia@eva.mpg.de

30 **Bodo Winter:** Dept. of Cognitive and Information Sciences, University of California, Merced.

31

32

## 33 **1 Introduction**

34

35 Speech is inherently variable. One mechanism that is thought to cope with this  
36 variability is categorical speech perception, where phonological categories are  
37 processed in a discontinuous fashion despite continuity in the phonetic domain.  
38 In studying the mechanisms that underlie categorical perception and other  
39 speech perception processes, researchers sometimes contrast the variability of  
40 speech with a postulated constancy of the listener's perceptual behavior (Liber-

man and Mattingly 1985; Stevens 2005). But, speech perception is far from constant: it is both flexible and plastic. Under “flexibility”, we understand the relatively short-term changes of the perceptual system that result from the immediate context such as the preceding or following sound (e.g., Repp and Liberman 1990). Under “plasticity”, we understand the relatively long-term changes such as in perceptual learning, where exposure to particular sounds produces changes that persist over time (Kraljic and Samuel 2005; Eisner and McQueen 2006). To permit the reliable perception of phonological categories amidst the inherent variability of speech, the perceptual system needs to be able to change its behavior in a principled manner, and only under certain conditions. This is where the interplay of flexibility and plasticity is crucial.

Dynamical systems theory is ideal for understanding these time-varying processes because it allows for the modeling of perception and learning as unfolding across multiple time scales. The perspective of dynamical systems has provided novel insights into numerous domains of human behavior, including movement coordination (e.g., Hock, Schöner, and Giese 2003), interpersonal coordination (e.g., Schmidt, Bienvenu, Fitzpatrick, and Amazeen 1998), learning (e.g., Zanone and Kelso 1992; Vallabha and McClelland 2007; Tuller, Jantzen, and Jirsa 2008), visual perception (e.g., Kawamoto and Anderson 1985) and speech perception (e.g., Grossberg and Myers 2000; Grossberg and Kazerounian 2011). With dynamical systems theory comes a commitment to look at the continuous aspects of perceptual and cognitive phenomena rather than purely looking at stable categories and discrete changes (Spivey 2007). Moreover, the perspective of dynamical systems focuses on studying the interactions between different processes rather than processes in isolation (Thelen and Smith 2006).

In this paper, we propose a connectionist model of speech identification, the behavior of which can be understood in terms of dynamical systems theory (cf., e.g., Spivey 2007; McClelland and Vallabha 2009). We test our model experimentally with a forced choice speech identification task, where listeners make perceptual decisions (such as, “Is this a /ba/ or /pa/?”) about phonetic items from an acoustic continuum. The response pattern is usually categorical, with listeners judging stimuli to be the same across a large variation along the continuum, but then abruptly providing a different response at a critical value of the acoustic parameter (e.g., Liberman, Harris, Hoffman, and Griffith 1957; Liberman, Cooper, Shankweiler, and Studdert-Kennedy 1967; Best, Morrongoiello, and Robson 1981).

Most experiments using this design try to avoid order effects by presenting different stimuli from the continuum in a random fashion. In contrast to this practice, Tuller, Case, Ding, and Kelso (1994) presented stimuli in an ordered fashion, starting at one end of the continuum and proceeding through all steps to the other end. They observed two different response patterns: contrastive behav-

1 ior and conservative behavior. With contrastive behavior, the perceptual switch  
2 from, say, /ba/ to /pa/, is anticipated, emphasizing the acoustic differences be-  
3 tween the stimuli on the initial extreme of the continuum and the stimuli which  
4 follow (Repp 1980). With conservative behavior, participants stick to the choice  
5 they made for the preceding stimulus, thus protracting the perceptual switch.

6 Tuller et al. (1994) modeled their data analytically with a differential equa-  
7 tion and proposed that the conservative behavior was due to a form of perceptual  
8 hysteresis (see also Case, Tuller, Ding, and Kelso 1995). Hysteresis is a typical phe-  
9 nomenon observed in many dynamical systems: if a system produces one re-  
10 sponse and this is also compatible with newly incoming input, the system tends  
11 to reproduce that response. The system changes its response only if the incoming  
12 input is incompatible with the last produced response. In both Tuller et al. (1994)  
13 and Case et al. (1995), the aim was to show the presence of contrastive and con-  
14 servative behavior and to interpret it as evidence for the dynamical nature of  
15 speech perception. Their analytical model used differential equations to account  
16 for both types of behavior, but contrastive behavior was only reproduced with an  
17 ad-hoc solution.<sup>1</sup>

18 Moreover, both studies did not consider response times, a crucial measure  
19 that offers a window into the temporal characteristics of the processes that are  
20 involved in categorical speech perception. It is known that response times are  
21 slow with maximally ambiguous stimuli (Studdert-Kennedy, Liberman, and Ste-  
22 vens 1963; Pisoni and Tash 1974; Repp 1981). Massaro and Cohen (1983) model  
23 this slowing down for speech identification experiments where stimuli are pre-  
24 sented at random. But, what happens if the category switch is shifted in an  
25 ordered presentation because of conservative or contrastive behavior? This is  
26 one aspect of categorical speech perception that we model and subsequently test  
27 experimentally.

28 We adopt the hypothesis that during speech identification, abstract linguistic  
29 representations compete for activation (cf. Grossberg 1973; McClelland and Val-  
30 labha 2009), and that this fast competition process is modulated by slower pro-  
31 cesses such as learning and habituation. Thus, in the model that we are going to  
32 present, the processes underlying fast perceptual competition and slow learning  
33 and slow habituation constrain each other in real time. We chose to investigate  
34

35

36

37 **1** In Tuller et al.'s (1994) model, the input to the dynamic equation which drives the behavior of  
38 the model depends both on the stimuli acoustics and on the position of the stimuli in the  
39 sequence. The function which combines these two factors can be trimmed in such a way that the  
40 effect of the acoustic difference between two consecutive stimuli is stronger in the second half of  
41 a sequence of stimuli. When this happens the perceptual switch is anticipated in the second half  
42 of the presentation and a more contrastive behavior is observed.

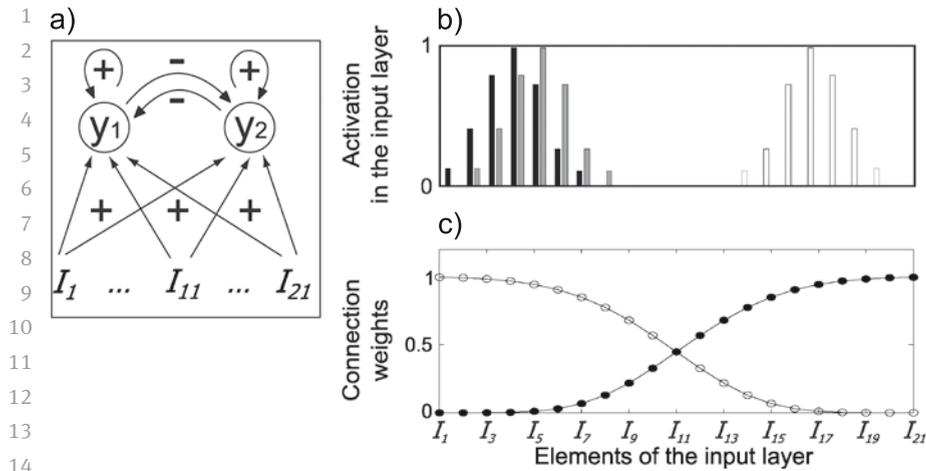
the interplay between these different processes because they provide explanations for the flexibility and plasticity of perception when studied in isolation (Kawamoto and Anderson 1985; Bogacz et al. 2006; Vallabha and McClelland 2007). When these mechanisms are taken together, they not only predict response times during identification tasks, but also other, more long-term effects: as the model learns, its sensitivity to small acoustic differences is heightened, mirroring the way in which listeners' perception of ambiguous, unfamiliar, or degraded speech sounds becomes gradually more reliable with experience (Maye, Aslin, and Tanenhaus 2008; Bradlow and Bent 2008; Winters and Pisoni 2008; Samuel and Kraljic 2009). And, as the model becomes more and more sensitive due to learning, it shifts from hysteresis to contrastive behavior (Tuller 2005; for a similar effect with audiovisual stimuli see Vroomen, van Linden, de Gelder, and Bertelson 2007).

In Section 2, we discuss the model in more detail. We discuss its architecture (Section 2.1) and the resulting dynamics (Section 2.2), as well as the model's behavior in a specific speech identification task (Section 2.3). Section 3 describes a pilot experiment that provides a first test of our model and that disentangles the effects of perceptual learning from other long-term effects such as fatigue or boredom. The main experiment (Section 4) then tests the full set of predictions that our model produces. Section 5 discusses the overall results, as well as some unexpected findings.

## 2 The model

### 2.1 Model architecture

Our model is composed of two layers (see Figure 1a). The input layer contains 21 nodes that represent 21 stimuli on an acoustic continuum. When no stimulus is presented, all of these nodes have zero activation. A stimulus at a position  $i$  on the continuum is modeled by a bell-shaped bump of activity centered at position  $i$  in the input layer. This means that with every stimulus located at a position on the continuum corresponding to a particular node, there is always some activation of the surrounding nodes (see Figure 1b). In Figure 1b, the two stimuli indicated by the black and the grey activation patterns are almost overlapping, corresponding to a high degree of similarity between these two stimuli. The stimulus indicated by the white activation pattern on the right side of the continuum has no overlap with the other stimuli.



**Fig. 1:** (a) The architecture of the model with a simplified depiction of the input layer ( $I_1, \dots, I_{21}$ ) and the two nodes of the competition layer ( $y_1$  and  $y_2$ ) that correspond to the two phonological categories. Plus signs indicate excitatory connections, minus signs inhibitory connections. (b) Three example stimuli that are indicated by bell-shaped activation patterns. (c) The weights of the connections between the input layer and node  $y_1$  (white dots) and node  $y_2$  (black dots).

The competition layer consists of only two nodes, one for each linguistic category in the forced choice identification task. Every node of the input layer is connected to both nodes of the competition layer. The incoming activation to the competition layer is multiplied by the weight of each connection. The activation of the two category nodes of the competition layer is a function of the sum of the weighted activations from the input layer. Before any stimuli are presented to the model, the weights are distributed so that towards each end of the continuum, input nodes tend to activate one category more than the other (see Figure 1c). This captures the observation that listeners, because of their linguistic background, already partition the acoustic continuum into two categories prior to the experiment. Towards the middle of the continuum, the two category nodes receive an increasingly similar amount of activation, rendering stimuli in this region more ambiguous. The connection weights can thus be likened to a filter that amplifies or attenuates the signals coming from a specific region of the input layer.

Within the competition layer, each category node has recurrent connections to itself. This means that strong activation provides additional input, allowing the system to settle into stable perceptual states more quickly. The nodes also have inhibitory connections between each other. The higher the activation of a node, the stronger its inhibitory effect on the other node. This introduces a “win-

ner takes all” dynamic, where after a short time in the competition process, only one node is strongly activated, and only one linguistic form at a time is perceived. This effectively leads to categorical speech perception, where acoustic/phonetic gradation along a continuum is not interpreted as phonological gradation.

## 2.2 Model dynamics

The activation levels within the input layer do not change during the presentation of a stimulus. The time-varying nature of the model emerges from the competition process, which in this model is governed by the competition equations introduced by Grossberg (1973). The model architecture depicted in Figure 1a corresponds to the schematic model formula in Equation 1. This equation is a simplified version of the actual model equation that is discussed in Appendix A (Eq. A1; see Grossberg [1973, 1976] for a detailed mathematical treatment).

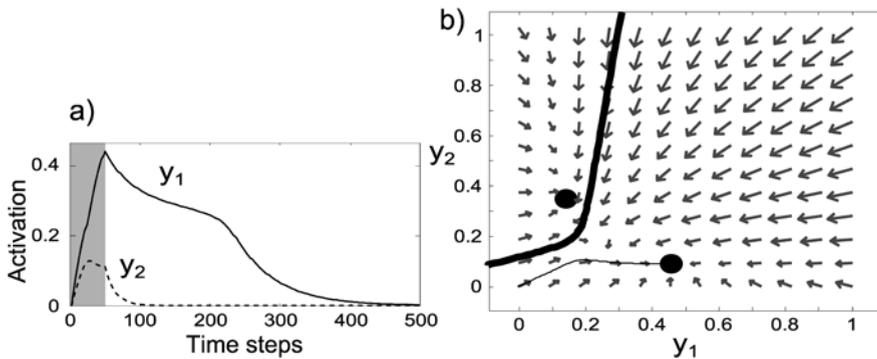
$$\Delta y_j = (1 - y_j)(input_j \cdot z_j + F_j) - y_j G_j - y_j L \quad (1)$$

Ideally the model changes in continuous time. However, our implementation proceeds in discrete time steps (1 step corresponds to 20 ms).  $\Delta y_j$  is the amount of change in the activation of node  $y_j$  observed at each time step. If this quantity is negative, the activation of  $y_j$  decreases. If this quantity is positive, the activation increases. The right-hand side of this equation is composed of three additive terms. The first additive term represents the positive contributions to the activation of  $y_j$ . Two kinds of positive signals reach the node:  $input_j$  represents the sum of the weighted signals from the input layer to  $y_j$ . This is multiplied by the efficiency variable  $z_j$ . Low values for this variable weaken the effect of incoming signals. The efficiency variable itself is changing at a slower time scale (Eq. A3, Appendix A). This variable models the effects of habituation: if a node becomes continually activated, its efficiency variable lowers and incoming bottom-up signals become decreased.  $F_j$  represents the recurrent (auto-excitatory) signal sent from  $y_j$  to itself (curved connections in Figure 1a).  $F_j$  is obtained by submitting the activation of node  $F_j$  to a sigmoid function. Finally, multiplying everything within this first additive term by  $(1 - y_j)$  assures that the activation value of  $y_j$  cannot exceed 1.

The second additive term (and the first negative contribution to the change in activation  $\Delta y_j$ ) is represented by  $G_j$ . This is a function of the activation of the competing node, represented by the lateral connections between the nodes in Figure 1a. This term implements the competition between the two nodes, where the activation of one node inhibits the activation of the other. Because  $G_j$  is multiplied by

1 the shunting term  $y_j$ , the activation of a node has a lower limit of 0 (as  $y_j$  ap-  
 2 proaches 0,  $G_j$  is multiplied by a smaller value and its effect decreases). The last  
 3 additive term of this equation,  $y_j L$ , represents the passive leakage of a node. It  
 4 corresponds to a constant multiplied by the activation  $y_j$ . When a stimulus is on  
 5 and a positive bottom-up signal reaches the node, passive leakage limits the  
 6 node's peak activation. After stimulus offset, the leakage term overcomes the re-  
 7 current auto-excitatory input, and without any further contributions from the  
 8 input layer, the activation of node  $y_j$  will rapidly recede to 0 at stimulus offset.  
 9 This decrease is contrasted by the self-sustaining signal traveling through the re-  
 10 current connections ( $F_j$ ): since the self-sustaining signal is a function of the  
 11 node's activation  $y_j$ , the higher the activation of a node at the stimulus offset, the  
 12 slower its decay.

13 The behavior of the two competitive nodes during the presentation of a stimu-  
 14 lus can be represented either by the time-dependent activation values of the two  
 15 nodes (Figure 2a) or by a trajectory on the two-dimensional state space of the  
 16 system (Figure 2b). The activations in Figure 2a are initially set to 0. They start  
 17 growing at stimulus onset and they start decaying back to 0 at stimulus offset.  
 18 The activations of the two nodes at instant  $t$  define the state of the system at that  
 19 instant and correspond to a point on the system's state space (Figure 2b). The axes  
 20 of the state space indicate the activation values of the two nodes. Therefore, the  
 21 coordinates of the points on that plane correspond to pairs of activation values,  
 22

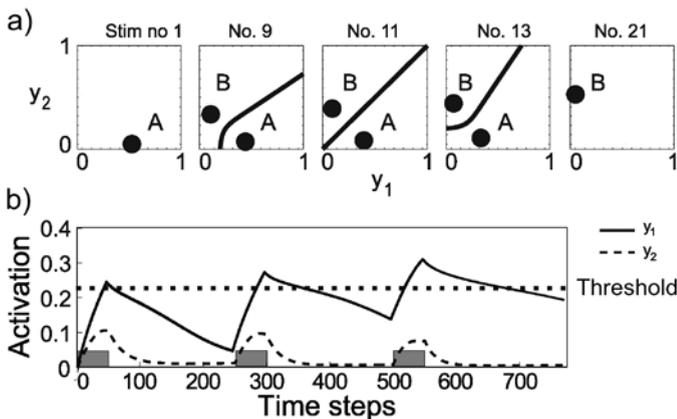


35 **Fig. 2:** (a) Evolution over time of the activation of the two nodes. The shaded area indicates the  
 36 time interval during which the stimulus is presented. (b) State space of the system. The state of  
 37 the system at a given instant is represented by a point on this plane. The coordinates of this  
 38 point represent the activation of node  $y_1$  and  $y_2$  respectively. Arrows indicate the direction of the  
 39 system dependent on its current state; the circles represent the system's attractors. Each  
 40 attractor is associated with one possible percept. The thick line is called a separatrix. When the  
 activations grow in Fig. 2a, the system follows the trajectory indicated by the thin line in Fig. 2b.

and a trajectory through the state space corresponds to the evolution over time of the activations of the two nodes.

The arrows in Figure 2b represent the velocity vectors of the system and depict the “pull” of the system towards certain patterns of activation. The system’s state changes according to the direction of the arrow located at the corresponding coordinates in the state space. The rate of change of the system state also depends on its position in the state space, and it is proportional to the length of the corresponding arrow (i.e., the magnitude of the velocity vector). The orientation of the arrows is such that the system tends to be drawn to either one of system’s attractors. The separatrix (the thick line in Figure 2b) partitions the state space into two attractor basins where the system that initially has its state located on one side of the separatrix is pulled towards the attractor on the same side. When the system is close to one attractor, only the activation of one node is high enough to exceed a perceptual threshold and trigger the perception of the corresponding category. This property lets us associate each attractor with one of the two perceptual categories. Once the system’s state has settled into one of these attractors, it does not change any more unless the input changes.

Each stimulus on the continuum creates a different state space with a different attractor layout (see Figure 3a). Each attractor layout defines the possible tra-



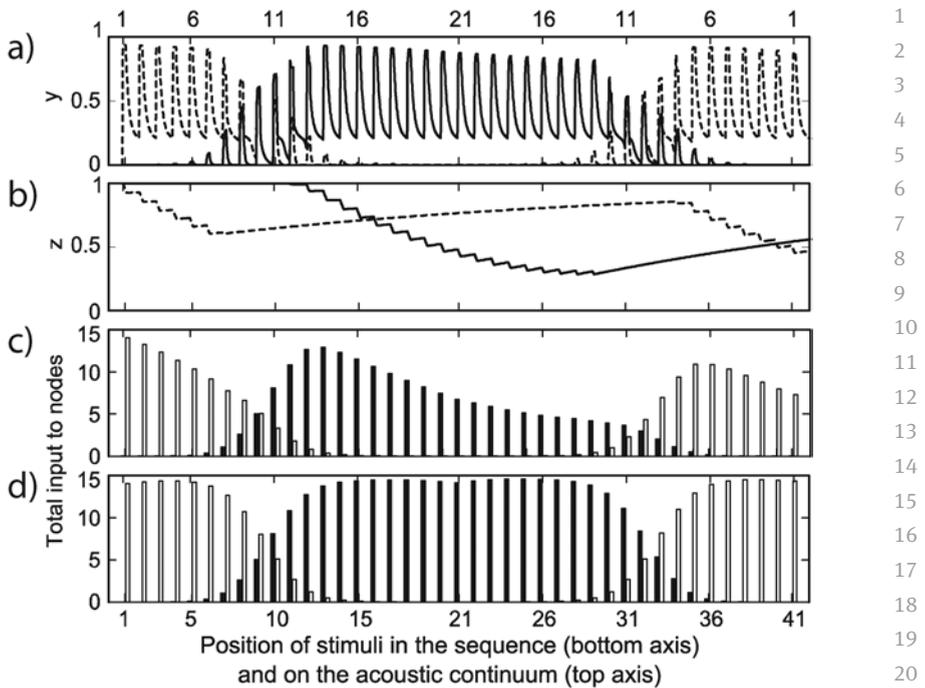
**Fig. 3:** (a) Different attractor landscapes with attractors (circles) and separatrices (lines) corresponding to five stimuli from the acoustic continuum. (b) Activation of one node when its most supporting stimulus (represented by grey rectangle) is presented repeatedly. Stimulus duration is 50 time steps (one step = 20 ms), during which the two nodes reach their peak activation values. The dashed line indicates the threshold for perception of the associated category. Note that the activation peaks of the winning node increase with repeated presentation of the same stimulus due to the fact that residual decay is slow.

jectories that the system can follow, ultimately leading to one of the two attractors. The attractor layout is not only affected by the incoming stimulus, but also by the value of the efficiency variables for each node ( $z_1$  and  $z_2$ ) and by the state of the learning process that affects the connection weights. The configuration of the connection weights at the beginning of the simulation is such that near the extremes of the continuum (stimulus 1 or stimulus 21), there is only one attractor and thus only one percept possible. However, stimuli near the center of the continuum have two attractors.

Due to the presence of the passive leakage term in Equation 1, each node is subject to gradual decay at stimulus offset. This means that the activation slowly recedes towards zero once no input is present. A node that has won the competition process will recede back more slowly to 0 than a node that has lost the competition process because its self-sustaining signal slows down the decay. If consecutive stimuli are presented with sufficiently small delay, the activation will not have fully receded back to 0. This residual activation is an additional input which biases the competition towards the results of the competition process from the preceding stimulus. Figure 3b shows the system's response to the repeated presentation of the same stimulus, where the bias towards the winning node tends to increase due to recurrent connection to itself and due to the fact that activation does not completely decay until the next stimulus is presented. This pattern leads to hysteresis, where the model tends to stick to a response that has already won the competition process.

If an arbitrarily defined perceptual threshold is crossed, the identification of the associated category is triggered and learning is allowed to happen. After a node has won the competition process, those connections from the input layer that contributed to its winning are strengthened, gradually increasing the node's sensitivity to specific regions of the input layer. Learning is thus simply a form of associative strengthening of the connection weights that depends on a combination of above-threshold activation of the category node and high activation of the input node. Such a learning law is compatible with Hebbian learning (a learning process where a synaptic connection between two nodes strengthens when both nodes are simultaneously active).

The third process that affects the behavior of the model is the habituation process. When a node becomes habituated, the signal coming from the incoming connections is reduced. The rate of change of habituation is slower than the rate of change of the competition process but faster than the rate of change of the learning process. Figure 4 illustrates the effects of habituation over time during the sequential presentation of 21 stimuli ordered as proposed by Tuller et al. (1994). Stimuli are presented sequentially from one extreme of the continuum to the other in the first half of the sequence, and then back again to the initial ex-



**Fig. 4:** Illustration of the habituation mechanism. (a) The activation levels of the two nodes with the habituation variable active. (b) The values of the efficiency variables of the two nodes over a sequence of stimuli. (c) The total amount of bottom-up input reaching the two nodes at each stimulus (differently colored bars represent input to different nodes). (d) Total input with habituation switched off by fixing the efficiency variable from Eq. 1 to the value 1.

trème in the second half. Figure 4a shows the activations of the two nodes. One node wins the competition at the beginning and at the end of the sequence, where the stimuli from one side of the continuum are presented; the other node wins the competition in the middle of the sequence, where stimuli from the other side of the continuum occur. Figure 4b tracks the development of the efficiency variables  $z_1$  and  $z_2$ . When a node reaches high activation values, its efficiency variable decays (Eq. A3, Appendix A). It gradually recovers back to 1 when the node has low activation values. However, the decay of the efficiency variable from 1 to 0 is much faster than the recovery. Thus, if a node wins the competition process over several consecutive stimuli (resulting in repeatedly high activation levels), the efficiency variable reaches very low levels and the node gets habituated. The efficiency variable of a node is multiplied by the total input arriving at the node. Figure 4c shows the total input reaching the nodes at each stimulus *with* habituation; 40

1 Figure 4d shows the total input prior to multiplication by the efficiency variable.  
2 In Figure 4d, moving from one extreme of the sequence towards the middle (i.e.,  
3 from one side of the continuum to the other), the input arriving at one node in-  
4 creases, while the input arriving at the other node decreases. This is due to the  
5 configuration of the connections' weights. In Figure 4c, on the other hand, the  
6 values of the input signals are shaped by the efficiency variable. The reduction of  
7 the input signal with habituation means that, slowly, the alternative node is fa-  
8 vored in the competition process.

9 In the model of speech categorization proposed by Tuller et al. (1994), the  
10 perceptual process is considered as a whole and nonlinearity is a macroscopic  
11 feature of speech identification necessary to obtain hysteresis. In our model, hys-  
12 teresis results primarily from the recurrent auto-excitatory connections of a node  
13 to itself, and nonlinearities are present in the microscopic aspects of the model  
14 behavior in a number of different ways. First, activity coming from recurrent au-  
15 to-excitatory and lateral inhibitory connections is transformed by a sigmoid func-  
16 tion (Eq. A1, Appendix A). The sigmoid function (Eq. A2, Appendix A) reduces the  
17 effects of small activation values and boosts the effects of large activation values.  
18 Learning also depends on a sigmoid function of the activation values (Eq. A4,  
19 Appendix A), ensuring that learning is not over-sensitive to small activation  
20 values and that only high activation levels associated with the winning node  
21 trigger learning. The rate of change of the habituation process also depends on a  
22 sigmoid transformation of that node's activation (cf. Eq. A3, Appendix A), assur-  
23 ing that habituation acts only when high activation values are reached.

24

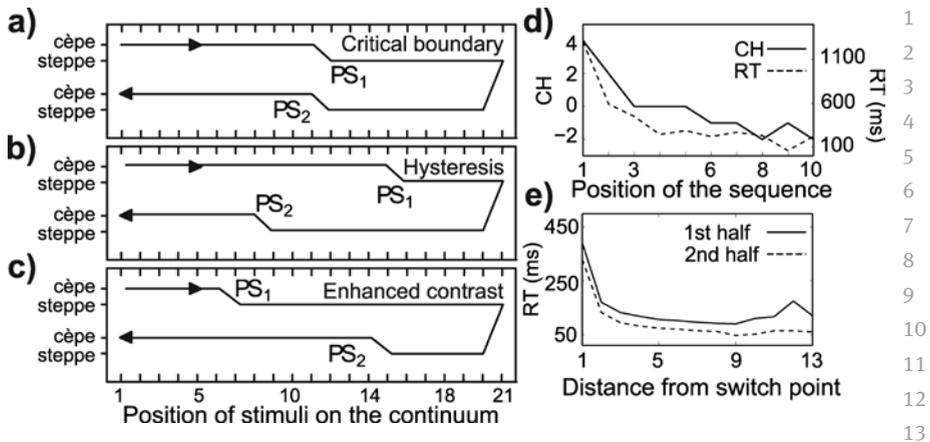
25

### 26 2.3 Task-specific behavior and predictions

27

28 Our model performed a binary forced choice task, where 21 stimuli from an acous-  
29 tic continuum had to be identified either as the French word *c pe* ([sep], a variety  
30 of mushroom), or as *steppe* ([step], English steppe). Stimuli were presented se-  
31 quentially in an increasing-decreasing order (henceforth, ID) or a decreasing-  
32 increasing order (DI). The increasing half of the ID sequence starts at the [sep]  
33 endpoint of the continuum and moves towards [step]. The decreasing half moves  
34 back from [step] to the starting point [sep]. Thus, a complete sequence contains 41  
35 stimuli in total (the stimulus occurring in the middle is not repeated). The  
36 decreasing-increasing sequence (DI) represents the opposite pattern (first half: 21  
37 to 1, second half: 1 to 21).

38 If the activation solely depended on the incoming input and static connec-  
39 tion weights, a perceptual switch from one category to another would always oc-  
40 cur at exactly the same point, regardless of whether the order was increasing or



**Fig. 5:** (a–c) Possible response patterns with ID order. The x-axis shows the position of a particular stimulus (indicated by tick marks); the y-axis shows a particular response choice. Arrows indicate the order of the responses. (d) CH index (continuous line with scale on left y-axis) and response times (dashed line with scale on right y-axis) for multiple sequences within a long experimental session. The horizontal axis represents the rank of each sequence in the experiment (e.g., “1” indicates the first ID or DI sequence). (e) Response times as a function of the distance from the perceptual switch point in the first half (continuous line) and in the second half (dashed line) of the sequences.

decreasing. This pattern is labeled as “critical boundary” in Figure 5a. Hysteresis, on the other hand, is characterized by a perceptual switch farther away from the initial end point of the continuum in each half of the sequence (Figure 5b), indicating that the system sticks to a response pattern. Switches closer to the initial end point in each half of the sequence represent contrastive behavior (Figure 5c). The amount of contrast or hysteresis can be quantified by measuring the difference between the position on the continuum of the switch point in the increasing half and of the switch point in the decreasing one, a measure which, following Nguyen et al. (2005), we call contrast-hysteresis index (CH index). If this index is negative, the model’s behavior can be characterized as contrastive; if the sign is positive, the behavior is conservative.

We ran the model on 5 increasing-decreasing (ID) and 5 decreasing-increasing (DI) sequences, as well as 10 sequences with random presentation order. ID and DI sequences were alternated, always interleaved with a random sequence (e.g., ID-random-DI-random-ID etc.). As the first stimulus of a sequence only contains the attractor for one category, the corresponding node wins the competition. It keeps winning the competition process so long as the attractor landscape for each stimulus contains the initial attractor. The vanishing of the

1 attractor on which the system resides is called critical instability. Starting from  
2 this point on the acoustic continuum, the state space produced by stimuli shows  
3 only the attractor far from the one visited by the system in previous trials. The  
4 system is then forced to move to the new attractor and a categorical switch is  
5 observed (cf. Figure 3a).

6 Although the presence of critical instability adds an element of discreteness  
7 to the system's behavior, the attractor landscape changes continuously across the  
8 stimulus continuum. Each attractor is characterized by a degree of relative stabil-  
9 ity, corresponding to its attraction strength. This attraction strength translates  
10 into the velocity of movement of the system through the attractor landscape. The  
11 system moves faster when directed towards a more stable attractor, i.e., an attrac-  
12 tor towards the end points of each continuum. Closer to the region of critical in-  
13 stability near the perceptual switch, the system is slower. This behavior is also  
14 known as “critical slowing down” (e.g., Kelso, Scholz, and Schöner 1986). In Fig-  
15 ure 5e, the continuous line represents the RTs<sup>2</sup> *before* the perceptual switch in the  
16 first half and the dashed line represents the RTs *after* the perceptual switch in the  
17 second half of each ID or DI sequence. We can observe that near the perceptual  
18 switch, responses tend to be slowed in both halves of the sequence. However, RTs  
19 following the second switch are faster than RTs preceding the first switch. This  
20 can be explained as an effect of the habituation process: Before the second per-  
21 ceptual switch, the competitor node has had high activation levels throughout a  
22 long sequence of stimuli (cf. Figure 4a), triggering the habituation mechanism.  
23 Then, once the second perceptual switch occurs, this node is not a good competi-  
24 tor anymore because it is habituated. Before the first perceptual switch in the  
25 first half of the sequence, the alternative node has not been activated and is thus  
26 not habituated. This means that it exerts more inhibitory influence, with the con-  
27 sequence that RTs are slower.

28 In Figure 5e we can also observe that RTs increase slightly for stimuli maxi-  
29 mally distant from the perceptual switch. At the beginning of a sequence, the  
30 model is relatively slow to respond to the very first stimulus because a node can-  
31 not benefit from residual activation. If the same node wins repeatedly, its residual  
32 activation and the increasing influence from auto-excitatory connections make  
33 the system respond faster. When, due to the effect of learning, the activation of  
34 the nodes are high enough to trigger the habituation mechanism and to anticipate

35

36

37

38

39 **2** In our simulations, we measure response times from the time point of stimulus onset to the  
40 time point when the activation of the winning node exceeds the perceptual threshold.

the perceptual switch in the second half of a sequence, the winning node becomes habituated at the very end of the sequence, slowing down its response.<sup>3</sup>

Throughout the simulation, overall connection weights strengthen as a result of learning. This has various effects: (1) the increased connection weights produce stronger dynamics, resulting in higher activation peaks, faster response times, and more resistance to noise (see discussion of noise below); (2) an increase in the average input arriving to the two nodes of the competition level produces a narrowing of the bistable region of the continuum (a relatively smaller number of stimuli produce a bistable attractor landscape), decreasing the value of the CH index (cf. Figure 5d); (3) the increased peak activation levels are strong enough to trigger the habituation process. This leads to a higher degree of contrastive behavior later in the experiment.

## 2.4 Internal noise

The behavior of the model described so far is completely deterministic. However, due to their inherent complexity, real systems (including the human perceptual system; cf. Haken 1983: 200–202) are characterized by internal noise. Such noise could randomly push the model away from its current position in the state space. We can model this by adding to Equation 1 a term that randomly takes values from a normal distribution at each time step, pushing the system away from its current position in the state space. If such a random term is added to the model, a response not only depends on the attractor landscape, but also on the degree of noise. If noise is strong and attractors are weak, the system can be randomly pushed from one attractor basin to another, which ultimately can result in a different response. Response changes due to noise are more generally instances of “critical fluctuations” (Kelso, Scholz, and Schöner 1986). These are expected to occur more frequently near the region of critical instability where the attractors are weaker and noise can have a larger effect. A strong attractor, on the other hand, is more resistant because noise is less likely to pull the system out of its attractor basin. Response changes due to critical fluctuations should be interpreted differently from response changes due to the critical instability itself (i.e., the vanishing of the attractor on which the system resides). Henceforth, we call the last response change in one half of a sequence a “perceptual switch” (which

---

<sup>3</sup> When the activation dynamics are strong enough to trigger habituation, nodes get habituated before each switch point. However the effect of habituation on response times before the switch points is confounded with the critical slowing down (due to increased competition) which precedes the switches.

1 is due to critical instability), and we call a response change before the last re-  
 2 sponse change a “flip-flop”. We can predict that the frequency of flip-flops, being  
 3 related to the relative stability of the attractors, increases as stimuli on the contin-  
 4 uum approach the perceptual switch (similar to the increase in RTs). And, we can  
 5 predict that the flip-flops should become less frequent as overall experience with  
 6 the stimuli and the task increases, due to the strengthened competition dynamics  
 7 and to the reduction of the bistable region of the continuum.

8 The presence of noise can have a seemingly counterintuitive effect on mul-  
 9 tistable systems such as the model we propose here: an increase of internal noise  
 10 can produce a decrease of the CH index. When a deterministic model is presented  
 11 with stimuli from the bistable region of the acoustic continuum, no category  
 12 switch is expected (the system switches only when leaving that region). However  
 13 in the presence of noise, the system can jump out of the initial attractor basin.

14 Figure 6 shows trajectories for five stimuli that have bistable attractor land-  
 15 scapes. Proceeding through the stimuli from left to right, the strength of the initial  
 16 attractor decreases while the strength of the alternative attractor increases. There-  
 17 fore the system is more and more likely to be pulled out from the initial attractor  
 18 (which gets weaker) to the alternative attractor (which gets stronger). With a  
 19 stronger noise component the perceptual switch is likely to occur earlier in the  
 20 ambiguous region, and thus we expect a reduction of the CH index. Thus, the  
 21 potential effect of noise is a source of ambiguity in the interpretation of the behav-  
 22 ior of the CH index: if a reduction of the CH index is observed after a prolonged  
 23 exposure to stimuli, this may be either due to the deterministic component of our  
 24 model (i.e., the effects of learning, increased competition dynamics, and in-  
 25 creased habituation), or it may be due to factors such as fatigue or boredom,  
 26 which can increase internal noise and which are likely to increase in importance  
 27 throughout a long experiment with human participants. Experiment 1 tries to dis-  
 28 entangle these two separate explanations of the reduction of the CH index.

29

30

31

32

33

34

35

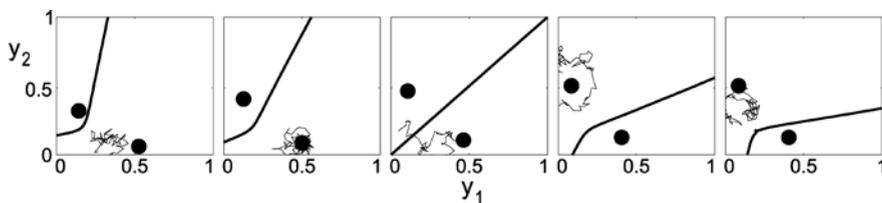
36

37

38

39

40



**Fig. 6:** Noisy trajectories for five stimuli of the continuum where there are two attractors (the bistable region). Stimuli were presented to the model in the left-right order. The middle picture represents the most ambiguous stimulus.

## 2.5 Ordinal nature of synthetic response patterns and pattern-oriented modeling

There are basically two approaches to comparing data obtained from simulations with empirical data: a quantitative and a qualitative, or ordinal, approach. With the quantitative approach, a model is evaluated by its goodness of fit to the data. With the ordinal approach, a model is evaluated through hypothesis testing on empirical data (Pitt, Kim, Navarro, and Myung 2006; Smolensky and Legendre 2006). Hypotheses correspond to predictions about the effect of the experimental conditions on the ordering of the data along some behavioral dimension (in our case values of the CH index and RTs). An ordinal approach is fruitful when, as in our case, the focus is on the general principles governing the modeled behavior. Adopting such an approach in the comparison of our model's behavior with empirical data means to test the following hypotheses: (1) the CH index, the RTs, and the amount of flip-flops decrease throughout the task; (2) the RTs and the frequency of flip-flops increase as consecutive stimuli approach the perceptual switch; and (3) responses are slower before the perceptual switch in the first half of a sequence than after the perceptual switch in the second half. If the ordinal relations listed here are inversed in the empirical data, our model should be considered inadequate. Note, moreover, that our validation approach complies with the principles of pattern-oriented modeling (Grimm et al. 2005), where it is considered insufficient if a model produces a single response pattern but instead, a model should be able to account for multiple response patterns.

To assess our model's predictions, we followed two directions. Because of the ambiguity of interpreting the CH index in noisy systems mentioned above (see Section 2.4), Experiment 1 was designed to provide evidence for a true effect of learning and to minimize the potential effects of noise due to fatigue or boredom. Also, this experiment serves to replicate Tuller et al.'s (1994) experiment with different stimuli, a different population (French speakers), and a different language (French). In Experiment 2, we looked at the effects of learning on the CH index and response times over the course of a long experiment, testing the full extent of our model's predictions.

## 3 Experiment 1

With Experiment 1, we tested the idea that the decrease of the CH index is in fact due to learning and not due to such factors as fatigue or boredom. To this end, we tested two populations with different levels of experience with the task. If the most experienced listeners show lower CH values, it is reasonable to assume that

1 this measure depends on listeners' experience and thus on learning. In this cir-  
2 cumstance, it would indeed be hard to interpret the lowering of the CH index as  
3 due to the effect of random noise, especially if the experienced participants have  
4 less variable response patterns.

5

6

### 7 **3.1 Participants**

8

9 Six experienced phoneticians and six academics from the humanities volun-  
10 teered to participate in the experiment (both groups come from the Université de  
11 Provence, France). There were 4 male and 8 female participants, equally distrib-  
12 uted across the two groups (mean age: 32). Phoneticians can be considered expert  
13 listeners, and the other group can be considered naïve listeners.

14

15

### 16 **3.2 Stimuli**

17

18 Stimuli form an acoustic continuum ranging from [sep] to [stɛp]. The two words  
19 differ only in the closure duration that follows the fricative [s]. A long closure  
20 duration indicates the presence of the stop in *steppe*; the absence of a closure  
21 duration indicates the absence of a stop in *cèpe*. We constructed a stimulus mid-  
22 way between [sep] to [stɛp] and systematically varied the closure duration in 4 ms  
23 steps from 0 to 56 ms, resulting in a 15-step continuum. The stimuli construction  
24 is detailed in Appendix B, Section 1.

25

26

### 27 **3.3 Procedure**

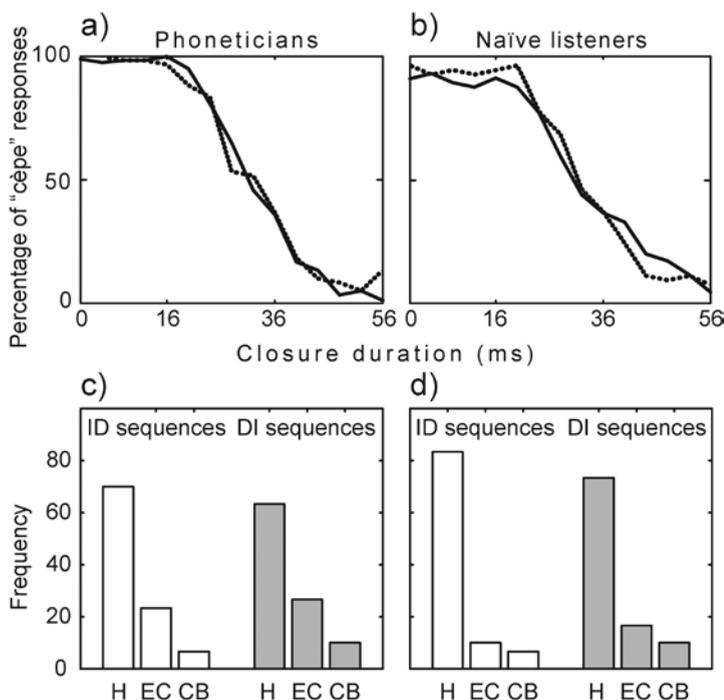
28

29 The experiment was preceded by a training session, where participants listened  
30 five times to the endpoint of the continuum with simultaneous orthographic pre-  
31 sentation of the respective word. In the main experiment, as in our simulations  
32 outlined in Section 2.3, stimuli were presented sequentially in 5 ID, 5 DI, or 10  
33 random sequences (20 sequences total). ID and DI sequences were randomly al-  
34 ternated but each ordered sequence was always preceded by a random sequence  
35 (e.g., ID-random-DI-random-DI etc.). This prevented participants from basing  
36 the perceptual switch on counts of the stimuli. Each sequence pair (random-DI or  
37 random-ID) was followed by a pause of 6 seconds. Within each sequence, the  
38 inter-stimulus-interval was set to 2 seconds. There was a 5-minute break in the  
39 middle of the experiment. Including pauses, the experiment lasted approximately  
40 30 minutes.

### 3.4 Results

Figure 7a and Figure 7b depict the sigmoidal identification curves for the two groups. Each curve was obtained by averaging across all responses (pooled across speakers and presentation orders). The curves are very similar for random versus ordered presentation of the stimuli, and they look similar to what has been found in previous studies on categorical perception (e.g., Pisoni and Tash 1974). However, the curves are not symmetrical around the middle value but shifted towards *steppe*, suggesting that the continuum was modeled in a non-optimal way.

Figure 7c and Figure 7d depict the relative frequencies of the three different response patterns. Hysteresis was by far the most frequent pattern (average CH index: 3.02), but it was distributed differently across phoneticians and naïve listeners: whereas the CH index was 1.3 for phoneticians, it was 4.75 for naïve lis-



**Fig. 7:** Top row: Proportion of *cèpe* responses for phoneticians and naïve listeners over the closure duration continuum. Solid lines are responses in ordered sequences; dotted lines are responses in random sequences. Bottom row: Percentage of response pattern showing hysteresis (H), enhanced contrasts (EC), and critical boundary (CB) in ordered sequences for phoneticians and naïve listeners. White bars are for ID sequences, gray bars for DI sequences.

1 teners, indicating that the latter group exhibited more hysteresis and less con-  
2 trastive behavior. The reliability of this difference was tested using a generalized  
3 linear mixed model with Gaussian error distribution and identity link function.<sup>4</sup>  
4 The model included Group as a between-participants fixed effect and random in-  
5 tercepts for participants. Group had a significant effect on the CH index ( $p = 0.011$ ),  
6 with experts having a CH index that was smaller by about  $3.45 \pm 1.17$  (standard  
7 error).

### 10 3.5 Discussion

12 While our participants exhibited both contrastive behavior and hysteresis, the  
13 latter response pattern was the most frequent. Moreover, experienced listeners  
14 exhibited less hysteresis and more contrastive behavior. This suggests that con-  
15 trast and hysteresis are not solely due to random noise affecting the identification  
16 process around the perceptual boundary, but that the occurrence of these pat-  
17 terns is affected by experience with the stimuli and the task. This reasoning is  
18 based on the assumption that participants from both groups are similarly affected  
19 by fatigue or boredom. This may not be true and indeed our data suggest that, as  
20 expected, phoneticians show an overall less variable behavior (the ambiguous  
21 region of the continuum is reduced for these listeners, corresponding to a lower  
22 absolute value for the CH index). This may be interpreted as evidence that pho-  
23 neticians were more attentive and less subject to fatigue or boredom. If the reduc-  
24 tion of the CH index was solely due to noise, this higher degree of attention should  
25 translate into a higher CH index for phoneticians – contrary to what we observed.  
26 We are thus confident in viewing the CH index in speech identification experi-  
27 ments as an actual reflection of learning and competition dynamics.

28 While Experiment 1 looked at the effects of learning in a between-  
29 participants design, our main Experiment 2 investigated the within-participants  
30 development of the CH index over the course of a long experiment. This tests the  
31 results from our model more directly.

---

36 <sup>4</sup> We used R (R Development Core Team 2005) and the lme4 package (Bates, Maechler, and  
37 Bolker 2012). Plots of residuals against fitted values revealed no obvious deviations from  
38 normality or homogeneity. *P*-values were computed with MCMC sampling implemented in the  
39 package languageR (Baayen 2011). Unless otherwise noted, this applies to all models with  
40 Gaussian error distributions throughout the paper.

## 4 Experiment 2

Experiment 2 was conducted to observe changes in the perception of phonological categories at several timescales in greater detail. We tracked the CH index and response times, as well as the development of these dependent measures over the course of a long experiment. Based on our model, we expect critical slowing down around the perceptual switch. Also, since internal noise has a stronger effect when perceptual attractors are weak, we expect to observe more flip-flops near the perceptual switch in the region of critical instability. On a longer time scale, we expect participants to get faster overall, and to exhibit a development towards more contrastive behavior (lower CH values) as exposure to the stimuli and the task increases.

### 4.1 Participants

We tested 14 participants (5 male, 9 female; mean age: 28). Three participants were excluded from the analyses because of fatigue (we explicitly asked participants to stop in case they found the task harder and harder as the experiment proceeded). Data from one additional participant was excluded from the RT analysis because of technical failure.

### 4.2 Stimuli

Because of the asymmetry of the identification curves in Experiment 1, we used a different method to construct the stimuli in Experiment 2 (see Appendix B, Section 2 for details). The crucial difference was that we used two continua: one based on the vowel of [sep], the other one based on [step] (we call this factor “vowel type”). Each of the two continua was composed of 21 stimuli, with the closure duration between the fricative and the vowel ranging from 0 to 100 ms in steps of 5 ms.

### 4.3 Procedure

For each vowel of the two types, we tested 5 ID sequences, 5 DI sequences, and 10 random sequences (40 sequences total). The experiment started with an ordered sequence, which we considered a training sequence. We balanced the initial order (ID or DI) and the vowel type of the initial sequence across participants.

1 Each ordered sequence (with the exception of the very first one) was preceded by  
 2 a random sequence and was followed by a 6-second pause. The ordering of the  
 3 sequences was the same as in Experiment 1 and our simulations (ID-random-DI,  
 4 etc.). There was a 5-minute pause midway through the experiment, resulting in an  
 5 experiment that was approximately one hour long.

6

7

#### 8 4.4 Results

9

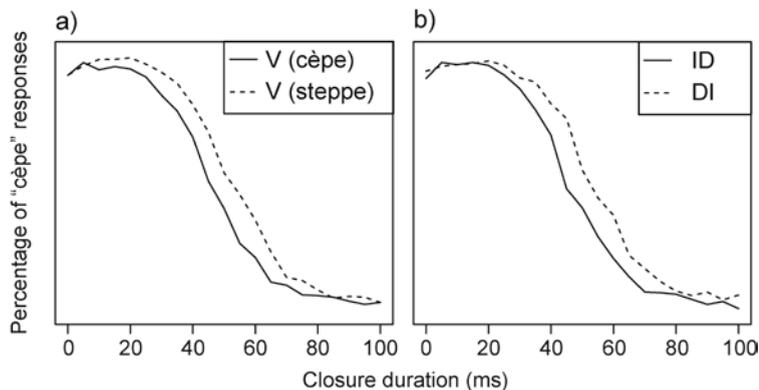
10 Figure 8a shows the identification curves separated with respect to the vowel  
 11 type. Although the new method of stimulus construction made the curves more  
 12 symmetrical, the trading relation between the closure duration and the vowel  
 13 goes in the opposite direction from what would be expected: for a given closure  
 14 duration, the stimuli which contained the vowel modeled on [sɛp] were identified  
 15 more often as [stɛp]. This result was unexpected.

16 Another unexpected result was that the identification curves were different  
 17 for participants starting the experiment with an ID sequence and for participants  
 18 starting the experiment with a DI sequence, shown in Figure 8b. Participants who  
 19 started with the ID sequence (whose initial stimulus was [sɛp]) tended to have an  
 20 identification curve with a higher proportion of [stɛp] responses. This surprising  
 21 result will be further discussed in Section 4.5.

22 To statistically validate our observations, we performed a series of general-  
 23 ized linear mixed models with binomial error structure and logit link function  
 24 (= mixed logistic regression) with the dependent variable response category

25

26



27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38 **Fig. 8:** Percentages of *cèpe* responses over the closure duration continua. (a) Curves for  
39 stimulus type, averaged over presentation orders. (b) Curves for participants who started with  
40 increasing-decreasing (ID) or decreasing-increasing (DI) sequences as the very first order.

([sep] vs. [step]).<sup>5</sup> Closure duration produced a significant effect for ordered 1  
 ( $p < 0.0001$ ,  $|Z| = 9.54$ ) and for random sequences ( $p < 0.0001$ ,  $|Z| = 11.14$ ). The 2  
 probability of a [sep] response decreased with longer closure duration by 0.99 3  
 from one extreme to the other of the continuum both in ordered sequences (log 4  
 odds ratio dropped by  $3.21 \pm 0.41$  for each increase in closure duration), and 5  
 in random sequences (log odds ratio  $-3.14 \pm 0.28$ ). Vowel type had a significant effect 6  
 for random ( $p < 0.0001$ ,  $|Z| = 4.82$ ) and ordered sequences ( $p < 0.0001$ ,  $|Z| = 8.42$ ) 7  
 as well: the probability of a [sep] response increased with vowels coming from 8  
 [step], by about 0.165 (log odds ratio =  $0.71 \pm 0.2$ ) for ordered and by 0.2 (log odds 9  
 ratio =  $0.87 \pm 0.18$ ) for random sequences. 10

In ordered sequences there was a significant interaction between the effect of 11  
 closure duration and sequence number ( $p < 0.0001$ ,  $|Z| = 7.37$ ), indicating that the 12  
 slope of the identification function became steeper with increasing exposure to 13  
 the stimuli. The initial sequence type (ID vs. DI) had a significant effect in both 14  
 ordered ( $p = 0.045$ ,  $|Z| = 2.00$ ) and random sequences ( $p < 0.01$ ,  $|Z| = 2.65$ ). The 15  
 order of presentation of the very first ordered sequence leads to a horizontal shift 16  
 of the identification curves. Participants who started the experiment with an ID 17  
 sequence were indeed less likely to perceive *cêpe* in ordered sequences by about 18  
 0.21 (log odds ratio =  $-0.88 \pm 0.44$ ) and by about 0.22 in random sequences (log 19  
 odds ratio =  $-0.94 \pm 0.356$ ). 20

To analyze flip-flops, the answers corresponding to the two different words 21  
 were considered separately. In the first half of an ID sequence, we considered the 22  
 [sep] responses before the perceptual switch observed in that half as an instance 23  
 of a flip-flop; in the second half of an ID sequence, we considered the [step] re- 24  
 sponses before the second perceptual switch as a flip-flop. The exact opposite 25

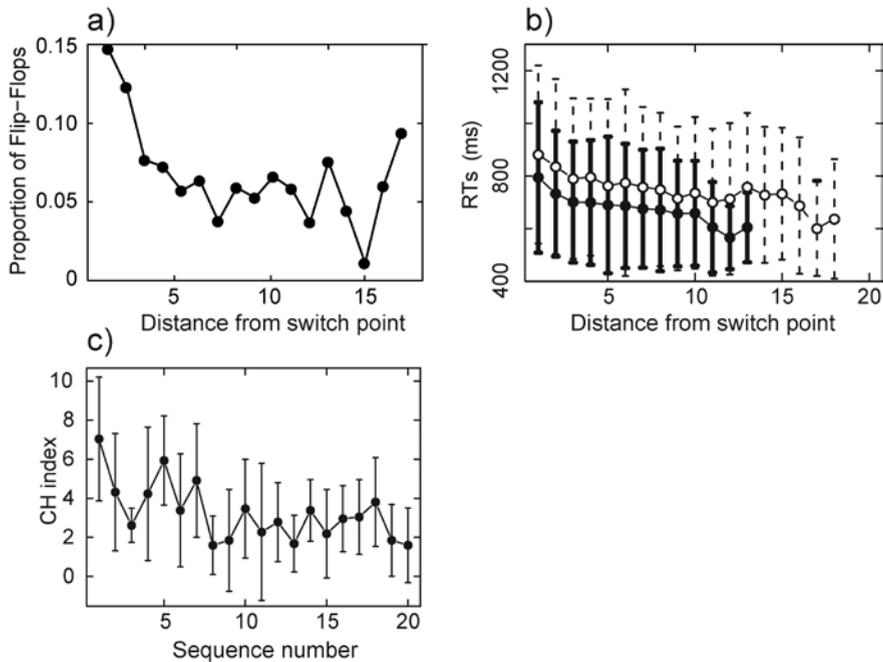
---

5 The model had the following fixed effects: closure duration, vowel type, sequence position in 29  
 the experiment ("sequence number", 1, 2, 3, etc.), initial presentation order (ID vs. DI), and initial 30  
 presentation vowel. We also included the following interactions: closure duration\*sequence 31  
 number, initial presentation order\*sequence number and initial vowel type\*sequence number. 32  
 The latter interaction did not become significant and was subsequently excluded from the model. 33  
 We included a random effect term for participants (random intercept) and participant-specific 34  
 random slopes for the distance to the switch point, for vowel type, and for sequence number. As 35  
 a general procedure for this and the following mixed models, we initially added a partici- 36  
 pant-specific random slope for each fixed effect and then removed those which did not produce 37  
 a significant decrease of the model's residuals (as assessed through likelihood ratio tests). We 38  
 analyzed the data obtained within random sequences and data obtained within ID and DI 39  
 sequences separately with different models of the same structure. Continuous predictors were 40  
 centered to a mean of 0 (corresponding to the average of the continuous predictor). Unless  
 otherwise noted, this applies to the continuous predictors of all subsequent models.

1 was done for DI sequences. In Figure 9a, the proportion of flip-flops is plotted as  
 2 a function of the distance of the stimulus to the switch point.

3 We performed mixed logistic regression on flip-flops with the fixed effects  
 4 “distance to the switch point” and “sequence number”, as well as the interaction  
 5 between these two factors. The effect of the distance from the switch point was  
 6 significant ( $p < 0.0001$ ,  $|Z| = 4.07$ ), as well as the effect of sequence number ( $p <$   
 7  $0.0001$ ,  $|Z| = 4.19$ ) and their interaction ( $p < 0.0001$ ,  $|Z| = 3.60$ ). The proportion of  
 8 flip-flops decreased from 0.064 for the stimuli positions closest to the switch  
 9 point to 0.003 for the most distant stimuli (log odds ratio per step on the contin-  
 10 uum:  $-0.17 \pm 0.04$ ). The flip-flop proportion decreased by 0.032 after 20 ordered  
 11 sequences of stimuli (log odds ratio per sequence:  $-0.06 \pm 0.01$ ).

12 Figure 9b shows the average response times against the distance from the  
 13 switch point. The very first and last stimuli of each sequence are not considered  
 14 here because, as noted in Section 2.3, these stimuli are expected to produce  
 15



16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36 **Fig. 9:** (a) Average proportions of flip-flops as a function of the distance from the switch point.  
 37 (b) Average response times and standard deviations against distance from the switch point.  
 38 Empty circles and dashed vertical bars: averages and standard deviations in the first halves of  
 39 the sequences. Filled circles and continuous bars: averages and standard deviations in the  
 40 second halves of the sequences. (c) Average value of the CH index and standard deviations with  
 respect to the position of the sequence in the experiment (sequence number).

slower response times. For the purpose of statistical analyses, we discarded response times shorter than 250 ms (cf. Jensen 2006). This resulted in 36 omitted data points (1.25%). To avoid deformation of the response time distribution, no ceiling was put on the data (cf. Ulrich and Miller 1994). We used the logarithm of response times as the dependent variable and the distance of the stimulus from the switch point, the half of the sequence (first vs. second), and the sequence number as fixed effects. We included participant as a random effect (random intercept), as well as participant-specific random slopes for the distance to the switch point and for sequence number.  $p$ -values were estimated using MCMC sampling.

Distance to the category switch had a significant effect on response times ( $p < 0.01$ ), which at the most distant stimuli from the category switch were about 171 ms faster than at the closest stimuli (log difference between consecutive stimuli:  $1.01 \pm 1.003$ ). Sequence half had a significant effect as well ( $p < 0.0001$ ), with RTs being 80 ms faster in the second half than in the first half (log difference between the two halves:  $1.06 \pm 0.04$ ). Sequence number also had a significant effect ( $p < 0.0001$ ): Participants were on average 139 ms faster at the end of the experiment (log difference between one sequence and the next:  $0.009 \pm 0.002$ ).

Figure 9c shows the development of the average CH value with respect to the sequence number, highlighting that there was a downward trend throughout the hour-long experiment. A Gaussian linear mixed effects model was run with CH as continuous dependent variable and sequence number as a fixed effect. We included random intercepts for participants (quantifying individual baseline differences in the CH index), as well as random slopes (quantifying individual differences in how sequence number affects the CH index). The position of the sequence had a significant influence on the CH index ( $p = 0.0017$ ), decreasing it by about  $0.15 \pm 0.05$  each sequence.

## 4.5 Experiment 2: Discussion

The analyses of response times and flip-flops confirmed that our data exhibited critical fluctuations and critical slowing down: participants responded more slowly and exhibited more flip-flops around their respective perceptual switch points. Overall, response times, the frequency of flip-flops, and the CH index decreased throughout the experiment. These results confirm the predictions made by our model. These results also refute the notion that the CH index in this experiment decreased purely because of increasing noise throughout the experiment: A decrease of the CH index due to noise should be paralleled by an increase of flip-flops, yet, the opposite was found.

1 There were also two unexpected findings. The first concerns the effect of the  
2 vowel type. When participants heard stimuli modeled from productions of *steppe*  
3 they were more likely to perceive *cèpe*. Although this unexpected effect is orthog-  
4 onal to our hypothesis, a possible explanation is suggested by looking at the de-  
5 tails of the stimuli of Experiment 2: the vowel for *cèpe* had a steeper rise in inten-  
6 sity at its onset compared to the vowel for *steppe*. The presence of a plosive  
7 between a fricative and a vowel usually involves an amplitude burst, correspond-  
8 ing to the plosive release. This sudden increase in intensity may have been inter-  
9 preted as a cue for the presence of a plosive. Also, we did not expect the initial  
10 presentation order to have a significant effect on identification curves throughout  
11 the experiment. In the following section we explore this novel result.

12

13

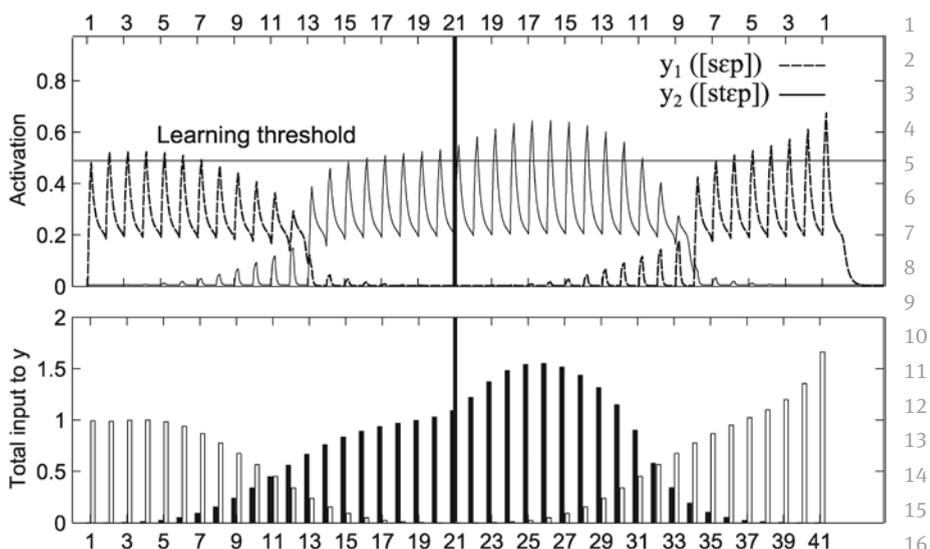
#### 14 4.5.1 The role of initial presentation order

15

16 At the end of the experiment, participants were more likely to identify a stimulus  
17 as an instance of the category they perceived at the middle of the very first se-  
18 quence of stimuli. For example, participants who started the experiment with an  
19 ID sequence (going from [sep] to [step] and back), were more likely to identify  
20 ambiguous stimuli as instances of the *steppe* category at the end of the experi-  
21 ment.

22 We explored this unexpected result with a post-hoc simulation. We could do  
23 this thanks to two features of our model. First, the presence of residual activation  
24 biases the competition in favor of the node that won recently. Second, stimuli at  
25 nearby positions on the continuum are represented by overlapping bell-shaped  
26 activation patterns in the input layer (cf. Figure 1b). Therefore, when a node  
27 learns to respond to a stimulus, learning strengthens its response not only to the  
28 present stimulus, but also to nearby stimuli. How precisely do these two aspects  
29 of our model lead to the observed first sequence effect?

30 Figure 10a shows the activation level of the two nodes during the very first  
31 ordered sequence within the simulation (in this case, ID order, starting with  
32 [sep]). A basic feature of the task employed is that in each sequence, the contin-  
33 uum is presented twice, with the stimuli presentation order reversed from one  
34 half to the other. Therefore the beginning and end of the ID sequence are most  
35 supportive for the [sep] node and the middle of the sequence is most supportive  
36 for the [step] node. During the presentation of this first sequence of stimuli, the  
37 overall input to the two nodes is weak because learning has not begun. As a con-  
38 sequence, the activation peaks reached by the two nodes are too small to trigger  
39 the learning process when one node wins the competition. However, once the  
40 node associated with [step] comes to win the competition near the middle of the



**Fig. 10:** Activation of the nodes and total input in the first sequence of the simulation with an ID order. (a) y-axis represents the activation of the two nodes (dashed line: [sep] node; continuous line: [step] node). x-axis represents position in the sequence (bottom axis) and on the acoustic continuum (top axis). The vertical bold line indicates the middle of the sequence. (b) Bars represent the total external input reaching each of the two nodes at each stimulus location (white bars: input to the [sep] node; black bars: input to [step]).

sequence (where its most supporting stimuli are located), it has already won the competition several times in a row and, as a result of the additional contribution coming from residual activation, can reach higher peaks and trigger the learning mechanism. When the [step] node triggers learning in response to the presentation of stimulus  $i$ , it also changes its connection weights with respect to stimulus  $i+1$ ,  $i+2$ , etc., and  $i-1$ ,  $i-2$ , etc. This happens because learning causes an increase of the connection weights that are active when a stimulus  $i$  is presented. Thus, due to the bell-shaped activation pattern, the surrounding connection weights can benefit from learning as well. And, when stimulus  $i+1$  is presented, it can benefit from learning triggered by the preceding stimulus at position  $i$ , triggering learning yet again. This way, learning can spread across the acoustic continuum to nearby regions.

The effects of this mutual reinforcement between activation and learning can be observed in Figure 10b, which depicts the total amount of input that reaches the two nodes at the onset of each stimulus. The figure shows how the effect of learning spreads gradually from the [step] extreme of the continuum (presented

1 in the middle of the sequence) to the initially ambiguous stimuli (presented in the  
2 middle of the second half of the sequence).

3 The resulting bias is not reversed in the following sequences because, due to  
4 the accumulated learning, each stimulus produces an input signal strong enough  
5 to trigger learning, regardless of the position of the stimulus on the continuum.  
6 These initial differences change the system in a way that persists throughout the  
7 experiment. This is another hallmark of complex dynamical systems: they are  
8 sensitive to initial conditions (see, e.g., Kelso 1995).

9

10

## 11 5 General discussion

12

13 The results from our experiments strongly support a dynamical account of speech  
14 perception and the processing of phonological categories. We observed crucial  
15 phenomena that are consistently reported in the dynamical literature, such as  
16 hysteresis, critical slowing down, critical fluctuations, and sensitivity to initial  
17 conditions. Such phenomena support the hypothesis that nonlinear dynamics  
18 regulate the perception of phonological elements. For speech identification tasks,  
19 this idea was first explicitly tested by Tuller and her colleagues (1994), who  
20 modeled the perceptual system as an inseparable whole governed by a simple  
21 differential equation. These authors showed how the inherent complexity of the  
22 perceptual processes is reduced in the accomplishment of the identification task.  
23 In the present paper, we extended this model, showing how the patterns investi-  
24 gated by Tuller et al. (1994) emerge out of the nonlinear interactions of different  
25 components of the speech perception system. Our model shows how the short-  
26 term flexibility and long-term plasticity exhibited in speech identification experi-  
27 ments are regulated by the interaction of known processes (perceptual competi-  
28 tion, habituation and learning). As our model focuses on the interactions between  
29 different processes rather than on individual processes themselves, it can be  
30 characterized as an interaction-dominant system rather than a component-  
31 dominant system (see Ihlen and Vereijken 2010).

32 Our model is also “consilient” in the sense of Thagard (1978) and Wilson  
33 (1999), in that it explains different classes of facts: it is able to account for slower  
34 response times and increased proportion of flip-flops around the perceptual  
35 boundary, the shift of this perceptual boundary, as well as the slow shift from  
36 conservative to contrastive behavior throughout a long experiment. Moreover, it  
37 is able to account for some unexpected findings, namely that the very initial pre-  
38 sentation order has long-term effects that persist throughout the whole experi-  
39 ment. The fact that multiple response patterns that are observed with human  
40 participants can be explained with a simple unified model also shows that our

model meets the basic requirements of pattern-oriented modeling (Grimm et al. 1  
2005), thus assuring a high degree of generalizability. 2

Although the model proposed here is more detailed than a macroscopic 3  
model, as proposed by Tuller et al (1994), its architecture is still much simpler 4  
than the architectures of many other models of speech perception based on per- 5  
ceptual competition (e.g., McClelland and Elman 1986; or the family of ART net- 6  
works summarized by Grossberg 1978). We adopted such a simplified model in 7  
order to focus on the general principles governing the perceptual behaviors and 8  
on their interactions. Due to the generality of the processes regulating the behav- 9  
ior of the model, it is tempting to extend its explanatory power beyond the task 10  
discussed in this paper. For example, it can be shown that the model proposed, 11  
due to its habituation component, can reproduce the results of selective adapta- 12  
tion experiments (Eimas and Corbit 1973). In these experiments, a sequence of 13  
adapter stimuli which unambiguously support a particular percept habituate that 14  
category and lead to more responses to the alternative category. Vroomen, van 15  
Linden, de Gelder, and Bertelson (2007) show a shift from conservative to con- 16  
trastive behavior in selective adaptation paradigms (based on data from Samuel 17  
2001), similar to the shift observed in our experiment and in our model. Such re- 18  
sults fall straightforwardly out of our model, where weak competition dynamics 19  
produce conservative behavior, and where more contrastive behavior arises when 20  
perceptual dynamics are strengthened due to learning, triggering habituation. 21

The unexpected result of the initial presentation order bias might also shed 22  
light on some results by Tuller, Jantzen, and Jirsa (2008), as well as by Vallabha 23  
and McClelland (2007). These researchers explicitly looked at the dependence of 24  
learning on the initial configuration of the perceptual space at the beginning of 25  
the experiment. However, the initial perceptual behavior was not explicitly ma- 26  
nipulated but rather controlled post-hoc by correlating the outcomes of the learn- 27  
ing process with the initial listeners' perception of unfamiliar sounds. The results 28  
presented in the current paper constitute more compelling evidence for a causal 29  
relation between the initial perception of unfamiliar sounds and the outcomes of 30  
the learning process, showcasing the known sensitivity to initial conditions of 31  
dynamical systems. Moreover, we can account for this causal relation with our 32  
model (Section 4.5.1). The sensitivity to initial conditions emerges from the inter- 33  
action of bell-shaped activation patterns that come from the stimuli, and from 34  
residual activation and learning of the model. However, as this result was un- 35  
expected, and as our experiment was not designed to explicitly test this hypothe- 36  
sis, the finding needs to be verified with additional experiments. 37

38

39

40

## 6 Conclusions

Differences in the temporal evolution of speech perception, learning, and habituation lead us to conceive their underlying processes as distinct. However in our model, these processes interact with each other and constrain each other. We thus consider perceptual competition, learning, and habituation in a unified framework, in line with viewing language and cognition as interaction-dominant systems. In this way, causal relations among processes occurring on different time scales can be accounted for by modeling.

We showed that the interaction among perception, learning, and habituation gives rise to different phases of behavior. In a first phase, perceptual dynamics are characterized by bistability on a wide region of the continuum, where the attractors for the two perceptual categories are relatively weak. Perceptual learning then strengthens the system's attractors, resulting in more stable and faster performances due to a reduction of the bistable region of the continuum. In its early phase, the learning process heavily depends on contextual factors, and changing the stimulus presentation order can affect the direction of learning, as shown by the initial presentation order bias. As learning proceeds, responses first become less dependent on their context, and then context dependency starts to increase again, this time with a contrastive nature.

This pattern shows how the perception of phonological elements changes in a principled manner. At the beginning, behavior is rapidly stabilized, and there is a tendency to respond in the same way despite variations in stimuli acoustics (= hysteresis). In the long run, perseverance to particular responses potentially loses its utility because it creates the conditions for the system to get stuck in the behavior already produced. But before this happens, a new form of flexibility arises (= contrastive behavior), and this helps the system to respond to small differences in stimuli (= increased perceptual sensitivity). This highlights how speech perception is adaptive at multiple time-scales, assuring a high degree of robustness (cf. Winter and Christiansen 2012). The system is characterized by flexibility due to rapid perceptual competition. But, due to initial perseverance, this flexibility does not lead to instability. Larger changes that lead to different regimes of response patterns happen more slowly. This plasticity allows speech perception to cope on a long-term basis with the massive amount of variation that characterizes the way in which phonological systems are reflected in everyday vocal communication.

37  
38  
39  
40

## Acknowledgments

The authors wish to thank Noel Nguyen and Betty Tuller for their contribution to the design of the experiments. We thank Susanne Fuchs, Caterina Petrone, Amelie Rochet-Capellan, and Pascal Perrier for feedback and suggestions. We want to thank Martine Grice, Doris Mücke, and the anonymous reviewers for helpful comments and suggestions, and Kayla Hewitt for proofreading. This work was partly supported by the ACI Systèmes complexes en SHS Research Program (CNRS & French Ministry of Research).

## References

- Baayen, Harald. 2011. languageR: Data sets and functions with “Analyzing Linguistic Data: A practical introduction to statistics”. R package version 1.4.
- Bates, Douglas, Martin Maechler, & Benjamin Bolker. 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.
- Best, Catherine, Barbara Morrongiello, & Rick Robson. 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics* 29(3). 191–211.
- Bogacz, Rafal, Eric Brown, Jeff Moehlis, Philip Holmes, & Jonathan Cohen. 2006. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review* 113(4). 700–765.
- Bradlow, Ann, & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106(2). 707–729.
- Carpenter, Gail, & Stephen Grossberg. 1987. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing* 37(1). 54–115.
- Case, Pamela, Betty Tuller, Mingzhou Ding, & Scott Kelso. 1995. Evaluation of a dynamical model of speech categorization. *Perception & Psychophysics* 57(7). 977–988.
- Coleman, John, & Andrew Slater. 2001. Estimation of parameters for the Klatt formant synthesizer. In Robert Dampier (ed.), *Data mining techniques in speech synthesis*, 215–238. Boston, USA: Kluwer.
- Eimas, Peter, & John Corbit. 1973. Selective adaptation of linguistic feature detectors. *Cognitive Psychology* 4(9). 99–109.
- Eisner, Frank, & James McQueen. 2006. Perceptual learning in speech: Stability over time. *Journal of the Acoustic Society of America* 119(4). 1950–1953.
- Grimm, Volker, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf Mooij, Steven Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, & Donald DeAngelis. 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from Ecology. *Science* 310(11). 987–991.
- Grossberg, Stephen. 1973. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics* 52(3). 213–257.

- 1 Grossberg, Stephen. 1976. Adaptive pattern classification and universal recoding: I. Parallel  
2 development and coding of neural feature detectors. *Biological Cybernetics* 23(3).  
3 121–134.
- 4 Grossberg, S. 1978. A theory of human memory: Self-organization and performance of  
5 sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (eds.), *Progress in*  
6 *Theoretical Biology*, Vol. 5, 233–374. New York, USA: Academic Press.
- 7 Grossberg, Stephen, & Christopher Myers. 2000. The resonant dynamics of speech perception:  
8 Inter word integration and duration-dependent backward effects. *Psychological Review*  
9 107(4). 735–767.
- 10 Grossberg, Stephen, & Sohrob Kazerounian. 2011. Laminar cortical dynamics of conscious  
11 speech perception: A neural model of phonemic restoration using subsequent context in  
12 noise. *Journal of the Acoustical Society of America* 130(1). 440–460.
- 13 Haken, Hermann. 1983. *Synergetics. An Introduction. Non-Equilibrium Phase Transitions and*  
14 *Self Organization in Physics, Chemistry and Biology*. Berlin, Germany: Springer-Verlag.
- 15 Ihlen, Espen, & Beatrix Vereijken. 2010. Interaction-dominant dynamics in human cognition:  
16 Beyond  $1/f\alpha$  fluctuation. *Journal of Experimental Psychology: General* 139(3). 436–463.
- 17 Jensen, Arthur. 2006. *Clocking the Mind: Mental Chronometry and Individual Differences*.  
18 Amsterdam, The Netherlands: Elsevier Science Ltd.
- 19 Hock, Howard, Gregor Schöner, & Martin Giese. 2003. The dynamical foundations of motion  
20 pattern formation: Stability, selective adaptation, and perceptual continuity. *Perception &*  
21 *Psychophysics* 65(3). 429–457.
- 22 Kawamoto, Alan, & James Anderson. 1985. A neural network model of multistable perception.  
23 *Acta Psychologica* 59(1). 35–65.
- 24 Kelso, Scott. 1995. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge,  
25 MA: MIT Press.
- 26 Kelso, Scott, John Scholz, & Gregor Schöner. 1986. Non equilibrium phase transitions in  
27 coordinated biological motion: critical fluctuations. *Physics Letters A* 118(6). 279–284.
- 28 Klatt, Dennis, & Laura Klatt. 1990. Design for a cascade-parallel formant synthesizer. *Journal of*  
29 *Acoustical Society of America* 87(2). 820–857.
- 30 Kraljic, Tanya, & Arthur Samuel. 2005. Perceptual learning for speech: Is there a return to  
31 normal? *Cognitive Psychology* 51(2). 141–178.
- 32 Liberman, Alvin, Franklin Cooper, Donald Shankweiler, & Michael Studdert-Kennedy. 1967.  
33 Perception of speech code. *Psychological Review* 74(6). 431–461.
- 34 Liberman, Alvin, Katherine Harris, Howard Hoffman, & Belver Griffith. 1957. The discrimination  
35 of speech sounds within and across phoneme boundaries. *Journal of Experimental*  
36 *Psychology* 54(5). 358–368.
- 37 Liberman, Alvin M., & Ignatius Mattingly. 1985. The motor theory of speech perception revised.  
38 *Cognition* 21(1). 1–36.
- 39 Massaro, Dominic, & Michael Cohen. 1983. Evaluation and integration of visual and auditory  
40 information in speech perception. *Journal of Experimental Psychology: Human Perception*  
*and Performance* 9(5). 753–771.
- Maye, Jessica, Richard Aslin, & Michael Tanenhaus. 2008. The weckud wetch of the wast:  
Lexical adaptation to a novel accent. *Cognitive Science* 32(3). 543–562.
- McClelland, James, & Jeffrey Elman. (1986). The TRACE model of speech perception. *Cognitive*  
*Psychology* 18(1). 1–86.

- McClelland, James, & Gautam Vallabha. 2009. Connectionist models of development: Mechanistic dynamical models with emergent dynamical properties. In John Spencer, Michael Thomas, & James McClelland (eds.), *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered*, 3–24. New York, USA: Oxford Scholarship Online Monographs. 1
- Moulines, Eric, & Francis Charpentier. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9(5–6). 453–467. 2
- Nguyen, Noel, Leonardo Lancia, Maïtine Bergounioux, Sophie Wauquier-Gravelines, & Betty Tuller. 2005. Role of training and short-term context effects in the perception of /s/ and /st/ in French. In Valerie Hazan & Paul Iverson (eds.), *ISCA workshop on Plasticity in Speech Perception*, A38–39. London, UK. 3
- Pisoni, David, & Jeffrey Tash. 1974. Reaction times to comparisons within and across phonemic categories. *Perception & Psychophysics* 15(2). 285–290. 4
- Pitt, Mark, Woojae Kim, Daniel Navarro, & Jay Myung. 2006. Global model analysis by parameter space partitioning. *Psychological Review* 113(1). 57–83. 5
- R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 6
- Repp, Bruno. 1980. A range-frequency effect on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research* 61. 151–165. 7
- Repp, Bruno. 1981. Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society* 18(1). 12–14. 8
- Repp, Bruno, & Alvin Liberman. 1990. Phonetic category boundaries are flexible. In S. R. Harnad (ed.), *Categorical perception: The groundwork of cognition*, 89–112. Cambridge, England: Cambridge University Press. 9
- Sakoe, Hiroaki, & Seibi Chiba. 1978. Dynamic Programming algorithm optimization for spoken word recognition. *IEEE Transactions of Acoustics, Speech, and Signal Processing ASSP-26*(1). 43–49. 10
- Samuel, Arthur. 2001. Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science* 12(4). 348–351. 11
- Samuel, Arthur, & Tanya Kraljic. 2009. Perceptual learning for speech. *Attention, Perception & Psychophysics* 71(6). 1207–1218. 12
- Schmidt, Richard, Michelle Bienvenu, Paula Fitzpatrick, & Polemnia Amazeen. 1998. A comparison of intra- and interpersonal interlimb coordination – Coordination breakdowns and coupling strengths. *Journal of Experimental Psychology: Human Perception and Performance* 24(3). 884–900. 13
- Smolensky, Paul, & Géraldine Legendre. 2006. Foundational implications of the ICS Architecture: Unification in cognitive science. In Paul Smolensky & Géraldine Legendre (eds.), *The Harmonic Mind*, Vol. 1, 100–121. Cambridge, MA: MIT Press. 14
- Spivey, Michael. 2007. *The Continuity of Mind*. Oxford, UK: Oxford University Press. 15
- Stevens, Kenneth. 2005. Features in speech perception and lexical access. In David Pisoni & Robert Remez (eds.), *The Handbook of Speech Perception*, 124–155. Oxford, UK: Blackwell. 16
- Studdert-Kennedy, Michael, Alvin Liberman, & Kenneth Stevens. 1963. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America* 35(11). 1900 (Abstract). 17
- Thagard, Paul. 1978. The best explanation: Criteria for theory choice. *The Journal of Philosophy* 75(2). 76–92. 18

- 1 Thelen, Esther, & Linda Smith. 2006. Dynamical systems theories. In Richard Lerner & Damon  
2 William (eds.), *Handbook of Child Psychology: Theoretical Models of Human Development*,  
3 258–312. Hoboken, NJ: John Wiley & Sons.
- 4 Tuller, Betty, Pamela Case, Mingzhou Ding, & Scott Kelso. 1994. The nonlinear dynamics of  
5 speech categorization. *Journal of Experimental Psychology: Human Perception and  
6 Performance* 20(1). 3–16.
- 7 Tuller, Betty. 2005. Categorization and learning in speech perception as a dynamical process. In  
8 Michael Riley & Guy van Orden (eds.), *Tutorials in Contemporary Nonlinear Methods for the  
9 Behavioral Sciences*. Arlington, VA, USA: National Science Foundation.
- 10 Tuller, Betty, McNeel Jantzen, & Victor Jirsa. 2008. A dynamical approach to speech  
11 categorization: Two routes to learning. *New Ideas in Psychology* 26(2). 208–226.
- 12 Ulrich, Rolf, & Jeff Miller. 1994. Effects of truncation on reaction time analysis. *Journal of  
13 Experimental Psychology: General* 123(1). 34–80.
- 14 Vallabha, Gautam, & James McClelland. 2007. Success and failure of new speech category  
15 learning in adulthood: Consequences of learned Hebbian attractors in topographic maps.  
16 *Cognitive, Affective, & Behavioral Neuroscience* 7(1). 53–73.
- 17 Vroomen, Jean., Sabine van Linden, Béatrice de Gelder, & Paul Bertelson. 2007. Visual  
18 recalibration and selective adaptation in auditory–visual speech perception: Contrasting  
19 build-up courses. *Neuropsychologia* 45(3). 572–577.
- 20 Wilson, Edward. 1999. *Consilience: The Unity of Knowledge*. New York: Knopf.
- 21 Winter, Bodo, & Morten Christiansen. 2012. Robustness as a design feature of speech  
22 communication. In Thomas Scott-Phillips, Mónica Tamariz, Erica Cartmill, & James Hurford  
23 (eds.), *Proceedings of the 9th International Conference on the Evolution of Language*,  
24 384–391. New Jersey: World Scientific.
- 25 Winters, Stephen, & David Pisoni. 2004. Perception and comprehension of synthetic speech. In  
26 *Research on Spoken Language Processing Report No. 26*, 95–138. Bloomington: Indiana  
27 University, Speech Research Laboratory, Department of Psychology.
- 28 Zanone, Pier-Giorgio, & Scott Kelso. 1992. The evolution of behavioral attractors with learning:  
29 Nonequilibrium phase transitions. *Journal of Experimental Psychology: Human Perception  
30 and Performance* 18(2). 403–421.

## 31 Appendix A

### 32 1 Model equations

33  
34 The equation which governs the activity of the two nodes was introduced by  
35 Grossberg (1973). It is given by:

$$36 \dot{y}_j = \psi_j \left\{ -\alpha y_j + (1 - y_j) \left[ \beta f(y_j) + \sum_{i=1}^n w_{i,j} I_i z_j \right] - y_j \left[ c \sum_{l \neq j} f(y_l) \right] \right\} \quad (A1)$$

37  
38  
39  
40

$y_j$  is the activity of a node, and  $\dot{y}_j$  the time derivative of that activity. When  $\dot{y}_j$  is positive, the activity  $y_j$  increases with a rate equal to  $\dot{y}_j$ . When  $\dot{y}_j$  is negative,  $y_j$  decreases.  $\psi_1$  is a rate coefficient regulating the global rate of the competition process (equal for both nodes). The dynamics of the nodes are defined by the three additive terms inside the curly brackets. The first additive term ( $-\alpha y_j$ ) represents passive leakage.

The second additive term is the total positive input arriving to the node. This input is multiplied by a “shunting term” ( $1 - y_j$ ) to keep  $y_j$  lower than 1. The total positive input itself is given by the sum of two terms. The first term is the activation value  $y_j$  transformed by the nonlinear sigmoid function  $f$  (described below) and multiplied by a free parameter ( $\beta$ ). This term determines the influence of the recurrent signal on the activation. The term  $\sum_{i=1}^n w_{i,j} I_i$  is the sum of the weighted activations coming from the input layer.  $I_i$  is the  $i$ th element of the input layer, and  $w_{i,j}$  is the weight of the connection from  $I_i$  to  $y_j$ . The effect of the external input on a given node  $y_j$  is modulated by the value of the efficiency variable  $z_j$ , which itself is changing dependent on the system state (see description below).

The third and final additive term in the equation ( $c \sum_{l \neq j} f(y_l)$ ) represents the inhibitory signal from the competitor node. Here, the signal strength depends on a sigmoidized value of the activation of the competitor node multiplied by the free parameter  $c$  which determines the influence of the incoming inhibitory signals onto the activation dynamics. The whole third additive term is multiplied by the value of  $y_j$ , which here has a shunting function constraining  $y_j$  to positive values.

The sigmoid function  $f$  is given by

$$f(y) = \frac{1}{1 + e^{-\rho(10y - \sigma)}} \quad (\text{A2})$$

where the parameters  $\rho$  and  $\sigma$  determine the slope and the horizontal shift of  $f$ . The other source of nonlinearity in the main Equation (A1) is represented by the shunting terms ( $1 - y_j$ ) and  $y_j$ . The choice of competition dynamics bounded by shunting terms was motivated by the normalization properties of these equations (cf. Grossberg 1973). Since the total activity over the nodes does not depend on the number of nodes, our model can in principle be extended, e.g., to incorporate multiple responses.

The efficiency variable  $z_j$  varies between 0 ( $= y_j$  is not affected by habituation at all) and 1 ( $= y_j$  is completely habituated and its activation cannot grow even in the presence of a positive input).

$$\dot{z}_j = \psi_2[-\mu f(y_j) + \varepsilon] \quad (\text{A3})$$

1  $\varphi_2$  regulates the overall rate of the habituation process,  $\mu$  determines the rate  
 2 at which the efficiency variable decreases when  $y_j$  is active, and  $\varepsilon$  regulates the  
 3 rate at which the efficiency variable  $z_j$  recovers when  $y_j$  is inactive. The habitua-  
 4 tion process is driven by the sigmoid function  $f$  of the value of  $y_j$ .

5 The following learning law regulates the weights of the connections between  
 6 the component  $I_i$  of the input layer and the node  $y_j$ :

$$\dot{w}_{i,j} = \psi_3 \left[ \left( \varphi - \sum_{k=1}^m w_{i,k} \right) f(y_j) I_i \right] \tag{A4}$$

7  
 8  
 9  
 10  
 11  
 12  
 13 This equation states that the speed of change of  $w_{i,j}$  depends on the product  
 14  $f(y_j)I_i$ . When the activation  $y_j$  is small, this activation is minimized by the sigmoid  
 15 function  $f$ , resulting in a negligible product (= no learning). When  $y_j$  reaches high  
 16 values,  $f$  outputs high values and learning occurs.  $\psi_3$  regulates the overall learn-  
 17 ing rate,  $m = 2$  is the number of the competing nodes, and the term  $(\varphi - \sum_{k=1}^m w_{i,k})$   
 18 defines an upper threshold (equal to  $\varphi$ ) to the total amount of the weights on the  
 19 connections starting from the  $i$ th component of the input layer. The presence of  
 20 this threshold implements in the simplest possible way the idea that learning re-  
 21 sources are limited (Carpenter and Grossberg 1987).

22  
 23  
 24 **2 Stimuli**

25  
 26 When a stimulus located at position  $\omega$  on the continuum is presented to the sys-  
 27 tem, the activation of the  $i$ th element of the input layer is given by:

$$I(i, \omega) = e^{-\frac{(\omega-i)^2}{v}}; \quad i \in \{1 \dots n\} \tag{A5}$$

28  
 29  
 30  
 31  
 32  
 33 Here,  $n$  is the number of the components of the input layer  $I$ , and  $v = 4$  is a  
 34 constant determining the narrowing and the height of the activation bump. In  
 35 order to keep the norm of the input vector constant across the stimuli of the con-  
 36 tinuum, each component of the input vector is divided by  $\sum_{\omega=1}^{21} I(i, \omega)$ .

37 The duration of the stimuli, the ISI, and the duration of the silent interval  
 38 between a sequence and the following one were matched to the values adopted in  
 39 the second experiment of this paper. A single time step (one iteration of the equa-  
 40 tion) corresponds to 20 ms.

### 3 Parameter settings

All simulations are based on the same set of parameters:  $\psi_1 = 0.036$ ,  $\alpha = 0.95$ ,  $\beta = 0.25$ ,  $c = 2.5$ ,  $\psi_2 = 0.01$ ,  $\mu = 0.14$ ,  $\varepsilon = 0.03$ ,  $\psi_3 = 0.01$ ,  $\phi = 15$ . The two parameters of the sigmoid function  $f$  (Equation A2) are set to  $\rho = 7$  and  $\sigma = 1.8$  for Equation A1, to  $\rho = 7.5$  and  $\sigma = 8$  for Equation (A3), and to  $\rho = 11$  and  $\sigma = 5.2$  for Equation A4.

## Appendix B Construction of stimuli

### 1 Stimuli for Experiment 1

We recorded five productions of [sep] and five productions of [step] from a native speaker of French in a soundproof booth. We measured a number of acoustic variables and chose the most prototypical token of each category by looking at the distance for each measured variable from the median of all five productions. We considered the durations of the fricative, the coronal closure, the bilabial closure, and the vowel, as well as the ratios between the intensity peak of each consonant<sup>6</sup> and the intensity peak of the vowel. The acoustic material for building the consonantal part of the stimuli was extracted from the selected [sep] token. This signal was then modified so that the value of each relevant acoustic parameter corresponded to the average of the median value of [step] and the median value of [sep]. Durations were modified using the PSOLA algorithm (Moulines and Charpentier 1990).

To synthesize the vowel we measured the first four formants of the vowel, its fundamental frequency ( $F_0$ ), and its intensity at four different points in time (the beginning and end of the vowel, the vowel center, and the point where the derivative of the concerned trajectory changes sign). We averaged over the measured trajectory parameters of both words to obtain values for a vowel that should sound midway between the vowels from [sep] and [step], which we subsequently produced using the Klatt synthesizer (Klatt and Klatt 1990).

### 2 Stimuli for Experiment 2

The materials that were used to construct the stimuli comprised 10 tokens of each word; five tokens were produced within the carrier sentence: *on dira [word] en-*

<sup>6</sup> To compute the intensity of the consonants, we first extracted the aperiodic component of the signal following the method illustrated by Coleman and Slater (2001). Then we computed the intensity of the aperiodic component using a sliding window of 12.5 ms. Finally, we selected the peak intensity value in the portion of signal that corresponds to the consonant.

1 *core* (“I am going to say [word] again”), and five were produced in isolation. Static  
2 parameters (e.g., duration values) were considered separately from the trajecto-  
3 ries (represented by time series such as in the case of  $F_0$  curves). The first group of  
4 parameters included the fricative, coronal closure, vowel, and bilabial closure  
5 duration, as well as the ratio between the intensity peak of the vowel to each con-  
6 sonant. We then considered the trajectories of the intrinsic  $F_0$  of the vowel, and  
7 the first six formants and bandwidths.

8 For each of the two words, we created a vector of the median values of the  
9 static parameters observed across the tokens. We then selected the token of each  
10 word whose vector of parameters was closest to the vector of the median values  
11 for that word. We extracted the acoustic material needed to build the consonantal  
12 part of the stimuli from the selected token of the word [sep]. As for Experiment 1,  
13 we modified the extracted signal to sound midway between [sep] and [step] by  
14 modifying the static parameter values. The set of prototypic trajectories needed to  
15 synthesize the vowel was deduced separately for each of the two words using a  
16 method proposed by Coleman and Slater (2001, to which the reader is referred for  
17 details). For each acoustic parameter, the trajectories from the different tokens  
18 were compared via dynamic time warping (Sakoe and Chiba 1978) to the median  
19 trajectory of all tokens. We chose the token which was most similar to the set of  
20 median trajectories as a prototype for the stimuli trajectories. Via this process, we  
21 obtained two sets of trajectories which defined the prototypes of the vowels of the  
22 two words. Each one of these prototypes was used as a basis for one continuum.  
23 The values of the static parameters and the intensity contour of the fricative were  
24 fixed across the stimuli. The intensity contour for the fricative was obtained by  
25 averaging the intensity contours of the two fricatives corresponding to the two  
26 selected vectors of static parameters.

27 The two vowels were synthesized at 16,000 Hz by activating the resonances  
28 corresponding to the first eight formants of the Klatt synthesizer. The acoustic  
29 material necessary for the construction of [s] and [p] was extracted from the token  
30 of the word [sep] corresponding to the prototypical vector of static parameters.  
31 For the construction of the fricative, only the stable portion of the signal was con-  
32 sidered and its duration was modified using PSOLA. The intensity contour was  
33 then substituted with the average contour described above. The duration of the  
34 last plosive closure was changed according to the modeled value. The intensity  
35 peaks of the two consonants were changed to match the desired ratio between the  
36 intensity of each consonant and the intensity of the vowel. Concatenating the  
37 acoustic chunks obtained, we produced two acoustic signals which sounded like  
38 *cèpe* or *steppe* depending on the duration of the closure duration.

39

40

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40