DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

# Time-Warp–Invariant Neuronal Processing

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| | |
|---|---|
| Citation | Gütig, Robert, Haim Sompolinsky, and Michael Robert DeWeese. 2009. Time-Warp–Invariant Neuronal Processing. PLoS Biology 7: e1000141. |
| Published Version | doi: 10.1371/journal.pbio.1000141 |
| Citable link | http://nrs.harvard.edu/urn-3:HUL.InstRepos:4551666 |
| Terms of Use | This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA |

# Time-Warp–Invariant Neuronal Processing

**Robert Gütig**[1,2]*, **Haim Sompolinsky**[1,2,3]

**1** Racah Institute of Physics, Hebrew University, Jerusalem, Israel, **2** Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel, **3** Center for Brain Science, Harvard University, Cambridge, Massachusetts, United States of America

## Abstract

Fluctuations in the temporal durations of sensory signals constitute a major source of variability within natural stimulus ensembles. The neuronal mechanisms through which sensory systems can stabilize perception against such fluctuations are largely unknown. An intriguing instantiation of such robustness occurs in human speech perception, which relies critically on temporal acoustic cues that are embedded in signals with highly variable duration. Across different instances of natural speech, auditory cues can undergo temporal warping that ranges from 2-fold compression to 2-fold dilation without significant perceptual impairment. Here, we report that time-warp–invariant neuronal processing can be subserved by the shunting action of synaptic conductances that automatically rescales the effective integration time of postsynaptic neurons. We propose a novel spike-based learning rule for synaptic conductances that adjusts the degree of synaptic shunting to the temporal processing requirements of a given task. Applying this general biophysical mechanism to the example of speech processing, we propose a neuronal network model for time-warp–invariant word discrimination and demonstrate its excellent performance on a standard benchmark speech-recognition task. Our results demonstrate the important functional role of synaptic conductances in spike-based neuronal information processing and learning. The biophysics of temporal integration at neuronal membranes can endow sensory pathways with powerful time-warp–invariant computational capabilities.

**Citation:** Gütig R, Sompolinsky H (2009) Time-Warp–Invariant Neuronal Processing. PLoS Biol 7(7): e1000141. doi:10.1371/journal.pbio.1000141

**Academic Editor:** Michael Robert DeWeese, UC Berkeley, United States of America

**Received** August 13, 2008; **Accepted** May 18, 2009; **Published** July 7, 2009

**Copyright:** © 2009 Gütig, Sompolinsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** RG was funded through fellowships from the Minerva Foundation (Hans-Jensen Fellowship, www.minerva.mpg.de) and the German Science Foundation (GU 605/2-1, www.dfg.de). HS received funding through the Israeli Science Foundation (www.isf.org.il) and MAFAT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The Hebrew University technology transfer unit (Yissum) has filed the learning rule described in this work and its application to speech processing for patent.

* E-mail: guetig@cc.huji.ac.il

## Introduction

Robustness of neuronal information processing to temporal warping of natural stimuli poses a difficult computational challenge to the brain [1–9]. This is particularly true for auditory stimuli, which often carry perceptually relevant information in fine differences between temporal cues [10,11]. For instance in speech, perceptual discriminations between consonants often rely on differences in voice onset times, burst durations, or durations of spectral transitions [12,13]. A striking feature of human performance on such tasks is that it is resilient to a large temporal variability in the absolute timing of these cues. Specifically, changes in speaking rate in ongoing natural speech introduce temporal warping of the acoustic signal on a scale of hundreds of milliseconds, encompassing temporal distortions of acoustic cues that range from 2-fold compression to 2-fold dilation [14,15]. Figure 1 shows examples of time warp in natural speech. The utterance of the word "one" in (A) is compressed by nearly a factor of one-half relative to the utterance shown in (B), causing a concomitant compression in the duration of prominent spectral features, such as the transitions of the peaks in the frequency spectra. Notably, the pattern of temporal warping in speech can vary within a single utterance on a scale of hundreds of milliseconds. For example, the local time warp of the word "eight" in (C) relative to (D), reverses from compression in the initial and final segments to strong dilation of the gap between them. Although it has long been demonstrated that speech perception in humans normalizes durations of temporal cues to the rate of speech [2,16–18], the neural mechanisms underlying this perceptual constancy have remained mysterious.

A general solution of the time-warp problem is to undo stimulus rate variations by comodulating the internal "perceptual" clock of a sensory processing system. This clock should run slowly when the rate of the incoming signal is low and embedded temporal cues are dilated, but accelerate when the rate is fast and the temporal cues are compressed. Here, we propose a neural implementation of this solution, exploiting a basic biophysical property of synaptic inputs, namely, that in addition to charging the postsynaptic neuronal membrane, synaptic conductances modulate its effective time constant. To utilize this mechanism for time-warp robust information processing in the context of a particular perceptual task, synaptic peak conductances at the site of temporal cue integration need to be adjusted to match the range of incoming spike rates. We show that such adjustments can be achieved by a novel conductance-based supervised learning rule. We first demonstrate the computational power of the proposed mechanism by testing our neuron model on a synthetic instantiation of a generic time-warp–invariant neuronal computation, namely, time-warp–invariant classification of random spike latency patterns. We then present a novel neuronal network model for word recognition and show that it yields excellent performance on a benchmark speech-recognition task, comparable to that achieved by highly elaborate, biologically implausible state-of-the-art speech-recognition algorithms.

## Author Summary

The brain has a robust ability to process sensory stimuli, even when those stimuli are warped in time. The most prominent example of such perceptual robustness occurs in speech communication. Rates of speech can be highly variable both within and across speakers, yet our perceptions of words remain stable. The neuronal mechanisms that subserve invariance to time warping without compromising our ability to discriminate between fine temporal cues have puzzled neuroscientists for several decades. Here, we describe a cellular process whereby auditory neurons recalibrate, on the fly, their perceptual clocks and allows them effectively to correct for temporal fluctuations in the rate of incoming sensory events. We demonstrate that this basic biophysical mechanism allows simple neural architectures to solve a standard benchmark speech-recognition task with near perfect performance. This proposed mechanism for time-warp–invariant neural processing leads to novel hypotheses about the origin of speech perception pathologies.

## Results

### Time Rescaling in Neuronal Circuits

Whereas the net current flow into a neuron is determined by the balance between excitatory and inhibitory synaptic inputs, both types of inputs increase the total synaptic conductance, which in turn modulates the effective integration time of the postsynaptic cell [19–21] (an effect known as synaptic shunting). Specifically, when the total synaptic conductance of a neuron is large relative to the resting conductance (leak) and is generated by linear summation of incoming synaptic events, the neuron's effective integration time scales inversely to the rate of inputs spikes. Hence, the shunting action of synaptic conductances can counter variations in afferent spike rates by automatically rescaling the effective integration time of the postsynaptic neuron.

We implement this mechanism in a leaky integrate-and-fire model neuron driven by $N$ exponentially decaying synaptic conductances $g_i(t)=g_i^{\max}\exp(-t/\tau_s)(i=1,\ldots,N)$. Here, $g_i^{\max}$ denotes the peak conductance of the $i$th synapse in units of $sec^{-1}$, and $\tau_s$ is the synaptic time constant. The total synaptic current, measured at rest, is given by

$$I_{\rm syn}(t,\beta)=\sum_{i=1}^{N}\sum_{t_i<t}V_i^{\rm rev}g_i(t-\beta t_i)$$

where $V_i^{\rm rev}$ denotes the reversal potential of the $i$th synapse relative to resting potential and $t_i$ denote the arrival times of the spikes of the $i$th afferent. The factor $\beta$ denotes a global scaling of all incoming spike times; $\beta=1$ is the unwarped inputs. The total synaptic conductance, $G_{\rm syn}(t,\beta)$, is

$$G_{\rm syn}(t,\beta)=\sum_{i=1}^{N}\sum_{t_i<t}g_i(t-\beta t_i).$$

For fast synapses, the total synaptic current is essentially a train of pulses, each of which occurs at the time of an incoming spike and delivers a total charge of $g_i\tau_s V_i^{\rm rev}$. Changing the rate of the incoming spikes will induce a corresponding change in the timing of these pulses but not their charge. Therefore, ignoring the effect of time warp on the time scale of $\tau_s$, which is short relative to the time scale of voltage modulations, the total synaptic current obeys the following time-warp scaling relation, $I_{\rm syn}(\beta t,\beta)=\beta^{-1}I_{\rm syn}(t,1)$. A similar scaling relation holds for the total synaptic conductance. The evolution in time of the subthreshold voltage is given by

$$\frac{\rm d}{{\rm d}t}V(t,\beta)=-V(t,\beta)\left(g_{\rm leak}+G_{\rm syn}(t,\beta)\right)+I_{\rm syn}(t,\beta). \quad (1)$$

Thus, $V$ integrates the synaptic current with an effective time constant whose inverse is $1/\tau_{\rm eff}=g_{\rm leak}+G_{\rm syn}(t,\beta)$. If the contribution of $G_{\rm syn}$ is significantly larger than the leak conductance, then $1/\tau_{\rm eff}$ is rescaled by time-warp similar to $G_{\rm syn}$ and $I_{\rm syn}$, and, hence, the solution of Equation 1 is approximately time-warp invariant, namely, $V(\beta t,\beta)=V(t,1)$. This result is illustrated in Figure 2, which compares the voltage traces induced by a random spike pattern for $\beta=1$ and $\beta=0.5$.

To perform time-warp–invariant tasks, peak synaptic conductances must be in the range of values appropriate for the statistics of the stimulus ensemble of the given task. To achieve this, we have devised a novel spike-based learning rule for synaptic conductances, the conductance-based tempotron. This model neuron learns to discriminate between two classes of spatiotemporal input spike patterns. The tempotron's classification rule requires it to fire at least one spike in response to each of its target stimuli but to remain silent when driven by a stimulus from the null class. Spike patterns from both classes are iteratively presented to the neuron, and peak synaptic conductances are modified after each error trial by an amount proportional to their contribution to the maximum value of the postsynaptic potential over time (see Materials and Methods). This contribution is sensitive to the time courses of the total conductance and voltage of the postsynaptic neuron. Therefore, the conductance-based tempotron learns to adjust, not only the magnitude of the synaptic inputs, but also its effective integration time to the statistics of the task at hand.

### Learning to Classify Time-Warped Latency Patterns

We first quantified the time-warp robustness of the conductance-based tempotron on a synthetic discrimination task. We randomly assigned 1,250 spike pattern templates to target and null classes. The templates consisted of 500 afferents, each firing once at a fixed time chosen randomly from a uniform distribution between 0 and 500 ms. Upon each presentation during training and testing, the templates underwent global temporal warping by a random factor $\beta$ ranging from compression by $1/\beta_{\max}$ to dilation by $\beta_{\max}$ (see Materials and Methods). Consistent with the psychophysical range, $\beta_{\max}$ was varied between 1 and 2.5. Remarkably, with physiologically plausible parameters, the error frequency remained almost zero up to $\beta_{\max}\approx 2$ (Figure 3A, blue curve). Importantly, the performance of the conductance-based tempotron showed little change when the temporal warping applied to the spike templates was dynamic (see Materials and Methods) (Figure 3A). The time-warp robustness of the neural classification depends on the resting membrane time constant $\tau_m$ and the synaptic time constant $\tau_s$. Increases in $\tau_m$ or decreases in $\tau_s$ both enhance the dominance of shunting in governing the cell's effective time constant. As a result, the performance for $\beta_{\max}=2.5$ improved with increasing $\tau_m$ (Figure 3B, left) and decreasing $\tau_s$ (Figure 3B, right). The time-warp robustness of the conductance-based tempotron was also reflected in the shape of its subthreshold voltage traces (Figure 3C, top row) and generalized to novel spike templates with the same input statistics that were not used during training (Figure 3C, second row).

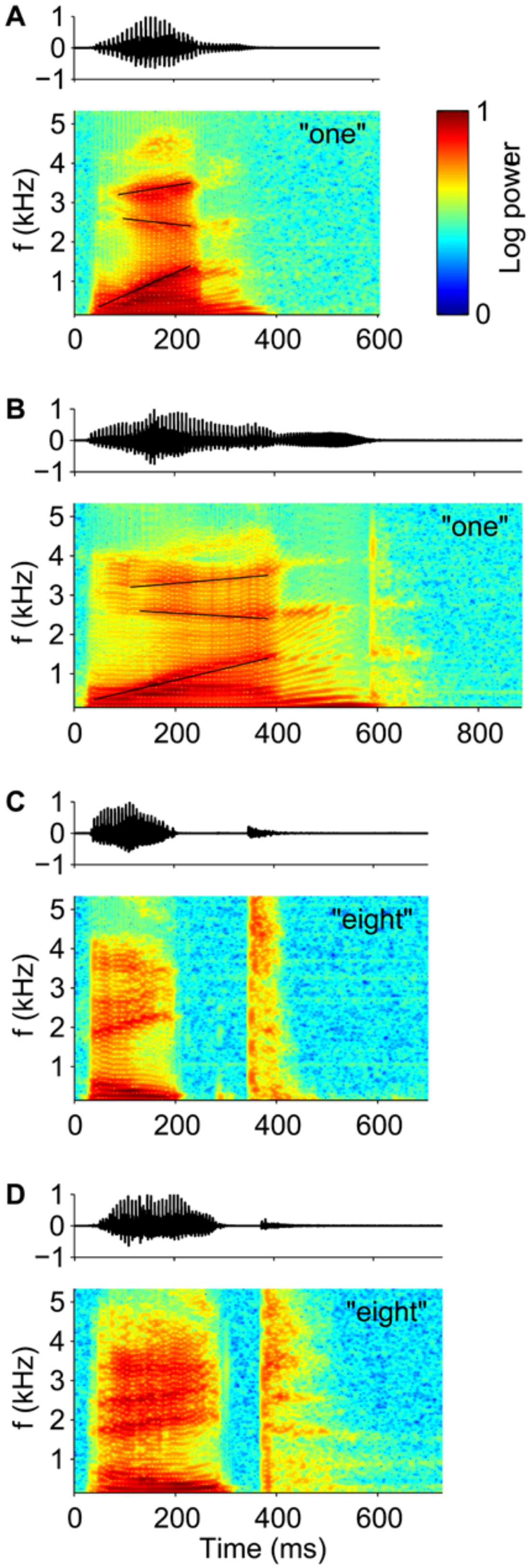

**Figure 1. Time warp in natural speech.** Sound pressure waveforms (upper panels, arbitrary units) and spectrograms (lower panels, color-code scaled between the minimum and maximum log power) of speech samples from the TI46 Word corpus [24], spoken by different male speakers. (A and B) Utterances of the word "one." Thin black lines highlight the transients of the second, third, and fourth (bottom to top) spectral peaks (formants). The lines in (A) are compressed relative to (B) by a common factor of 0.53. (C and D) Utterances of the word "eight."
doi:10.1371/journal.pbio.1000141.g001

Synaptic conductances were crucial in generating the neuron's robustness to temporal warping. Athough an analogous neuron model with a fixed integration time, the current-based tempotron [22] (see Materials and Methods) also performed the task perfectly in the absence of time-warp ($\beta_{\rm max}=1$); its error frequency was sensitive even to modest temporal warping and deteriorated further when the applied time warp was dynamic (Figure 3A, red curve). Similarly, the voltage traces of this current-based neuron showed strong dependence on the degree of temporal warping applied to an input spike train (Figure 3C, bottom trace pair). Finally, the error frequency of the current-based neuron at $\beta_{\rm max}=2.5$ showed only negligible improvement upon varying the values of the membrane and synaptic time constants (Figure 3B), highlighting the limited capabilities of fixed neural kinetics to subserve time-warp–invariant spike-pattern classification.

Note that in the present classification task, the degree of time-warp robustness depends also on the learning load, i.e., number of

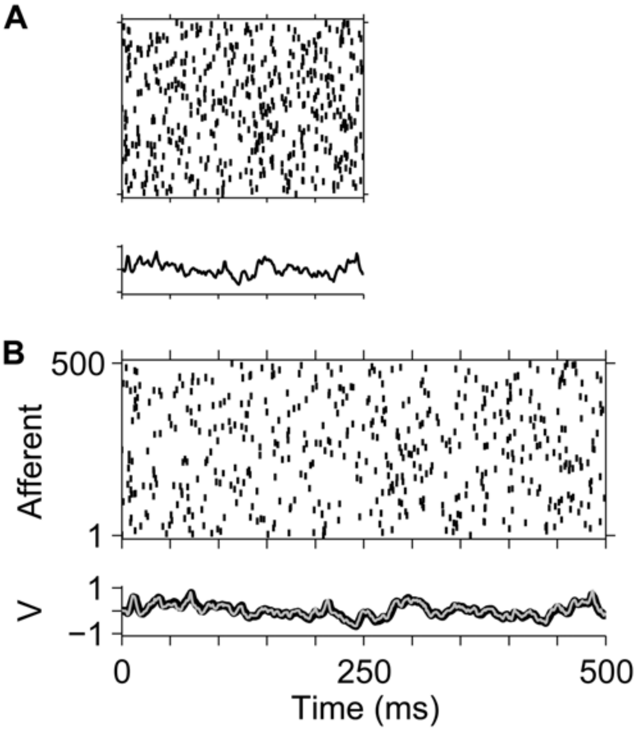

**Figure 2. Time-warp–invariant voltage traces.** Spike rasters show a random spike pattern across $N=500$ afferents ($N_{\rm ex}=250$ excitatory and $N_{\rm in}=250$ inhibitory), each of which fires a single action potential at a random time chosen uniformly between 0 and 500 ms. Whereas the original spike pattern ($\beta=1$) is shown in (B), the pattern displayed in (A) is compressed by a factor of $\beta=0.5$. In each panel, the lower trace depicts the voltage $V(t,\beta)$ induced by the spike patterns in our model neuron with balanced uniform synaptic peak conductances that resulted in a zero mean synaptic current at rest set to $g_{\rm ex}^{\rm max}=6/(N_{\rm ex}\tau_{\rm s})$ for excitatory synapses and $g_{\rm in}^{\rm max}=5g_{\rm ex}^{\rm max}$ for inhibitory synapses. These values result in a mean total synaptic conductance of $\overline{G}_{\rm syn}\approx 7g_{\rm leak}$. In (B), the voltage trace $V(t,1)$ (thin grey line) is superimposed on the rescaled voltage trace $V(\beta t,\beta)$ (thick black line) from (A).
doi:10.1371/journal.pbio.1000141.g002

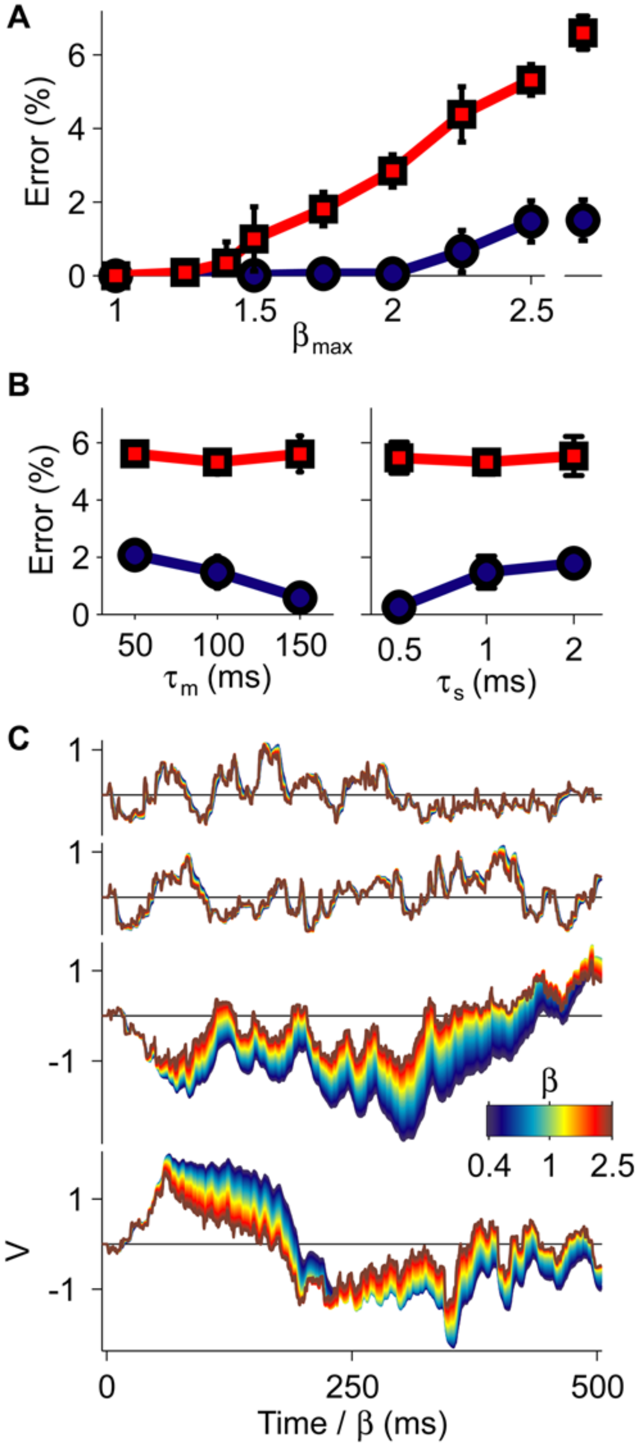

**Figure 3. Classification of time-warped random latency patterns.** (A) Error probabilities versus the scale of global time-warp $\beta_{max}$ for the conductance-based (blue) and the current-based (red) neurons. Errors were averaged over 20 realizations, error bars depict ±1 standard deviation (s.d.). Isolated points on the right were obtained under dynamic time warp with $\beta_{max}=2.5$ (see Materials and Methods). (B) Dependence of the error frequency at $\beta_{max}=2.5$ on the resting membrane time constant $\tau_m$ (left) and the synaptic time constant $\tau_s$ (right). Colors and statistics as in (A). (C) Voltage traces of a conductance-based (top and second rows) and a current-based neuron (third and bottom rows). Each trace was computed under global time warp with a temporal scaling factor $\beta$ (see Materials and Methods) (color bar) and plotted versus a common rescaled time axis. For each neuron model, the upper traces were elicited by a target and the lower traces by an untrained spike template.

doi:10.1371/journal.pbio.1000141.g003

patterns that have to classified by a neuron (unpublished data). A given degree of time warp translates into a finite range of distortions of the intracellular voltage traces. If these distortions remain smaller than the margins separating the neuronal firing threshold and the intracellular peak voltages, a neuron's classification will be time-warp invariant. Since the maximal possible margins increase with decreasing learning load, time-warp invariance can be traded for storage capacity. This tradeoff is governed by the susceptibility of the voltage traces to time warp. If the susceptibility is high, as in the current-based tempotron, robustness to time warp comes at the expense of a substantial reduction in storage capacity. If it is low, as in the conductance-based tempotron, time-warp invariance can be achieved even when operating close to the neuron's maximal storage capacity for unwarped patterns.

## Adaptive Plasticity Window

In the conductance-based tempotron, synaptic conductances controlled, not only the effective integration time of the neuron, but also the temporal selectivity of the synaptic update during learning. The tempotron learning rule modifies only the efficacies of the synapses that were activated in a temporal window prior to the peak in the postsynaptic voltage trace. However, the width of this temporal plasticity window is not fixed but depends on the effective integration time of the postsynaptic neuron at the time of each synaptic update trial, which in turn varies with the input firing rate at each trial and the strength of the peak synaptic conductances at this stage of learning (Figure 4). During epochs of high conductance (warm colors), only synapses that fired shortly before the voltage maximum were appreciably modified. In contrast, when the membrane conductance was low (cool colors), the plasticity window was broad. The ability of the plasticity window to adjust to the effective time constant of the postsynaptic voltage is crucial for the success of the learning. As is evident from Figure 4, the membrane's effective time constant varies consider-

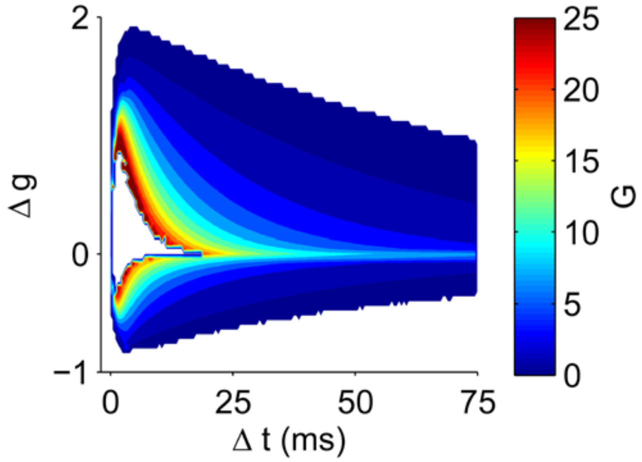

**Figure 4. Adaptive learning kernel.** Change in synaptic peak conductance $\Delta g$ versus the time difference $\Delta t$ between synaptic firing and the voltage maximum, as a function of the mean total synaptic conductance $G$ during this interval (color bar). Data were collected during the initial 100 cycles of learning with $\beta_{max}=2.5$ and averaged over 100 realizations.

doi:10.1371/journal.pbio.1000141.g004

ably during the learning epochs; hence, a plasticity rule that does not take this into account fails to credit appropriately the different synapses.

## Task Dependence of Learned Synaptic Conductance

The evolution of synaptic peak conductances during learning was driven by task requirements. When we replaced the temporal warping of the spike templates by random Gaussian jitter [22] (see Materials and Methods), conductance-based tempotrons that had acquired high synaptic peak conductances during initial training on the time-warp task readjusted their synaptic peak conductances to low values (Figure 5, inset). The concomitant increase in their effective integration time constants from roughly 10 ms to 50 ms improved the neurons' ability to average out the temporal spike jitter and substantially enhanced their task performance (Figure 5).

## Neuronal Model of Word Recognition

To address time-warp–invariant speech processing, we studied a neuronal module that learns to perform word-recognition tasks. Our model consists of two auditory processing stages. The first stage (Figure 6) consists of an afferent population of neurons that convert incoming acoustic signals into spike patterns by encoding the occurrences of elementary spectrotemporal events. This layer forms a 2-dimensional tonotopy-intensity auditory map. Each of

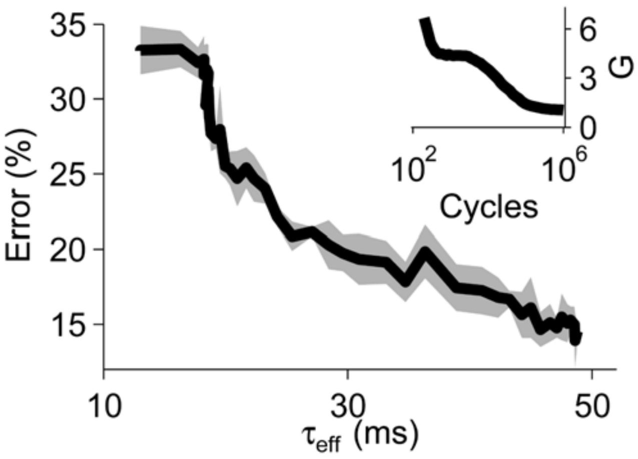

**Figure 5. Task dependence of the learned total synaptic conductance.** Error frequency of the conductance-based tempotron versus its effective integration time $\tau_{eff}$. After switching from time-warp to Gaussian spike jitter, $\tau_{eff}$ increased as the mean time-averaged total synaptic conductance $G$ decreased with learning time (inset).
doi:10.1371/journal.pbio.1000141.g005

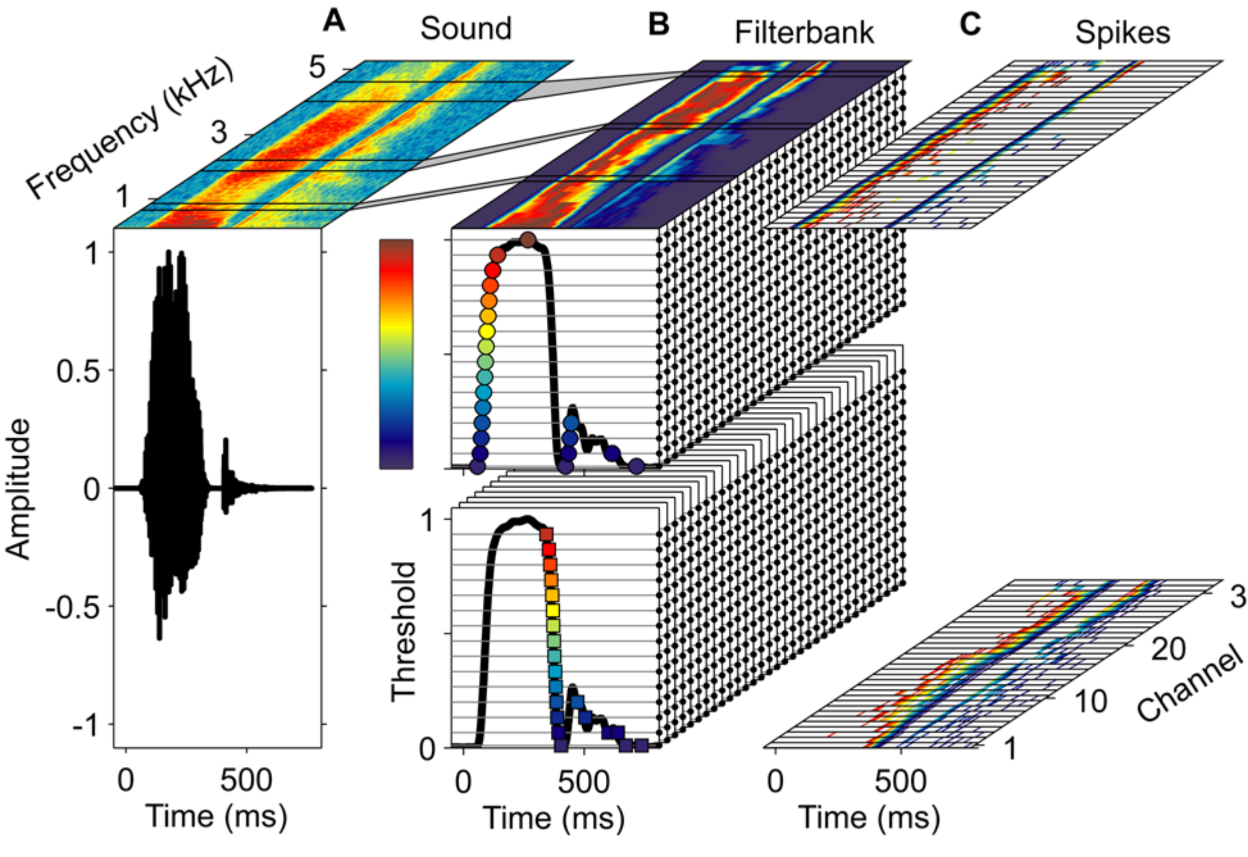

**Figure 6. Auditory front end.** (A and B) Incoming sound signal (bottom) and its spectrogram in linear scale (top) as in Figure 1D (A). Based on the spectrogram, the log signal power in 32 frequency channels (Mel scale, see Materials and Methods) is computed and normalized to unit peak amplitude in each channel ([B], top, colorbar). Black lines delineate filterbank channels 10, 20, and 30 and their respective support in the spectrogram (connected through grey areas). In each channel, spikes in 31 afferents (small black circles) are generated by 16 onset (upper block) and 15 offset (lower block) thresholds. For the signal in channel 1 (shown twice as thick black curves on the front sides of the upper and lower blocks), resulting spikes are marked by circles (onset) and squares (offset) with colors indicating respective threshold levels (colorbar). (C) Spikes (onset, top, and offset, bottom) from all 992 afferents plotted as a function of time (*x*-axis) and corresponding frequency channel (*y*-axis). The color of each spike (short thin lines) indicates the threshold level (as used for circles and squares in [B]) of the eliciting unit.
doi:10.1371/journal.pbio.1000141.g006

its afferents generates spikes by performing an onset or offset threshold operation on the power of the acoustic signal in a given frequency band. Whereas an onset afferent elicits a spike whenever the log signal power crosses its threshold level from below, offset afferents encode the occurrences of downward crossings (see Materials and Methods) (cf. also [6,23]). Different on and off neurons coding for the same frequency band differ in their threshold value, reflecting a systematic variation in their intensity tuning. The second, downstream, layer consists of neurons with plastic synaptic peak conductances that are governed by the conductance-based tempotron plasticity rule. These neurons are trained to perform word discrimination tasks. We tested this model on a digit-recognition benchmark task with the TI46 database [24]. We trained each of the 20 conductance-based tempotrons of the second layer to perform a distinct gender-specific binary classification, requiring it to fire in response to utterances of one digit and speaker gender, and to remain quiescent for all other stimuli. After training, the majority of these digit detector neurons (70%) achieved perfect classification of the test set, and the remaining ones performed their task with a low error (Table 1). Based on the spiking activity of this small population of digit detector neurons, a full digit classifier (see Materials and Methods) that weighted spikes according to each detector's individual performance, achieved an overall word error rate of 0.0017. This performance matches the error rates of state-of-the-art artificial speech-recognition systems such as the Hidden Markov model–based Sphinx-4 and HTK, which yield error rates of 0.0017 [25] and 0.0012 [26], respectively, on the same benchmark.

**Table 1.** Test set error fractions of individual detector neurons.

| Digit | Male | Female |
|---|---|---|
| 0 | 0.0 | 0.0 |
| 1 | 0.0 | 0.0 |
| 2 | 0.0008 | 0.0017 |
| 3 | 0.0 | 0.0 |
| 4 | 0.0 | 0.0 |
| 5 | 0.0029 | 0.0062 |
| 6 | 0.0 | 0.0 |
| 7 | 0.0004 | 0.0008 |
| 8 | 0.0 | 0.0 |
| 9 | 0.0 | 0.0 |

doi:10.1371/journal.pbio.1000141.t001

## Learned Spectrotemporal Target Features

To reveal qualitatively some of the mechanisms used by our digit detector neurons to selectively detect their target word, we compared the learned synaptic distributions (Figure 7A) of two digit detector neurons ("one" and "four") to the average spectrograms of each neuron's target stimuli aligned to the times of its output spikes (Figure 7B; see Materials and Methods). The spectrotemporal features that preceeded the output spikes (time zero, grey vertical lines) corresponded to the frequency-specific onset and offset selectivity of the excitatory afferents (Figure 7A, warm colors). These examples demonstrate how the patterned excitatory and inhibitory inputs from both onset and offset neurons are tuned to features of the speech signal. For instance, a prominent feature in the averaged spectrogram of the word "one" (male speakers) was the increase in onset time of the power in the low-frequency channels with the frequency of the channel (Figure 7B, left, channels 1–16). This gradual onset was encoded by a diagonal band of excitatory onset afferents whose thresholds decreased with increasing frequency (Figure 7A, left). By compensating for the temporal lag between the different lower-frequency channels, this arrangement ensured a strong excitatory drive when a target stimulus was presented to the neuron. The spectrotemporal feature learned by the word "four" (male speakers) detector neuron combined decreasing power in the low-frequency range with rising power in the mid-frequency range (Figure 7B, right). This feature was encoded by synaptic efficacies through a combination of excitatory offset afferents in the low-frequency range (Figure 7A, right, channels 1–11) and excitatory onset afferents in the mid-frequency range (channels 12–19). Excitatory synaptic populations were complemented by inhibitory inputs (Figure 7A, blue patches) that prevented spiking in response to null stimuli and also increased the total synaptic conductance. The substantial differences between the mean spike-triggered voltage traces for target stimuli (Figure 7C, blue) and the mean maximum-triggered voltage traces for null stimuli (red) underline the high target word selectivity of the learned synaptic distributions as well as the relatively short temporal extend of the learned target features.

In the examples shown, the average position of the neural decision relative to the target stimuli varied from early to late (Figure 7B, left vs. right). This important degree of freedom stems from the fact that the tempotron decision rule does not constrain the time of the neural decision. As a result, the learning process in each neuron can select the spectrotemporal target features from any time window within the word. The selection of the target feature by the learning takes into account both the requirement of triggering output spikes in response to target stimuli as well as the demand to remain silent during null stimuli. Thus, for each target neuron, the selected features reflect the statistics of both the target and the null stimuli.

## Generalization Abilities of Word Detector Neurons

We have performed several tests designed to assess the ability of the model word detector neurons to perform well on new input sets, different in statistics from the trained database. First, we assessed the ability of the neurons to generalize to unfamiliar speakers and dialects. After training the model with the TI46 database, as described above, we measured its digit-recognition performance on utterances from another database, the TIDIGITS database [27], which includes speech samples from a variety of English dialects (see Materials and Methods). This test has been done without any retraining of the network synapses. The resulting word error rate of 0.0949 compares favorably to the performance of the HTK system, which resulted in an error rate of 0.2156 when subjected to the same generalization test (see Materials and Methods). Across all dialects, our model performed perfectly for roughly one-quarter of all speakers and with at most one error for half of them. Within the best dialect group, an error of at most one word was achieved for as many as 80% of the speakers (Table S1). These results underline the ability of our neuronal word-recognition model to generalize to unfamiliar speakers across a wide range of different unfamiliar dialects.

An interesting question is whether our model neurons are able to generalize their performance to novel time-warped versions of the trained inputs. To address this question, we have tested their performance on randomly generated time-warped versions of the input spikes corresponding to the trained word utterances, without retraining. As shown in Figure 8, the neurons exhibited considerable time-warp–robust performance on the digit-recogni-

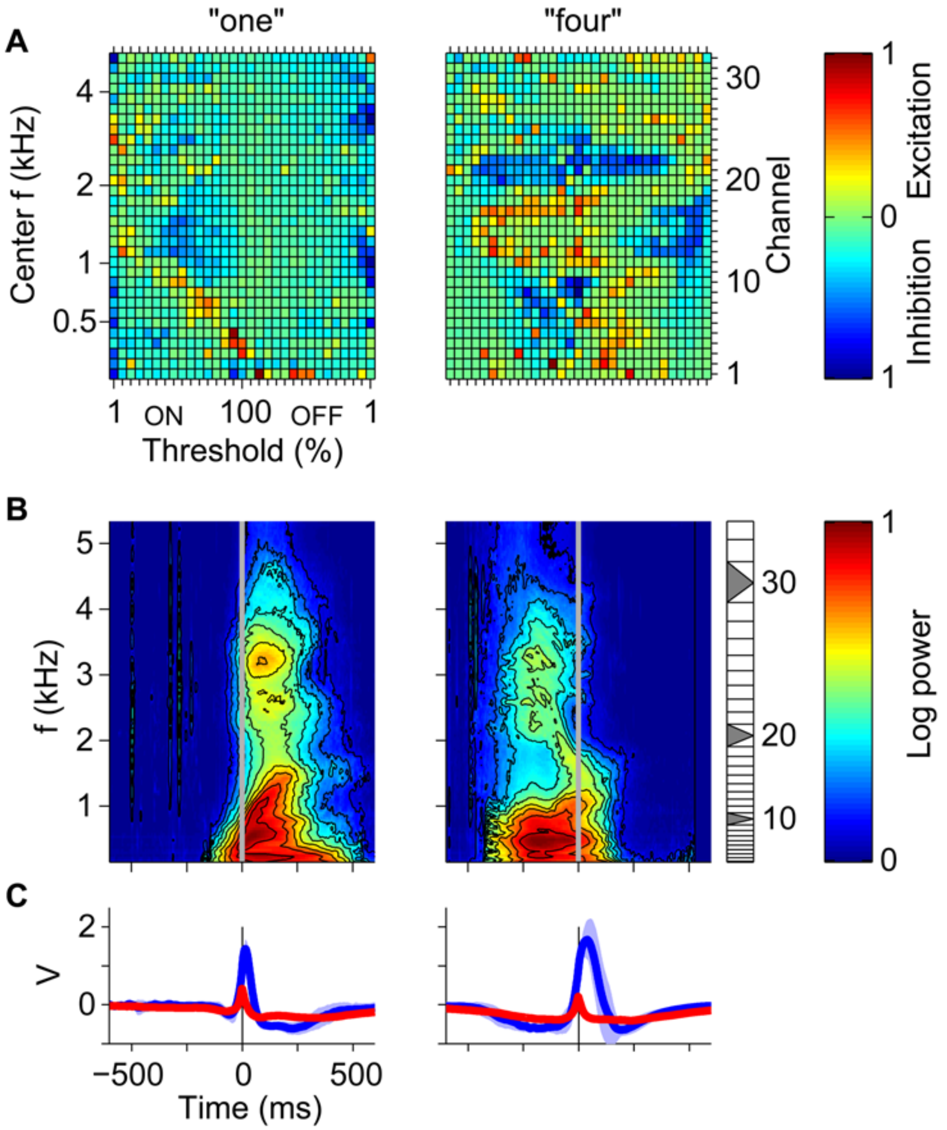

**Figure 7. Speech-recognition task.** (A) Learned synaptic peak conductances. Each pixel corresponds to one synapse characterized by its frequency channel (right *y*-axis) and its onset (ON) or offset (OFF) afferent power threshold level (*x*-axis, in percent of maximum signal powers [see Materials and Methods]). Learned peak conductances were color coded with excitatory (warm colors) and inhibitory conductances (cool colors) separately normalized to their respective maximal values (color bar). The left *y*-axis shows the logarithmically spaced center frequencies (Mel scale) of the frequency channels. (B) Spike-triggered target stimuli (color-code scaled between the minimum and maximum mean log power). (C) Mean voltage traces for target (blue, light blue ±1 s.d.; spike triggered) and null stimuli (red; maximum triggered).

doi:10.1371/journal.pbio.1000141.g007

tion task. For instance, the errors for the "one" (Figure 8A, black line) and "four" (blue line) detector neurons (cf. Figure 7) were insensitive to a 2-fold time warp of the input spike trains. The "seven" detector neuron (male, red line) showed higher sensitivity to such warping; nevertheless, its error rate remained low. Consistent with the proposed role of synaptic conductances, the degree of time-warp robustness was correlated with the total synaptic conductance, here quantified through the mean effective integration time $\tau_{\mathrm{eff}}$ (Figure 8B). Additionally, the mean voltage traces induced by the target stimuli (Figure 8C, lower traces) showed a substantially smaller sensitivity to temporal warping than their current-based analogs (see Materials and Methods) (Figure 8C, upper traces).

We also found that our model word detector neurons are robust to the introduction of spike failures in their input patterns. For each neuron, we have measured its performance on inputs which were corrupted by randomly deleting a fraction of the incoming spikes, again without retraining. For the majority of neurons, the error percentage increased by less than 0.01% for each percent increase in spike failures (Figure 9). This high robustness reflects the fact that each classification is based on integrating information from many presynaptic sources.

## Discussion

### Automatic Rescaling of Effective Integration Time by Synaptic Conductances

The proposed conductance-based time-rescaling mechanism is based on the biophysical property of neurons that their effective integration time is shaped by synaptic conductances and therefore

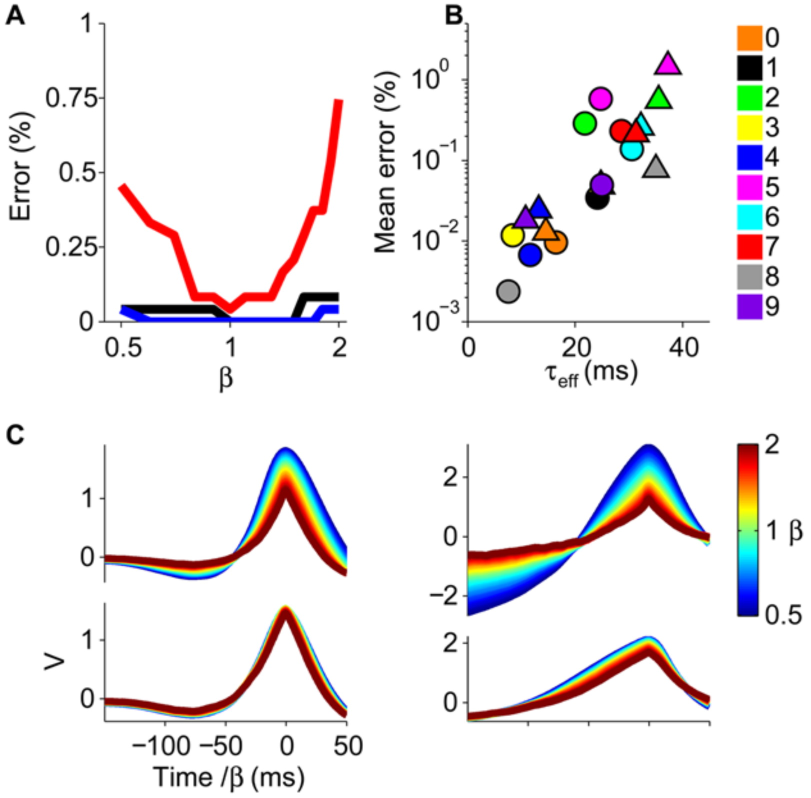

**Figure 8. Time-warp robustness.** (A) Error versus time-warp factor $\beta$. (B) Mean errors over the range of $\beta$ shown in (A) (digit color code; triangles: female speakers, circles: male speakers) versus the mean effective time constant $\tau_{\rm eff}$ calculated for $\beta = 1$ by averaging the total synaptic conductance over 100-ms time windows prior to either the output spikes (target stimuli) or the voltage maxima (null stimuli). (C) Mean voltage traces for time-warped target patterns for the neurons shown in Figure 7. Bottom row: conductance-based neurons, upper row: current-based neurons (see Materials and Methods).

doi:10.1371/journal.pbio.1000141.g008

can be modulated by the firing rate of its afferents. To utilize these modulations for time-warp–invariant processing, a central requirement is a large evoked total synaptic conductance that dominates the effective integration time constant of the postsynaptic cell through shunting. In our speech-processing model, large synaptic conductances with a median value of a 3-fold leak conductance across all digit detector neurons (cf. Figure 8B) result from a combination of excitatory and inhibitory inputs. This is consistent with high total synaptic conductances, comprising excitation and inhibition, that have been observed in several regions of cortex [28] including auditory [29,30], visual [31,32], and also prefrontal [33,34] (but see ref. [35]). Our model predicts that in cortical sensory areas, the time-rescaled intracellular voltage traces (cf. Figure 3C), and consequently, also the rescaled spiking responses of neurons that operate in the proposed fashion, remain invariant under temporal warping of the neurons' input spike patterns. These predictions can be tested by intra- and extracellular recordings of neuronal responses to temporally warped sensory stimuli.

A large total synaptic conductance is associated with a substantial reduction in a neuron's effective integration time relative to its resting value. Therefore, the resting membrane time constant of a neuron that implements the automatic time-rescaling mechanism must substantially exceed the temporal resolution that is required by a given processing task. Because the word-recognition benchmark task used here comprises whole-word stimuli that favored effective time constants on the order of several tens of milliseconds, we used a resting membrane time constant of $\tau_{\rm m} = 100$ ms. Whereas values of this order have been reported in hippocampus [36] and cerebellum [21,37], it exceeds current estimates for neocortical neurons, which range between 10 and 30 ms [35,38,39]. Note, however, that the correspondence of our passive membrane model and the experimental values that typically include contributions from various voltage-dependent conductances is not straightforward. Our model predicts that neurons specialized for time-warp–invariant processing at the whole-word level have relatively long resting membrane time constants. It is likely that the auditory system solves the problem of time-warp–invariant processing of the sound signal primarily at the level of shorter speech segments such as phonemes. This is supported by evidence that primary auditory cortex has a special role in speech processing at a resolution of milliseconds to tens of milliseconds [11–13]. Our mechanism would enable time-warp–invariant processing of phonetic segments with resting membrane time constants in the range of tens of milliseconds, and much shorter effective integration times.

The proposed neuronal time-rescaling mechanism assumes linear summation of synaptic conductances. This assumption is challenged by the presence of voltage-dependent conductances in neuronal membranes. Since the potential implications for our model depend on the specific nonlinearity induced by a cell-type–specific composition of different ionic channels, it is hard to evaluate the overall effect on our model in general terms. Nevertheless, because of its immanence, we expect the conductance-based time-rescaling mechanism to cope gracefully with moderate levels of nonlinearity. As an example, we tested its

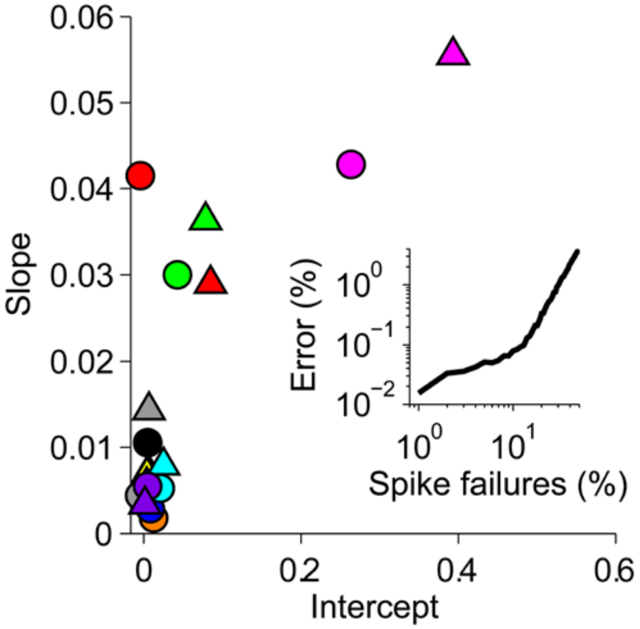

**Figure 9. Robustness to spike failures.** The error fraction of each digit detector neuron was measured as a function of the spike failure probability over the range from 0% to 10% and fitted by linear regression. For each neuron, the resulting slope (median 0.0069) is plotted versus the intercept (median 0.0061) with symbols and colors as in Figure 8B. The median $R^2$ of the linear regression fits was 0.94. The inset shows the median error fraction of the population as a function of the spike failure probability in the range of 1% to 50% with the robust regime braking down at approximately 20%.
doi:10.1371/journal.pbio.1000141.g009

behavior in the presence of an h-like conductance (see Materials and Methods) that opposes conductance changes induced by depolarizing excitatory synaptic inputs and is active at the resting potential. As expected, we found that physiological levels of h-conductances resulted in only moderate impairment of the automatic time-rescaling mechanism (Figure S1).

For the sake of simplicity as well as numerical efficiency, we have assumed symmetric roles of excitation and inhibition in our model architecture. We have checked that this assumption is not crucial for the operation of the automatic time-rescaling mechanism and the learning of time-warped random latency patterns. Specifically, we have implemented the random latency classification task for a control architecture in which all synapses were confined to be excitatory except a single global inhibitory input that, mimicking a global inhibitory network, received a separate copy of all incoming spikes. In this architecture, all spike patterns have to be encoded by the excitatory synaptic population, and the role of inhibition is reduced to a global signal that has equal strength for all input patterns. Due to the limitations of this architecture, this model showed some reduction of storage capacity relative to the symmetric case, but the automatic time-rescaling mechanism remained intact. For a time-warp scale of $\beta_{\max} = 2.5$ (cf. Figure 3), the global inhibition model roughly matched the performance of the symmetric model when the learning load was lowered to 1.5 spike patterns per synapse, with an error fraction of 0.18%.

## Supervised Learning of Synaptic Conductances

To utilize synaptic conductances as efficient controls of the neuron's clock, the peak synaptic conductances must be plastic so that they adjust to the range of integration times relevant for a given perceptual task. This was achieved in our model by our novel supervised spike-based learning rule. This plasticity posits that the temporal window during which pre- and postsynaptic activity interact continuously adapts to the effective integration time of the postsynaptic cell (Figure 4). The polarity of synaptic changes is determined by a supervisory signal that we hypothesize to be realized through neuromodulatory control [22]. Because present experimental measurements of spike-timing–dependent synaptic plasticity rules have assumed an unsupervised setting, i.e., have not controlled for neuromodulatory signals (but see [40]), existing results do not directly apply to our model. Nevertheless, recent data have revealed complex interactions between the statistics of pre- and postsynaptic spiking activity and the expression of synaptic changes [41–44]. Our model offers a novel computational rationale for such interactions, predicting that for fixed supervisory signaling, the temporal window of plasticity shrinks with growing levels of postsynaptic shunting. One challenge for the biological implementation of the tempotron learning rule is the need to compute the time of the maximum of the postsynaptic voltage. We have previously shown for a current-based neuron model that this temporally global operation can be approximated by temporally local computations that are based on the postsynaptic voltage traces following input spikes [22]. We have extended this approach to plastic synaptic conductances and checked that the resulting biologically plausible implementation of conductance-based tempotron learning can readily subserve time-warp–invariant classification of spike patterns. Specifically, in this implementation, the induction of synaptic plasticity is controled by the correlation of the postsynaptic voltage and a synaptic learning kernel (see Materials and Methods) whose temporal extend is controlled by the average conductance throughout a given error trial. A synaptic peak conductance is changed by a uniform amount whenever this correlation exceeds a fixed plasticity induction threshold. When tested on the time-warped latency patterns with $\beta_{\max} = 2.5$ (cf. Figure 3), the correlation-based tempotron roughly matched the voltage maximum–based version at a reduced learning load of 1.5 patterns per synapse with an error fractions of 0.35%.

## Time-Warp Invariance Is Task Dependent

In our model, dynamic time-warp–invariant capabilities become avaliable through a conductance-based learning rule that tunes the shunting action of synaptic conductances. This learning rule enables neurons to adjust the degree of synaptic shunting to the requirements of a given processing task. As a result, our model can naturally encompass a continuum of functional specializations ranging from neurons that are sensitive to absolute stimulus durations by employing low total synaptic conductances, to time-warp–invariant feature detectors that operate in a high-conductance regime. In the context of auditory processing, such a functional segregation into neurons with slower and faster effective integration times is reminiscent of reports suggesting that rapid temporal processing in time frames of tens of milliseconds is localized in left lateralized language areas, whereas processing of slower temporal features is attributed to right hemispheric areas [45–47]. Although anatomical and morphological asymmetries between left and right human auditory cortices are well documented [48], it remains to be seen whether these differences form the physiological substrate for a left lateralized implementation of the proposed time-rescaling mechanism. Consistent with this picture, the general tradeoff between high temporal resolution and robustness to temporal jitter that is predicted by our model (Figure 5) parallels reports of the vulnerability of the lateralizion of language processing with respect to background acoustic noise [49] as well as to abnormal timing of auditory brainstem responses [50].

## Neuronal Circuitry for Time-Warp–Invariant Feature Detection

The architecture of our speech-processing model encompasses two auditory processing stages. The first stage transforms acoustic signals into spatiotemporal patterns of spikes. To engage the proposed automatic time-rescaling mechanism, the population rate of spikes elicited in this afferent layer must track variations in the rate of incoming speech. Such behavior emerges naturally in a sparse coding scheme in which each neuron responds transiently to the occurrences of a specific acoustic event within the auditory input. As a result, variations in the rate of acoustic events are directly translated into concomitant variations in the population rate of elicited spikes. In our model, the elementary acoustic events correspond to onset and offset threshold crossings of signal power within specific frequency channels. Such frequency-tuned onset and offset responses featuring a wide range of dynamic thresholds have been observed in the inferior colliculus (IC) of the auditory midbrain [51]. This nucleus is the site of convergence of projections from the majority of lower auditory nuclei and is often referred to as the interface between the lower brain stem auditory pathways and the auditory cortex. Correspondingly, we hypothesize that the layer of time-warp–invariant feature detector neurons in our model implements neurons located downstream of the IC, most probably in primary auditory cortex. Current studies on the functional role of the auditory periphery in speech perception and its pathologies have been limited by the lack of biologically plausible neuronal readout architectures; a limitation overcome by our model, which allows evaluation of specific components of the auditory pathway in a functional context.

## Implications for Speech Processing

Psychoacoustic studies have indicated that the neural mechanism underlying the perceptual normalization of temporal speech cues is involuntary, i.e., it is cognitively impenetrable [16], controlled by physical rather than perceived speaking rate [17], confined to a temporally local context [2,18], not specific to speech sounds [52], and already operational in prearticulate infants [53]. The proposed conductance-based time-rescaling mechanism is consistent with these constraints. Moreover, our model posits a direct functional relation between high synaptic conductances and the time-warp robustness of human speech perception. This relation gives rise to a novel mechanistic hypothesis explaining the impaired capabilities of elderly listeners to process time-compressed speech [54,55]. We hypothesize that the down-regulation of inhibitory neurotransmitter systems in aging mammalian auditory pathways [56,57] limits the total synaptic conductance and therefore prevents the time-rescaling mechanism from generating short, effective time constants through synaptic shunting. Furthermore, our model implies that comprehension deficits in older adults should be linked specifically to the processing of phonetic segments that contain fast time-compressed temporal cues. Our hypothesis is consistent with two interrelated lines of evidence. First, comprehension difficulties of time-compressed speech in older adults are more likely a consequence of an age-related decline in central auditory processing than attributes of a general cognitive slowing [56,58]. Second, recent reports have indicated that recognition differences between young and elderly listeners originate mainly from the temporal compression of consonants, which often feature rapid spectral transitions, but not from steady-state segments [54,55,58] of speech. Finally, our hypothesis posits that speaking rate–induced shifts in perceptual category boundaries [2,16,17] should be age-dependent, i.e., their magnitude should decrease with increasing listener age. This prediction is straightforwardly testable within established psychoacoustic paradigms.

## Connections to Other Models of Time-Warp–Invariant Processing

In a previous neuronal model of time-warp–invariant speech processing [5,6], sequences of acoustic events are converted into patterns of transiently matching firing rates in subsets of neurons within a population, which trigger synchronous firing in a downstream readout circuit. The identity of neurons whose firing rates converge to an identical value during an input pattern, and hence also the pattern of synchrony emerging in the readout layer, depends only on the relative timing of the events, not on the absolute duration of the auditory signal. However, for this model to recognize multiple input patterns, the convergence of firing rates is only approximate. Therefore, the resulting time-warp robustness is limited and, as in our model, dependent on the learning load. Testing this model on our synthetic classification task (cf. Figure 3) indicated a substantially smaller storage capacity than is realizable by the conductance-based tempotron (Text S1). An additional disadvantage of this approach is that it copes only with global (uniform) temporal warping. Invariant processing of dynamic time warp as is exhibited by natural speech (cf. Figure 1C and 1D) is more challenging since it requires robustness to local temporal distortions of a certain statistical character. Established algorithms that can cope with dynamically time-warped signals are typically based on minimizing the deviation between an observed signal and a stored reference template [59–61]. These algorithms are computationally expensive and lack biologically plausible neuronal implementations. By contrast, our conductance-based time-rescaling mechanism results naturally from the biophysical properties of input integration at the neuronal membrane and does not require dedicated computational resources. Importantly, our model does not rely on a comparison between the incoming signal and a stored reference template. Rather, after synaptic conductances have adjusted to the statistics of a given stimulus ensemble, the mechanism generalizes and automatically stabilizes neuronal voltage responses against dynamic time warp even when processing novel stimuli (cf. Figure 3C). The architecture of our neuronal model also fundamentally departs from the decades-old layout of Hidden Markov Model–based artificial speech-recognition systems, which employ probabilistic models of state sequences. These systems are hard to reconcile with the biological reality of neuronal system architecture, dynamics, and plasticity. The similarity in performance between our model and such state-of-the-art systems on a small vocabulary task highlights the powerful processing capabilities of spike-based neural representations and computation.

## Generality of Mechanism

Although the present work focuses on the concrete and well-documented example of time-warp robustness in the context of neural speech processing, the proposed mechanism of automatic rescaling of integration time is general and applies also to other problems of neuronal temporal processing such as birdsong recognition [3], insect communication [9], and other ethologically important natural auditory signals. Moreover, robustness of neuronal processing to temporal distortions of spike patterns is not only important for the processing of stimulus time dependencies, but also in the context of spike-timing–based neuronal codes in which the precise temporal structure of spiking activity encodes information about nontemporal physical stimulus dimensions [62]. Evidence for such temporal neural codes have been reported in the visual [63–65], auditory [66], and somatosensory [67], as well as the olfactory [68] pathways. As a result, we expect mechanisms of time-warp–invariant processing to also play a role in generating perceptual constancies along nontemporal stimulus dimensions such as contrast invariance in vision or concentration invariance in olfaction [4]. Finally, time warp has also been described in intrinsically generated brain signals. Specifically, the replay of hippocampal and cortical spiking activity at variable temporal warping [69,70] suggests that our model has applicability beyond sensory processing, possibly also encompassing memory storage and retrieval.

## Materials and Methods

### Conductance-Based Neuron Model

Numerical simulations of the conductance-based tempotron were based on exact integration [71] of the conductance-based voltage dynamics defined in Equation 1. With the membrane capacitance set to 1, the resting membrane time constant in this model is $\tau_{\mathrm{m}} = 1/g_{\mathrm{leak}}$. Implementing an integrate-and-fire neuron model, an output spike was elicited when $V(t)$ crossed the firing threshold $V_{\mathrm{thr}}$. After a spike at $t_{\mathrm{spike}}$, the voltage is smoothly reset to the resting value by shunting all synaptic inputs that arrive after $t_{\mathrm{spike}}$ (cf. [22]). We used $V_{\mathrm{thr}} = 1$, $V_{\mathrm{rest}} = 0$, and reversal potentials $V_{\mathrm{ex}}^{\mathrm{rev}} = 5$ and $V_{\mathrm{in}}^{\mathrm{rev}} = -1$ for excitatory and inhibitory conductances, respectively. Unless stated otherwise, the resting membrane time constant was set to $\tau_{\mathrm{m}} = 100$ ms throughout our work [20]. For the synaptic time constant, we used $\tau_{\mathrm{s}} = 1$ ms for the random latency task, which minimized the error of the current-based neuron, and to $\tau_{\mathrm{s}} = 5$ ms in the speech-recognition tasks.

### H-Current

To check the effect of nonsynaptic voltage-dependent conductances on the automatic time-rescaling mechanism, we implemented an h-like current $I_{\mathrm{h}}$ after [72] as a noninactivating current with HH-type dynamics of the form

$$I_{\mathrm{h}} = g_{\mathrm{h}}^{\max} m\left(V - V_{\mathrm{h}}^{\mathrm{rev}}\right).$$

Here, $g_{\mathrm{h}}^{\max}$ is the maximal h-conductance, with reversal potential $V_{\mathrm{h}}^{\mathrm{rev}} = -20$ mV and $m$ is its voltage-dependent activation variable with kinetics

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{m_{\infty}(V) - m}{\tau_{\mathrm{h}}(V)}$$

where

$$\tau_{\mathrm{h}}(V) = \frac{1}{\alpha(V) + \beta(V)}$$

and

$$m_{\infty}(V) = \frac{\alpha(V)}{\alpha(V) + \beta(V)}.$$

The voltage dependence of the rate constants $\alpha$ and $\beta$ were described by the form

$$\alpha,\, \beta(V) = \frac{a_{\alpha,\beta} V + b_{\alpha,\beta}}{1 - \exp\left[\left(V + b_{\alpha,\beta}/a_{\alpha,\beta}\right)/k_{\alpha,\beta}\right]}$$

with parameters $a_{\alpha} = -39.015\ \mathrm{s}^{-1}$, $b_{\alpha} = -259.925\ \mathrm{s}^{-1}$, $k_{\alpha} = 1.77926$ and $a_{\beta} = 365.85\ \mathrm{s}^{-1}$, $b_{\beta} = -2853.25\ \mathrm{s}^{-1}$, $k_{\beta} = -1.28889$.

In Figure S1, we quantified the effect of the h-conductance on the fidelity of the time-rescaling mechanism by measuring the time-warp–induced distortions of neuronal voltage traces for different values of the maximal h-conductance $g_{\mathrm{h}}^{\max}$. Specifically, for a given value of $g_{\mathrm{h}}^{\max}$ and a time warp $\beta$, we measure the voltage traces $V_{g_{\mathrm{h}}^{\max}}(t, 1)$ and $V_{g_{\mathrm{h}}^{\max}}(t, \beta)$ and their standard deviations across time $\sigma_1$ and $\sigma_{\beta}$, respectively. We define the time-warp distortion index $\Lambda\left(g_{\mathrm{h}}^{\max}, \beta\right)$ as the mean magnitude of the time-warp–induced voltage difference across time normalized by the mean standard deviation, $\bar{\sigma} = \left(\sigma_1 + \sigma_{\beta}\right)/2$,

$$\Lambda\left(g_{\mathrm{h}}^{\max}, \beta\right) = \frac{\left\langle \left| V_{g_{\mathrm{h}}^{\max}}(t, 1) - V_{g_{\mathrm{h}}^{\max}}(\beta t, \beta) \right| \right\rangle_t}{\bar{\sigma}}.$$

In Figure S1, values of $\Lambda\left(g_{\mathrm{h}}^{\max}, \beta\right)$ are normalized by $\Lambda(0,\beta)$. The voltage traces were generated by random latency patterns and uniform synaptic peak conductances as used in Figure 2. As increasing values of $g_{\mathrm{h}}^{\max}$ depolarized the neuron's resting potential, excitatory and inhibitory synaptic conductances were balanced separately for each value of $g_{\mathrm{h}}^{\max}$.

### Current-Based Neuron Model

In the current-based tempotron that was implemented as described in [22], each input spike evoked an exponentially decaying synaptic current that gave rise to a postsynaptic potential with a fixed temporal profile. In Figure 8C (upper row), voltage traces of a current-based analog of a conductance-based tempotron with learned synaptic conductances $g_i^{\max}$, reversal potentials $V_i^{\mathrm{rev}}$, and effective membrane integration time $\tau_{\mathrm{eff}}$ (cf. Figure 8B) were computed by setting the synaptic efficacies $\omega_i$ of the current-based neuron to $\omega_i = g_i^{\max} V_i^{\mathrm{rev}}$ and its membrane time constant to $\tau_{\mathrm{m}} = \tau_{\mathrm{eff}}$. The resulting current-based voltage traces were scaled such that for each pair of models, the mean voltage maxima for unwarped stimuli ($\beta = 1$) were equal.

### Tempotron Learning

Following [22], changes in the synaptic peak conductance $g_i^{\max}$ of the $i$th synapse after an error trial were given by the gradient of the postsynaptic potential, $\Delta g_i^{\max} \propto -\mathrm{d}V(t_{\max})/\mathrm{d}g_i^{\max}$, at the time of its maximal value $t_{\max}$. To compute the synaptic update for a given error trial, the exact solution of Equation 1 was differentiated with respect to $g_i^{\max}$ and evaluated at $t_{\max}$, which was determined numerically for each error trial. Whenever a synaptic peak conductance attempted to cross to a negative value, its reversal potential was switched.

### Voltage Correlation-Based Learning

A voltage correlation-based approximation of tempotron learning was implemented by extending the approach in [22] such that the change in the synaptic peak conductance $g_i^{\max}$ of the $i$th synapse due to a spike at time $t_i$ was governed by the correlation $v_i = \int_{t_i}^{\infty} \mathrm{d}t\, V(t) K_{\mathrm{learn}}(t - t_i)$ of the postsynaptic potential $V(t)$ with a synaptic learning kernel $K_{\mathrm{learn}}(t) = (\exp(-t/\tau_{\mathrm{learn}}) - \exp(-t/\tau_{\mathrm{s}}))/(\tau_{\mathrm{learn}} - \tau_{\mathrm{s}})$. The two time constants of the synaptic learning kernel were the synaptic time constant $\tau_s$ and the learning time constant $\tau_{\mathrm{learn}} = 1/\left(g_{\mathrm{leak}} + \overline{G}_{\mathrm{syn}}\right)$, which was determined by the time-averaged synaptic conductance $\overline{G}_{\mathrm{syn}}$ of the current error trial and approximated the effective membrane time constant during that trial. The voltage maximum operation was approximated by thresholding $v_i$, yielding

$$\Delta g_i^{\max} \propto \begin{cases} \pm 1 & v_i > \kappa \\ 0 & v_i \leq \kappa \end{cases}$$

for changes of excitatory conductances on target and null patterns, respectively, and changes with the reversed polarity, ±1, for inhibitory conductances. The plasticity induction threshold was set to $\kappa = 0.45$.

## Learning Rate and Momentum Term

As in [22], we employed a momentum heuristic to accelerate learning in all learning rules. In this scheme, synaptic updates $\left[\Delta g_i^{\max}\right]_{\mathrm{current}}$ consisted, not only of the correction $\lambda \Delta g_i^{\max}$, which was given by the learning rule and the learning rate $\lambda$, but also incorporated a fraction $\mu$ of the previous synaptic change $\left[\Delta g_i^{\max}\right]_{\mathrm{previous}}$. Hence, $\left[\Delta g_i^{\max}\right]_{\mathrm{current}} = \lambda \Delta g_i^{\max} + \mu [\Delta g^{\max}]_{\mathrm{previous}}$. We used an adaptive learning rate that decreased from its initial value $\lambda_{\mathrm{ini}}$ as the number of learning cycles $l$ grew, $\lambda = \lambda_{\mathrm{ini}}/(1+10^{-4}(l-1))$. A learning cycle corresponded to one iteration through the batch of templates in the random latency task or the training set in the speech task.

**Random latency task training.** To ensure a fair comparison between the conductance-based and the current-based tempotrons (cf. Figure 3A), the learning rule parameters $\lambda_{\mathrm{ini}}$ and $\mu$ were optimized for each model. Specifically, for each value of $\beta_{\max}$, optimal values over a 2-dimensional grid were determined by the minimal error frequency achieved during runs over $10^5$ cycles, with synaptic efficacies starting from Gaussian distributions with zero mean and standard deviations of 0.001. The optimization was performed over five realizations.

## Global Time Warp

Global time warp was implemented by multiplying all firing times of a spike template by a constant scaling factor $\beta$. In Figure 3A, random global time warp between compression by $1/\beta_{\max}$ and dilation by $\beta_{\max}$ was generated by setting $\beta = \exp(q \ln(\beta_{\max}))$ with $q$ drawn from a uniform distribution between $-1$ and 1 for each presentation.

## Dynamic Time Warp

Dynamic time warp was implemented by scaling successive interspike intervals $t_j - t_{j-1}$ of a given template with a time-dependent warping factor $\tilde{\beta}(t)$, such that warped spike times $t'_j = t'_{j-1} + \tilde{\beta}(t_j)\left(t_j - t_{j-1}\right)$ with $t'_1 \equiv t_1$ and $\tilde{\beta}(t) = \exp(\tilde{q}(t) \ln(\beta_{\max}))$. The time-dependent factor $\tilde{q}(t) = \mathrm{erfc}(\xi(t)) - 1$ resulted from an equilibrated Ornstein-Uhlenbeck process $\xi(t)$ with a relaxation time of $\tau = 200$ ms that was rescaled by the complementary error function erfc to transform the normal distribution of $\xi(t)$ into a uniform distribution over $[-1\ 1]$ at each $t$.

## Global Inhibition Model

To ensure that the symmetry of excitation and inhibition in our model architecture was not crucial for the time-warp–invariant processing of spike patterns, we implemented a control architecture in which all afferents were confined to be excitatory, except one additional inhibitory synapse, which mimicked the effect of a global inhibitory network. In the time-warped random latency task, spike patterns were fed into the excitatory population as before; however, in addition, the inhibitory synapse received a copy of each incoming spike. All synaptic peak conductances were plastic and controlled by the conductance-based tempotron rule. In this model, synaptic sign changes were prohibited.

## Gaussian Spike Time Jitter

Spike time jitter [22] was implemented by adding independent Gaussian noise with zero mean and a standard deviation of 5 ms to each spike of a template before each presentation.

## Acoustic Front-End

Sound signals were normalized to unit peak amplitude and converted into spectrograms over $N_{\mathrm{FTT}} = 129$ linearly spaced frequencies $f_j = f_{\min} + j(f_{\max} + f_{\min})/(N_{\mathrm{FTT}}+1)$ $(j = 1 \ldots N_{\mathrm{FTT}})$ between $f_{\min} = 130$ Hz and $f_{\max} = 5{,}400$ Hz by a sliding fast Fourier transform with a window size of 256 samples and a temporal step size of 1 ms. The resulting spectrograms were filtered into $N_{\mathrm{f}} = 32$ logarithmically spaced Mel frequency channels by overlapping triangular frequency kernels. Specifically, $N_{\mathrm{f}}+2$ linearly spaced frequencies given by $h_j = h_{\min} + j(h_{\max} - h_{\min})/(N_{\mathrm{f}}+1)$ with $j = 0 \ldots N_{\mathrm{f}}+1$ and $h_{\max,\min} = 2{,}595 \log(1 + f_{\max,\min}/700)$ were transformed to a Mel frequency scale $f_j^{\mathrm{Mel}} = 700\left(\exp\left(h_j/2595\right) - 1\right)$ between $f_{\min}$ and $f_{\max}$. Based on these, signals in $N_{\mathrm{f}}$ channels resulted from triangular frequency filters over intervals $\left[f_{j-1}^{\mathrm{Mel}}, f_{j+1}^{\mathrm{Mel}}\right]$ with center peaks at $f_j^{\mathrm{Mel}}$ $(j = 1 \ldots N_{\mathrm{f}})$. After normalization of the resulting Mel spectrogram $S^{\mathrm{Mel}}$ to unit peak amplitude, the logarithm was taken through $\log(S^{\mathrm{Mel}} = \varepsilon) - \log(\varepsilon)$ with $\varepsilon = 10^{-5}$ and the signal in each frequency channel smoothed in time by a Gaussian kernel with a time constant of 10 ms. Spikes were generated by thresholding of the resulting signals by a total of 31 onset and offset threshold-crossing detector units. Whereas each onset afferent emitted a spike whenever the signal crossed its threshold in the upward direction, offset afferents fired when the signal dropped below the threshold from above. For each frequency channel and each utterance, threshold levels for onset and offset afferents were set relative to the maximum signal over time to $\vartheta_1 = 0.01$ and $\vartheta_j = j/15$ $(j = 1 \ldots 15)$. For $\vartheta_{15} = 1$, onset and offset afferents were reduced to a single afferent whose spikes encoded the time of the maximum signal for a given frequency channel.

## Speech Databases

We used the digit subset of the TI46 Word speech database [24]. This clear speech dataset comprises 26 isolated utterances of each English digit from zero to nine spoken by 16 adult speakers (eight male and eight female). The data is partitioned into a fixed training set, comprising 10 utterances per digit and speaker, and a fixed test set holding the remaining 16 utterances per digit and speaker. We also tested our neuronal word-recognition model on the adult speaker, isolated-digit subset of the TIDIGITS database [27]. This subset comprises two utterances per digit and speaker, i.e., a total of 20 utterances from 225 adult speakers (111 male and 114 female), that are dialectically balanced across 21 dialectical regions (tiling the continental United States). Because the TI46 database only provides utterances of the word "zero" for the digit 0, we excluded the utterances of "oh" from our TIDIGITS sample.

## Digit Classification

Based on the spiking activity of all binary digit detector neurons, a full digit classifier was implemented by ranking the digit detectors according to their individual task performances. As a result, a given stimulus was classified as the target digit of the most reliable of all responding digit detector neurons. If all neurons remained silent, a stimulus was classified as the target digit of the least reliable neuron.

### Spike-Triggered Target Features

To preserve the timing relations between the learned spectro-temporal features and the target words, we refrained from correcting the spike-triggered stimuli for stimulus autocorrelations [73].

### Speech Task Training

Test errors in the speech tasks were substantially reduced by training with a Gaussian spike jitter with a standard deviation of $\sigma$ added to the input spikes as well a symmetric threshold margin $v$ that required the maximum postsynaptic voltage on target stimuli to exceed $V_{\mathrm{thr}}+v$ and to remain below $V_{\mathrm{thr}}-v$ during null stimuli. Values of $\lambda_{\mathrm{ini}}$, $\mu$, $\sigma$, and $v$ were optimized on a 4-dimensional grid. Because for each grid point, only short runs over maximally 200 cycles were performed, we also varied the mean values of initial Gaussian distributions of the excitatory and inhibitory synaptic peak conductances, keeping their standard deviations fixed at 0.001. The reported performances are based on the solutions that had the smallest errors fractions over the test set. If not unique, we selected the solution with the highest robustness to time warp (cf. Figure 8B). Note that this naive optimization of the training parameters did not maintain a separate holdout test set for cross-validation and might therefore overestimate the true generalization performance. We adopted this optimization scheme from [25,26] to ensure comparability of the resulting performance measures.

### Comparison to the HTK

HTK generalization performance was tested with the HTK package version 3.4.1 [74] with front-end and HMM model parameters following [26]. Specifically, speech data from the TI46 and TIDIGITS databases were converted to 13 Mel-cepstral coefficients (including the *0th* order coefficient) along with their first and second derivatives at a frame rate of 5 ms. Mel-coefficients were computed over 30 channels in 25-ms windows with zero mean normalization enabled (TARGETKIND = MFCC_D_A_Z_0). In addition, the following parameters were set: USEHAMMING = T; PREEMPCOEF = 0.97; and CEPLIFTER = 22. Ten HMM models, one for each digit plus one HMM model for silence, were used. Each model consisted of five states (including the the two terminal states) with eight Gaussian mixtures per state and left-to-right (no skip) transition topology.

## Supporting Information

**Figure S1 Effect of h-conductance on time rescaling.** Time-warp distortion index computed for random latency patterns (see Materials and Methods) versus the maximal h-conductance for different values of the mean synaptic conductance $\overline{G}_{\mathrm{syn}}/g_{\mathrm{leak}}$: 7.2 (triangles), 10.8 (squares), and 14.4 (circles). Curves were averaged over 2,000 spike-pattern realizations.
Found at: doi:10.1371/journal.pbio.1000141.s001 (0.70 MB TIF)

**Table S1 Generalization from TI46 to TIDIGITS.** For each dialect group, the table lists the percentages of speakers for which our model committed a given number of word-recognition errors.
Found at: doi:10.1371/journal.pbio.1000141.s002 (0.01 MB PDF)

**Text S1 Comparison to the Hopfield-Brody model of time-warp–invariant neuronal processing.**
Found at: doi:10.1371/journal.pbio.1000141.s003 (0.03 MB PDF)

## Acknowledgments

We thank C. Brody, D. Buonomano, D. Hansel, M. Kilgard, M. London, M. Merzenich, J. Miller, I. Nelken, A. Roth, S. Shamma, and M. Tsodyks for discussions and comments.

## Author Contributions

The author(s) have made the following declarations about their contributions: Conceived and designed the experiments: RG HS. Performed the experiments: RG HS. Analyzed the data: RG HS. Contributed reagents/materials/analysis tools: RG HS. Wrote the paper: RG HS.

## References

1. Sakoe H, Chiba S (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 26: 43–49.
2. Miller JL (1981) Effects of speaking rate on segmental distinctions. In: Eimas PD, Miller JL, eds. Perspectives on the study of speech. Hilsdale (New Jersey): Lawrence Erlbaum Associates. pp 39–74.
3. Anderson S, Dave A, Margoliash D (1996) Template-based automatic recognition of birdsong syllables from continuous recordings. J Acoust Soc Am 100: 1209–1219.
4. Hopfield JJ (1996) Transforming neural computations and representing time. Proc Natl Acad Sci U S A 93: 15440–15444.
5. Hopfield JJ, Brody CD (2000) What is a moment? "Cortical" sensory integration over a brief interval. Proc Natl Acad Sci U S A 97: 13919–131924.
6. Hopfield JJ, Brody CD (2001) What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. Proc Natl Acad Sci U S A 98: 1282–1287.
7. Herz AVM (2005) How is time represented in the brain? In: van Hemmen J, Sejnowski T, eds. 23 problems in systems neuroscience. Oxford (United Kingdom): Oxford University Press. pp 266–283.
8. Brown J, Miller P (2007) Automatic classification of killer whale vocalizations using dynamic time warping. J Acoust Soc Am 122: 1201–1207.
9. Gollisch T (2008) Time-warp invariant pattern detection with bursting neurons. New J Phys 10: 015012.
10. Shannon R, Zeng F, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270: 303–304.
11. Merzenich M, Jenkins W, Johnston P, Schreiner C, Miller S, et al. (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271: 77–81.
12. Phillips D, Farmer M (1990) Acquired word deafness, and the temporal grain of sound representation in the primary auditory cortex. Behav Brain Res 40: 85–94.
13. Fitch RH, Miller S, Tallal P (1997) Neurobiology of speech perception. Annu Rev Neurosci 20: 331–351.
14. Miller JL, Grosjean F, Lomanto C (1984) Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. Phonetica 41: 215–225.
15. Miller JL, Grosjean F, Lomanto C (1986) Speaking rate and segments: a look at the relation between speech production and speech perception for voicing contrast. Phonetica 43: 106–115.
16. Miller JL, Green K, Schermer TM (1984) A distinction between the effects of sentential speaking rate and semantic congruity on word identification. Percept Psychophys 36: 329–337.
17. Miller JL, Aibel IL, Green K (1984) On the nature of rate-dependent processing during phonetic perception. Percept Psychophys 35: 5–15.
18. Newman R, Sawusch J (1996) Perceptual normalization for speaking rate: effects of temporal distance. Percept Psychophys 58: 540–560.
19. Bernander O, Douglas R, Martin K, Koch C (1991) Synaptic background activity influences spatiotemporal integration in single pyramidal cells. Proc Natl Acad Sci U S A 88: 11569–11573.
20. Koch C, Rapp M, Segev I (1996) A brief history of time (constants). Cereb Cortex 6: 93–101.
21. Häusser M, Clark BA (1997) Tonic synaptic inhibition modulates neuronal output pattern and spatiotemporal synaptic integration. Neuron 19: 665–678.
22. Gütig R, Sompolinsky H (2006) The tempotron: a neuron that learns spike timing-based decisions. Nat Neurosci 9: 420–428.

23. Hopfield JJ (2004) Encoding for computation: recognizing brief dynamical patterns by exploiting effects of weak rhythms on action-potential timing. Proc Natl Acad Sci U S A 101: 6255–6260.
24. Liberman M, Amsler R, Church K, Fox E, Hafner C, et al. (1993) TI 46-Word. Philadelphia (Pennsylvania): Linguistic Data Consortium.
25. Walker W, Lamere P, Kwok P, Raj B, Singh R, et al. (2004) Sphinx-4: a flexible open source framework for speech recognition. Technical Report SMLI TR-2004-139. Menlo Park (California): Sun Microsystems Laboratories. pp 1–15.
26. Deshmukh O, Espy-Wilson C, Juneja A (2002) Acoustic-phonetic speech parameters for speaker-independent speech recognition. In: Proceedings of IEEE ICASSP 2002; 13–17 May 2002. Orlando, Florida, United States. pp 593–596.
27. Leonard R, Doddington G (1993) TIDIGITS. Philadelphia (Pennsylvania): Linguistic Data Consortium.
28. Destexhe A, Rudolph M, Paré D (2003) The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci 4: 739–751.
29. Zhang L, Tan A, Schreiner C, Merzenich M (2003) Topography and synaptic shaping of direction selectivity in primary auditory cortex. Nature 424: 201–205.
30. Wehr M, Zador A (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426: 442–446.
31. Borg-Graham L, Monier C, Frégnac Y (1998) Visual input evokes transient and strong shunting inhibition in visual cortical neurons. Nature 393: 369–373.
32. Hirsch JA, Alonso JM, Reid R, Martinez L (1998) Synaptic integration in striate cortical simple cells. J Neurosci 18: 9517–9528.
33. Shu Y, Hasenstaub A, McCormick DA (2003) Turning on and off recurrent balanced cortical activity. Nature 423: 288–293.
34. Haider B, Duque A, Hasenstaub AR, McCormick DA (2006) Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. J Neurosci 26: 4535–4545.
35. Waters J, Helmchen F (2006) Background synaptic activity is sparse in neocortex. J Neurosci 26: 8267–8277.
36. Major G, Larkman A, Jonas P, Sakmann B, Jack J (1994) Detailed passive cable models of whole-cell recorded ca3 pyramidal neurons in rat hippocampal slices. J Neurosci 14: 4613–4638.
37. Roth A, Häusser M (2001) Compartmental models of rat cerebellar purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. J Physiol 535: 445–572.
38. Sarid L, Bruno R, Sakmann B, Segev I, Feldmeyer D (2007) Modeling a layer 4-to-layer 2/3 module of a single column in rat neocortex: interweaving in vitro and in vivo experimental observations. Proc Natl Acad Sci U S A 104: 16353–16358.
39. Oswald A, Reyes A (2008) Maturation of intrinsic and synaptic properties of layer 2/3 pyramidal neurons in mouse auditory cortex. J Neurophysiol 99: 2998–3008.
40. Froemke R, Merzenich M, Schreiner C (2007) A synaptic memory trace for cortical receptive field plasticity. Nature 450: 425–429.
41. Froemke R, Dan Y (2002) Spike-timing-dependent synaptic modification induced by natural spike trains. Nature 416: 433–438.
42. Wang HX, Gerkin RC, Nauen DW, Bi GQ (2005) Coactivation and timing-dependent integration of synaptic potentiation and depression. Nat Neurosci 8: 187–193.
43. Froemke R, Tsay I, Raad M, Long J, Dan Y (2006) Contribution of individual spikes in burstinduced long-term synaptic modification. J Neurophysiol 95: 1620–1629.
44. Wittenberg G, Wang S (2006) Malleability of spike-timing-dependent plasticity at the ca3-ca1 synapse. J Neurosci 26: 6610–6617.
45. Zatorre R, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11: 946–953.
46. Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8: 389–395.
47. Abrams D, Nicol T, Zecker S, Kraus N (2008) Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J Neurosci 28: 3958–3965.
48. Hutsler J, Galuske R (2003) Hemispheric asymmetries in cerebral cortical networks. Trends Neurosci 26: 429–435.
49. Shtyrov Y, Kujala T, Ahveninen J, Tervaniemi M, Alku P, et al. (1998) Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. Neurosci Lett 251: 141–144.
50. Abrams DA, Nicol T, Zecker SG, Kraus N (2006) Auditory brainstem timing predicts cerebral asymmetry for speech. J Neurosci 26: 11131–11137.
51. Casseday JH, Fremouw T, Covey E (2002) The inferior colliculus: a hub for the central auditory system. In: Oertel D, Fay R, Popper A, eds. Integrative functions in the mammalian auditory pathway. New York (New York): Springer. pp 238–318.
52. Jusczyk P, Pisoni D, Reed M, Fernald A, Myers M (1983) Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals. Science 222: 175–177.
53. Eimas PD, Miller JL (1980) Contextual effects in infant speech perception. Science 209: 1140–1141.
54. Gordon-Salant S, Fitzgibbons P (2001) Sources of age-related recognition difficulty for timecompressed speech. J Speech Lang Hear Res 44: 709–719.
55. Gordon-Salant S, Fitzgibbons P, Friedman S (2007) Recognition of time-compressed and natural speech with selective temporal enhancements by young and elderly listeners. J Speech Lang Hear Res 50: 1181–1193.
56. Caspary D, Schatteman T, Hughes L (2005) Age-related changes in the inhibitory response properties of dorsal cochlear nucleus output neurons: role of inhibitory inputs. J Neurosci 25: 10952–10959.
57. Caspary DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. J Exp Biol 211(Pt 11): 1781–1791.
58. Schneider BA, Daneman M, Murphy DR (2005) Speech comprehension difficulties in older adults: cognitive slowing or age-related changes in hearing? Psychol Aging 20: 261–271.
59. Itakura F (1975) Minimum prediction residual principle applied to speech recognition. IEEE Trans Acoust Speech Signal Proc ASSP-23: 67–72.
60. Myers C, Rabiner L, Rosenberg A (1980) Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. IEEE Acoust Speech Signal Process ASSP-28: 623–635.
61. Kavaler RA, Lowy M, Murveit H, Brodersen RW (1987) A dynamic-time-warp integrated circuit for a 1000-word speech recognition system. IEEE J Solid-State Circuits 22: 3–14.
62. Mauk M, Buonomano D (2004) The neural basis of temporal processing. Annu Rev Neurosci 27: 307–340.
63. Meister M, Lagnado L, Baylor DA (1995) Concerted signaling by retinal ganglion cells. Science 270: 1207–1210.
64. Neuenschwander S, Singer W (1996) Long-range synchronization of oscillatory light responses in the cat retina and lateral geniculate nucleus. Nature 379: 728–732.
65. Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. Science 319: 1108–1111.
66. deCharms RC, Merzenich MM (1996) Primary cortical representation of sounds by the coordination of action-potential timing. Nature 381: 610–613.
67. Johansson RS, Birznieks I (2004) First spikes in ensembles of human tactile afferents code complex spatial fingertip events. Nat Neurosci 7: 170–177.
68. Wehr M, Laurent G (1996) Odour encoding by temporal sequences of firing in oscillating neural assemblies. Nature 384: 162–166.
69. Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. Neuron 29: 145–156.
70. Ji D, Wilson MA (2007) Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat Neurosci 10: 100–107.
71. Brette R (2006) Exact simulation of integrate-and-fire models with synaptic conductances. Neural Computat 18: 2004–2027.
72. Dickson CT, Magistretti J, Shalinsky MH, Fransén E, Hasselmo ME, et al. (2000) Properties and role of Ih in the pacing of subthreshold oscillations in entorhinal cortex layer II neurons. J Neurophysiol 83: 2562–2579.
73. Klein DJ, Depireux DA, Simon JZ, Shamma SA (2000) Robust spectrotemporal reverse correlation for the auditory system: optimizing stimulus design. J Comput Neurosci 9: 85–111.
74. Woodland P (2009) Htk3. Available: http://htk.eng.cam.ac.uk. Accessed 1 April 2009.