

Neuromorphic computing with nanoscale spintronic oscillators

Jacob Torrejon¹, Mathieu Riou¹, Flavio Abreu Araujo¹, Sumito Tsunegi², Guru Khalsa^{3†}, Damien Querlioz⁴, Paolo Bortolotti¹, Vincent Cros¹, Kay Yakushiji², Akio Fukushima², Hitoshi Kubota², Shinji Yuasa², Mark D. Stiles³ & Julie Grollier¹

Neurons in the brain behave as nonlinear oscillators, which develop rhythmic activity and interact to process information¹. Taking inspiration from this behaviour to realize high-density, low-power neuromorphic computing will require very large numbers of nanoscale nonlinear oscillators. A simple estimation indicates that to fit 10^8 oscillators organized in a two-dimensional array inside a chip the size of a thumb, the lateral dimension of each oscillator must be smaller than one micrometre. However, nanoscale devices tend to be noisy and to lack the stability that is required to process data in a reliable way. For this reason, despite multiple theoretical proposals^{2–5} and several candidates, including memristive⁶ and superconducting⁷ oscillators, a proof of concept of neuromorphic computing using nanoscale oscillators has yet to be demonstrated. Here we show experimentally that a nanoscale spintronic oscillator (a magnetic tunnel junction)^{8,9} can be used to achieve spoken-digit

recognition with an accuracy similar to that of state-of-the-art neural networks. We also determine the regime of magnetization dynamics that leads to the greatest performance. These results, combined with the ability of the spintronic oscillators to interact with each other, and their long lifetime and low energy consumption, open up a path to fast, parallel, on-chip computation based on networks of oscillators.

Nanoscale spintronic oscillators (or spin-torque nano-oscillators) are nanoscale pillars composed of two ferromagnetic layers separated by a non-magnetic spacer (Fig. 1a). Charge currents become spin-polarized when they flow through these junctions and generate torques on the magnetizations^{10,11} that lead to sustained magnetization precession at frequencies of hundreds of megahertz to several tens of gigahertz. Magnetization oscillations are converted into voltage oscillations through magneto-resistance. The resulting radio-frequency oscillations, of up to tens of millivolts (ref. 12), can be detected by measuring

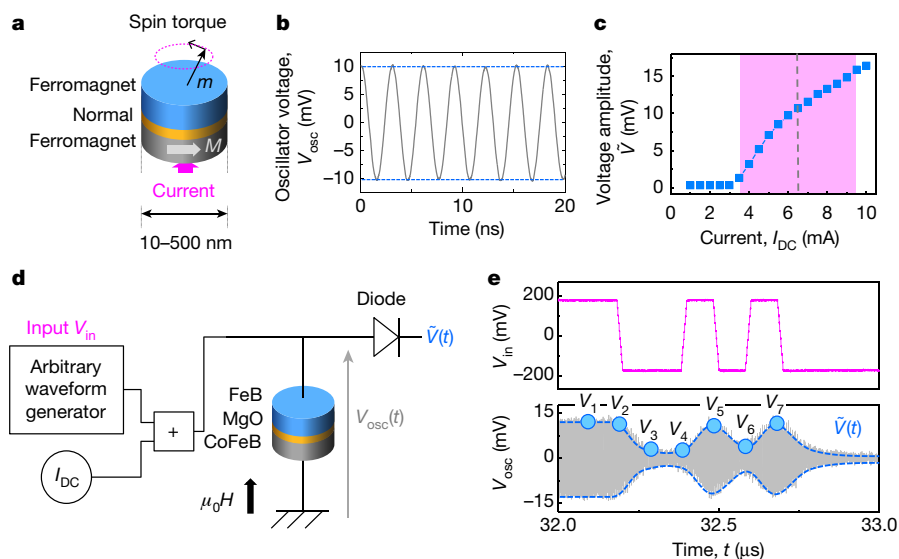


Figure 1 | Spin-torque nano-oscillator for neuromorphic computing. **a**, Schematic of a spin-torque nano-oscillator, consisting of a non-magnetic spacer (gold) between two ferromagnetic layers, with magnetization m for the free layer (blue) and M for the fixed layer (silver). A current injected into the oscillator induces magnetization precessions of m . For our experiments we used a nano-oscillator with a diameter of 375 nm; however, diameters of 10–500 nm are possible. **b**, Measured a.c. voltage emitted by the oscillator as a function of time, $V_{\text{osc}} = \tilde{V}(t)\cos(\omega t + \varphi)$, for a steady current injection of 7 mA at an external magnetic field $\mu_0 H = 430$ mT. The dotted blue lines highlight the amplitude \tilde{V} . **c**, Voltage amplitude \tilde{V} as a function of d.c. current I_{DC} at $\mu_0 H = 430$ mT (blue squares). The purple shaded area highlights the typical excursion in the

voltage amplitude that results when an input signal of $V_{\text{in}} = \pm 250$ mV is injected (here for $I_{\text{DC}} = 6.5$ mA (vertical dotted line) and $\mu_0 H = 430$ mT). **d**, Schematic of the experimental set-up. A d.c. current I_{DC} and a rapidly varying waveform that encodes the input V_{in} are injected into the spin-torque nano-oscillator. The microwave voltage V_{osc} emitted by the oscillator in response to the excitation is measured with an oscilloscope. For computing, the amplitude \tilde{V} of the oscillator is used, and measured directly with a microwave diode. **e**, Input V_{in} (top; magenta) and measured microwave voltage V_{osc} (bottom; grey) emitted by the oscillator as a function of time. Here $I_{\text{DC}} = 6$ mA and $\mu_0 H = 430$ mT. The envelope \tilde{V} of the oscillator signal is highlighted in blue. For computing it is sampled periodically, as shown by the blue circles labelled V_{1-7} .

¹Unité Mixte de Physique, CNRS, Thales, Université Paris-Sud, Université Paris-Saclay, 91767 Palaiseau, France. ²National Institute of Advanced Industrial Science and Technology (AIST), Spintronics Research Center, Tsukuba, Ibaraki 305-8568, Japan. ³Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-6202, USA. ⁴Centre de Nanosciences et de Nanotechnologies, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay, France. [†]Present address: Cornell University, Department of Materials Science and Engineering, Ithaca, New York 14853-1501, USA.

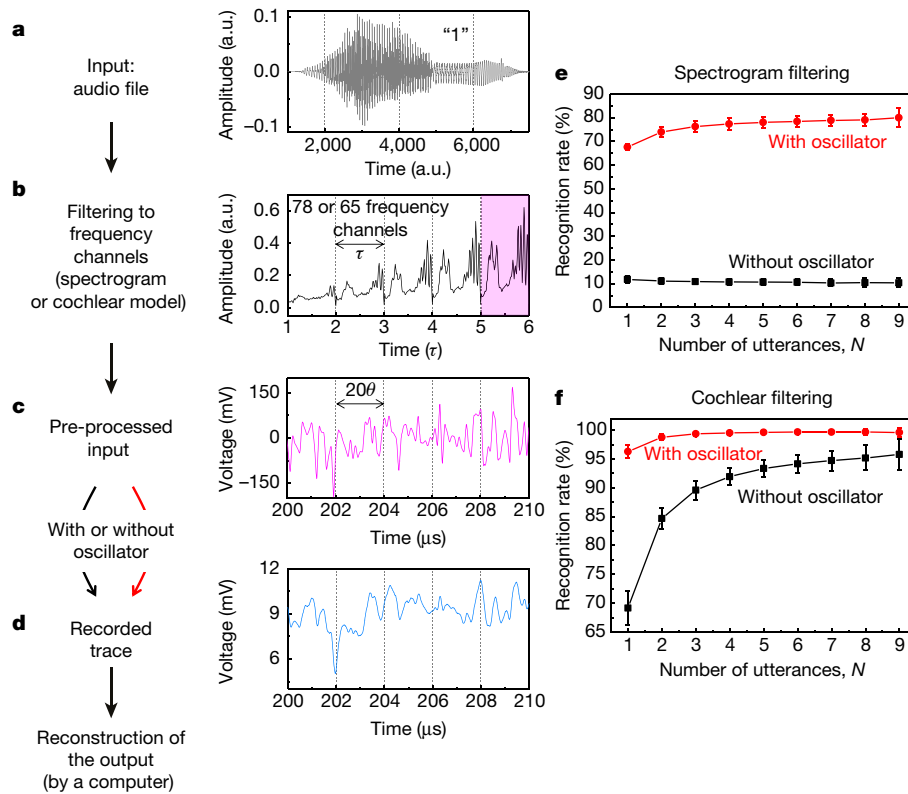


Figure 2 | Spoken-digit recognition. a–d, Principle of the experiment. a, Audio waveform corresponding to the digit 1 pronounced by speaker 1. b, Filtering to frequency channels for acoustic feature extraction. The audio waveform is divided in intervals of duration τ . The cochlear model filters each interval into 78 frequency channels (65 for the spectrogram model), which are then concatenated as 78 (65) values for each interval, to form the filtered input. c, Pre-processed input (transformed from the purple shaded region in b). The filtered input is multiplied by a randomly filled binary matrix (masking process), resulting in 400 points separated by a time step θ of 100 ns in each interval of duration τ ($\tau = 400\theta$). d, Oscillator output. The envelope $\tilde{V}(t)$ of the emitted voltage amplitude of the experimental oscillator is shown ($\mu_0H = 430$ mT, $I_{DC} = 6$ mA). The 400 values of $\tilde{V}(t)$ per interval τ (\tilde{V}_i , sampled with a time step θ) emulate 400 neurons. The reconstructed output ‘1’, corresponding to this digit, is

the voltage across the junction (Fig. 1b). Spin-torque nano-oscillators are therefore simple and ultra-compact: their lateral size can be scaled down to 10 nm and their power consumption reduced to 1 μ W (ref. 13). Because they have the same structure as present-day magnetic memory cells, they are compatible with complementary metal–oxide–semiconductor (CMOS) technology, have high endurance, operate at room temperature and can be fabricated in large numbers (currently up to hundreds of millions) on a single chip¹⁴. Just as the frequency of a neuron is modified by the spikes received from other neurons, the frequencies of spin-torque nano-oscillators are highly sensitive to the magnetization dynamics of neighbouring oscillators to which they are coupled^{15,16}. Together, these features of spin-torque nano-oscillators make them promising candidates for use in neuromorphic computing with large arrays of coupled oscillators^{17–21}. However, they have yet to be used to perform an actual computing task.

Our idea is to exploit the amplitude dynamics of spin-torque nano-oscillators for neuromorphic computing. Their oscillation amplitude \tilde{V} (dotted blue line in Fig. 1b) is robust to noise, owing to the confinement that is provided by the counteracting torques exerted by the injected current and magnetic damping²². In addition, \tilde{V} is highly nonlinear as a function of the injected current and depends intrinsically on past inputs¹⁵. Exploiting the amplitude dynamics of spin-torque nano-oscillators thus combines in one single nanodevice the two most crucial properties of neurons—nonlinearity and memory—the

obtained by linearly combining the 400 values of \tilde{V}_i , sampled from each interval τ . e, f, Spoken-digit recognition rates in the testing set as a function of the number of utterances N used for training for the spectrogram filtering (e; $\mu_0H = 430$ mT, $I_{DC} = 6$ mA) and for the cochlear filtering (f; $\mu_0H = 448$ mT, $I_{DC} = 7$ mA). Because there are many ways to pick the N utterances, the recognition rate is an average over all $10!/[(10-N)!N!]$ combinations of N utterances out of the 10 in the dataset. The red curves are the experimental results using the magnetic oscillator. The black curves are control trials, in which the pre-processed inputs are used for reconstructing the output on a computer directly, as described in Methods, without going through the experimental set-up. The error bars correspond to the standard deviation of the recognition rate, based on training with all possible combinations.

realization of which would otherwise require several electronics components and a much larger on-chip area using conventional CMOS²³. To compute, we encode neural inputs in the time-dependent current $I(t)$ that is injected into the oscillator and use the amplitude response $\tilde{V}(t)$ as the neural output.

Our nano-oscillators consist of circular magnetic tunnel junctions, with a 6-nm-thick free layer of FeB of 375-nm diameter, which have magnetic vortex ground states (see Methods). We measure the dynamics of the signal amplitude $\tilde{V}(t)$ directly using a microwave diode. In Fig. 1c we show the nonlinear response of the amplitude \tilde{V} to a d.c. current I_{DC} : $\tilde{V} \propto \sqrt{I_{DC} - I_{th}}$, where I_{th} is the current threshold for steady oscillations to occur¹⁵. Using an arbitrary waveform generator, we inject a varying current through the junctions in addition to the d.c. current, using the set-up schematized in Fig. 1d. The resulting voltage oscillations, recorded with an oscilloscope, are shown in Fig. 1e. The amplitude of the oscillator varies in response to the injected d.c. current, with a relaxation time that induces a few hundred nanoseconds memory of past inputs²².

Recent studies have revealed that time-multiplexing can enable a single oscillator to emulate a full neural network^{24–26}. Here we use this approach—a form of “reservoir computing”^{24,5} (see Methods)—to demonstrate the ability of spin-torque nano-oscillators to realize neuromorphic tasks. We perform a benchmark task of spoken-digit recognition. The input data, taken from the TI-46 database²⁷, are

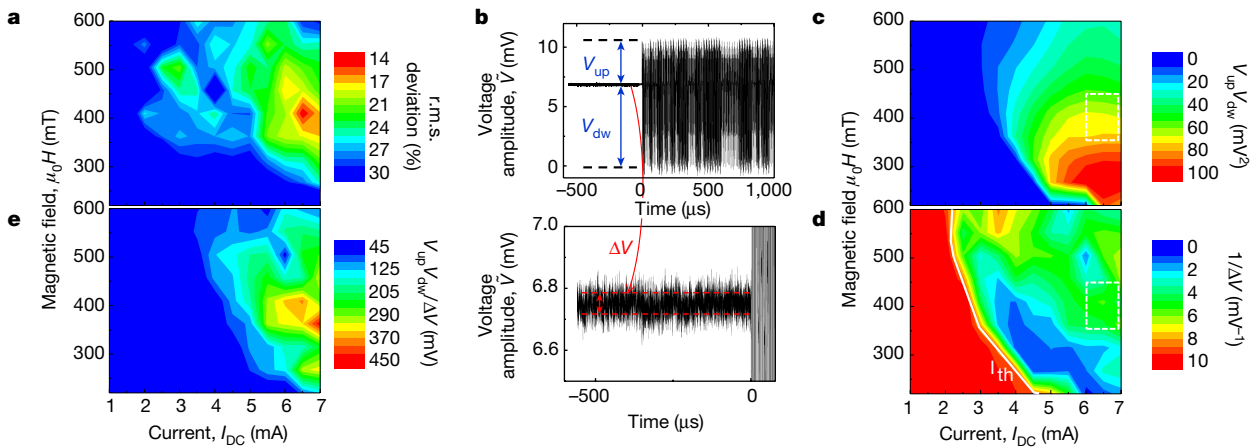


Figure 3 | Conditions for optimal waveform classification and identification of important oscillator properties. The task consists of recognizing sine waveforms from square ones with the same period. The target for the output that is reconstructed from the oscillator's response is one for square, zero for sine. We emulate 24 neurons \tilde{V}_i , $\tau = 24\theta$. **a**, Root-mean-square (r.m.s.) deviation of output-to-target deviations: map as a function of d.c. current I_{DC} and magnetic field $\mu_0 H$. **b**, Extraction of parameters from the time traces of the oscillator's response. Top, maximum positive (V_{up}) and negative (V_{dw}) variations in the oscillator's amplitude in response to the varying pre-processed input. Bottom, noise

ΔV of the voltage amplitude \tilde{V} at steady state under I_{DC} . **c**, Maximal response ($V_{up} V_{dw}$) of the oscillator to the input: map in the I_{DC} - $\mu_0 H$ plane. **d**, Inverse of the noise amplitude $1/\Delta V$: map in the I_{DC} - $\mu_0 H$ plane. The threshold current I_{th} is indicated by a white solid line. In **c** and **d**, the optimal range of bias conditions for waveform classification is marked by a white dashed rectangle (currents of 6–7 mA and magnetic fields of 350–450 mT). **e**, Map of the ratio of maximal amplitudes to noise $V_{up} V_{dw} / \Delta V$, showing that these parameters largely determine the performance of the oscillator (compare with **a**).

audio waveforms of isolated spoken digits (0 to 9) pronounced by five different female speakers (Fig. 2a). The goal is to recognize the digits, independent of the speaker.

Neural networks classify information through chain reactions: neuron after neuron, each input undergoes a series of nonlinear transformations²⁸. In a trained network, the same digit always triggers a similar chain reaction even if it is pronounced by different speakers, whereas different digits generate different chain reactions, thus allowing pattern recognition. An input can trigger a chain reaction in space by using ensembles of neurons, wherein the state of downstream neurons depends on the state of upstream neurons. But an input can also trigger a chain reaction in time by constantly exciting a single nonlinear oscillator with memory: in this case, the state of the oscillator in the future depends on the state of the oscillator in the past. We use the latter approach, which simplifies the hardware because only one oscillator is needed, but requires preprocessing of the input: each point of the audio waveform is converted into a fast-paced binary sequence that is designed to generate a chain reaction of amplitude variations in the oscillator²⁴.

The procedure is illustrated in Fig. 2a–d and detailed in Methods. Because acoustic features are mainly encoded in frequencies²⁹, we filter each audio file into N_f different frequency channels (a standard procedure in speech recognition), which are then concatenated in intervals of duration τ (Fig. 2b). For preprocessing, each of these segments is multiplied by a randomly filled binary matrix (of dimension $N_f \times N_\theta$). In this way, each point of the input audio waveform is converted into a binary sequence of duration τ that is composed of N_θ points separated by a time step θ ($\tau = N_\theta \theta$). When this preprocessed input (Fig. 2c) is applied as a current to our spin-torque nano-oscillator, the resulting amplitude variations $\tilde{V}(t)$ (Fig. 2d) function as a set of N_θ neurons coupled in time (we take N_θ samples \tilde{V}_i per interval τ). For spoken-digit recognition, we emulate $N_\theta = 400$ neurons and use $\theta = 100$ ns (about one-fifth of the relaxation time of the oscillators) to set the oscillator in a transient state.

The responses of the voltage amplitude $\tilde{V}(t)$ of the oscillator are recorded for each utterance of each spoken digit. The goal of the subsequent training process, performed on a computer, is to choose a linear combination of these responses (sets of \tilde{V}_i in each τ) for each digit such that the sum is one for that digit and zero for the rest

(see Methods). Because each digit has been pronounced ten times by each of the five speakers, we can use some of the data to determine the coefficients (training), and the rest to evaluate the recognition performance (testing); see Methods. To assess the effect of our oscillator on the quality of recognition, we always perform a control trial without the oscillator. In that case, the preprocessed input traces are used to reconstruct the outputs on the computer directly, without going through the experimental set-up.

The improvement shown in the experimental results over the control results (see Fig. 2e, f) indicates that the spin-torque nano-oscillator greatly improves the quality of spoken-digit recognition, despite the added noise that is concomitant to its nanometre-scale size. In Fig. 2e (linear spectrogram filtering), we present an example in which the extraction of acoustic features, achieved by Fourier transforming the audio waveform over finite time windows, plays a minimal part in classification. Without the oscillator (black line), the recognition rates are consistent with random choices; with the oscillator (red line), the recognition rate is improved by 70%, reaching values of up to 80%. This example highlights the crucial role of the oscillator in the recognition process. Using nonlinear cochlear filtering³⁰ (Fig. 2f), which is the standard in reservoir computing^{24–26} and has been optimized on the basis of the behaviour of biological ears, we achieve recognition rates of up to 99.6%, as high as the state-of-the-art. Compared to the control trial, the oscillator reduces the error rate by a factor of up to 15. Our results with a spin-torque nano-oscillator are therefore comparable to the recognition rates obtained with more complicated electronic or optical systems (between 95.7% and 99.8% for the same task with cochlear filtering)^{23–26,29}.

The optimal operating conditions for pattern recognition with our spin-torque nano-oscillator are determined by the oscillation amplitude and noise. We use a simpler task, classification of sine and square waveforms with the same period²⁵, to investigate the ability of the oscillator to classify waveforms in a wide range of injected d.c. currents I_{DC} and applied magnetic fields $\mu_0 H$ (see Methods). As can be seen in Fig. 3a, the quality of pattern recognition, characterized by the root-mean-square of deviations between the reconstructed output and the target, varies from 10% to more than 30% depending on the bias conditions. The oscillator performs well when it responds strongly to the time-varying preprocessed input, with large amplitude variations in both the

positive and negative directions, V_{up} and V_{dw} respectively (Fig. 3b, top). On the other hand, it performs poorly when the noise in the oscillator ΔV (the standard deviation of the noise in the voltage amplitude) is high (Fig. 3b, bottom). As shown in Fig. 3b, we extract these parameters from the time traces of the voltage emitted from the oscillator at each bias point, and plot $V_{\text{up}}V_{\text{dw}}$ (Fig. 3c) and $1/\Delta V$ (Fig. 3d) as a function of the d.c. current I_{DC} and field $\mu_0 H$. The red regions of large oscillation amplitudes in Fig. 3c correspond to low magnetic fields, in which the magnetization is weakly confined, and to high currents, for which the spin torque on magnetization is maximal. The blue regions of high noise in Fig. 3d correspond to areas just above the threshold current I_{th} for oscillation, in which the oscillation amplitude \bar{V} is growing rapidly as a function of current and is becoming sensitive to external fluctuations¹⁵. As can be seen by comparing Fig. 3c and d, the range of bias conditions highlighted by the dotted white boxes (currents of 6–7 mA and magnetic fields of 350–450 mT) features wide variations in oscillation amplitudes and low noise. In this region, root-mean-square deviations below 15% are achieved, and there are no classification errors between sine and square waveforms. The similarity between the map of $V_{\text{up}}V_{\text{dw}}/\Delta V$ (Fig. 3e) and that of the classification performance (Fig. 3a) confirms that the best conditions for classification correspond to regions of optimal compromise between low noise and large amplitude variations. The necessity of a high signal-to-noise ratio for efficient neuromorphic computing, highlighted here for magnetic oscillators, is a general guideline that applies to any type of nanoscale oscillator.

As a conclusion, our pattern-recognition results show that simple, ultra-compact spintronic oscillators have all of the properties that are needed to emulate collections of neurons: nonlinearity, memory and stability. The ability of groups of these oscillators to mimic neural connections by influencing the behaviour of one another through current and magnetic-field coupling opens up a route to realizing large-scale neural networks in hardware, which exploit magnetization dynamics for computing^{15–21}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 January; accepted 2 June 2017.

- Buzsaki, G. *Rhythms of the Brain* (Oxford Univ. Press, 2011).
- Hoppensteadt, F. C. & Izhikevich, E. M. Oscillatory neurocomputers with dynamic connectivity. *Phys. Rev. Lett.* **82**, 2983–2986 (1999).
- Aonishi, T., Kurata, K. & Okada, M. Statistical mechanics of an oscillator associative memory with scattered natural frequencies. *Phys. Rev. Lett.* **82**, 2800–2803 (1998).
- Jaeger, H. & Haas, H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).
- Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
- Pickett, M. D., Medeiros-Ribeiro, G. & Williams, R. S. A scalable neuristor built with Mott memristors. *Nat. Mater.* **12**, 114–117 (2013).
- Segall, K. *et al.* Synchronization dynamics on the picosecond time scale in coupled Josephson junction neurons. *Phys. Rev. E* **95**, 032220 (2017).
- Kiselev, S. I. *et al.* Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425**, 380–383 (2003).
- Rippard, W. H., Pufall, M. R., Kaka, S., Russek, S. E. & Silva, T. J. Direct-current induced dynamics in $\text{Co}_{90}\text{Fe}_{10}/\text{Ni}_{80}\text{Fe}_{20}$ point contacts. *Phys. Rev. Lett.* **92**, 027201 (2004).
- Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
- Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
- Tsunegi, S., Yakushiji, K., Fukushima, A., Yuasa, S. & Kubota, H. Microwave emission power exceeding 10 μW in spin torque vortex oscillator. *Appl. Phys. Lett.* **109**, 252402 (2016).
- Sato, H. *et al.* Properties of magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure down to junction diameter of 11 nm. *Appl. Phys. Lett.* **105**, 062403 (2014).
- Apalkov, D., Dieny, B. & Slaughter, J. M. Magnetoresistive random access memory. *Proc. IEEE* **104**, 1796–1830 (2016).
- Slavin, A. & Tiberkevich, V. Nonlinear auto-oscillator theory of microwave generation by spin-polarized current. *IEEE Trans. Magn.* **45**, 1875–1918 (2009).
- Houshang, A. *et al.* Spin-wave-beam driven synchronization of nanocontact spin-torque oscillators. *Nat. Nanotechnol.* **11**, 280–286 (2016).
- Macià, F., Kent, A. D. & Hoppensteadt, F. C. Spin-wave interference patterns created by spin-torque nano-oscillators for memory and computation. *Nanotechnology* **22**, 095301 (2011).
- Pufall, M. R. *et al.* Physical implementation of coherently coupled oscillator networks. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **1**, 76–84 (2015).
- Nikonov, D. E. *et al.* Coupled-oscillator associative memory array operation for pattern recognition. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **1**, 85–93 (2015).
- Yogendra, K., Fan, D. & Roy, K. Coupled spin torque nano oscillators for low power neural computation. *IEEE Trans. Magn.* **51**, 4003909 (2015).
- Grollier, J., Querlioz, D. & Stiles, M. D. Spintronic nanodevices for bioinspired computing. *Proc. IEEE* **104**, 2024–2039 (2016).
- Grimaldi, E. *et al.* Response to noise of a vortex based spin transfer nano-oscillator. *Phys. Rev. B* **89**, 104404 (2014).
- Soriano, M. C. *et al.* Delay-based reservoir computing: noise effects in a combined analog and digital implementation. *IEEE Trans. Neural Netw. Learn. Syst.* **26**, 388–393 (2015).
- Appeltant, L. *et al.* Information processing using a single dynamical node as complex system. *Nat. Commun.* **2**, 468 (2011).
- Paquot, Y. *et al.* Optoelectronic reservoir computing. *Sci. Rep.* **2**, 287 (2012).
- Martinenghi, R., Rybalko, S., Jacquot, M., Chembo, Y. K. & Larger, L. Photonic nonlinear transient computing with multiple-delay wavelength dynamics. *Phys. Rev. Lett.* **108**, 244101 (2012).
- Texas Instruments. 46-Word Speaker-Dependent Isolated Word Corpus (TI-46), NIST Speech Disc 7-1.1, <https://catalog.ldc.upenn.edu/LDC93S9> (NIST, 1991).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Yildiz, I. B., von Kriegstein, K. & Kiebel, S. J. From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLOS Comput. Biol.* **9**, e1003219 (2013).
- Lyon, R. A computational model of filtering, detection, and compression in the cochlea. in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 82)* Vol. 7, 1282–1285 (IEEE, 1982).

Acknowledgements This work was supported by the European Research Council (ERC) under grant bioSPINspired 682955. We thank L. Larger, B. Penkovsky and F. Duport for discussions.

Author Contributions The study was designed by J.G. and M.D.S., samples were optimized and fabricated by S.T. and K.Y., experiments were performed by J.T. and M.R., numerical studies were realized by F.A.A., M.R. and G.K., and all authors contributed to analysing the results and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.G. (julie.grollier@cncs-thales.fr).

Reviewer Information Nature thanks F. Hoppensteadt and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Samples. Magnetic tunnel junction (MTJ) films with a stacking structure of buffer/PtMn(15)/Co₇₁Fe₂₉(2.5)/Ru(0.9)/Co₆₀Fe₂₀B₂₀(1.6)/Co₇₀Fe₃₀(0.8)/MgO(1)/Fe₈₀B₂₀(6)/MgO(1)/Ta(8)/Ru(7) (with thicknesses given in parentheses in nanometres) were prepared by ultrahigh vacuum (UHV) magnetron sputtering. After annealing at 360 °C for 1 h, the resistance–area products (*RA*) were approximately 3.6 Ω μm². Circular-shape MTJs with a diameter of approximately 375 nm were patterned using Ar ion etching and e-beam lithography. The resistance of the samples is close to 40 Ω and the magneto-resistance ratio is about 135% at room temperature. The FeB layer presents a vortex structure as the ground state for the dimensions used here. In a small region called the core of the vortex, the magnetization spirals out of plane. Under d.c. current injection, the core of the vortex steadily gyrates around the centre of the dot with a frequency in the range 250–400 MHz for the oscillators we consider here. Vortex dynamics driven by spin torque are well understood, well controlled and have been shown to be particularly stable²².

Measurement set-up. The experimental implementation for spoken-digit recognition and sine/square classification tasks is illustrated in Fig. 1d. The pre-processed input signal V_{in} is generated by a high-frequency arbitrary-waveform generator and injected as a current through the magnetic nano-oscillator. The sampling rate of the source is set to 200 MHz (20 points per interval of time θ) for the spoken-digit recognition task and 500 MHz (50 points per interval of time θ) for the classification of sines and squares. The peak-to-peak variation in the input signal is 500 mV, which corresponds to peak-to-peak current variations of 6 mA, as illustrated in Fig. 1c (part of the incoming signal is reflected owing to impedance mismatch). The bias conditions of the oscillator are set by a d.c. current source and an electromagnet that applies a field perpendicular to the plane of the magnetic layers. The oscillating voltage emitted by the nano-oscillator is rectified by a planar tunnel microwave diode, with a bandwidth of 0.1–12.4 GHz and a response time of 5 ns. The input dynamic range of the diode is between 1 μW and 3.15 mW, corresponding to a d.c. output level of 0–400 mV. We use an amplifier to adjust the emitted power of the nano-oscillator to the working range of the diode. The output signal is then recorded by a real-time oscilloscope. In Figs 1b, c, e, 2d and 3b–e, the amplitude of the signal emitted by the oscillator is shown without amplification (the signal measured after the diode has been divided by the total amplification of the circuit, about +21 dB). If, owing to sampling errors, the measured envelope of the oscillators is shifted with respect to the input, classification accuracy can be degraded. We use alignment marks to align our measurements with the input when we reconstruct the output. The alignment precision is ±1 ns.

General concepts of reservoir computing. In machine learning, a reservoir is a network of recurrently and randomly connected nonlinear nodes^{4,5}. When an input signal is injected in the reservoir, it is mapped to a higher-dimensional space in which it can become linearly separable. The key insight behind reservoir computing is that the network does not need any tuning; all connections inside the reservoir are kept fixed. Only external connections (between the reservoir and an output layer) are trained to achieve the desired task.

In other words, reservoir computing requires the generation of complex nonlinear dynamics but, as a trade-off, learning is greatly simplified. For efficient reservoir computing, several requirements related to the dynamical properties of the network should be satisfied. First, different inputs should trigger different dynamics (separation property) and similar inputs should generate similar dynamics (approximation property), enabling efficient classification. Second, the reservoir state should not depend only on present inputs but also on recent past inputs. This short-term memory, called fading memory, is essential for processing temporal sequences for which the history of the signal is important.

A single nonlinear oscillator can emulate a reservoir when it is set in transient dynamics by a rapidly varying input²⁴. The loss of parallelism is compensated by an additional pre-processing input step: the input is multiplied by a rapidly varying mask, which enables virtual nodes to be defined, interconnected in time through the resultant oscillator dynamics. This approach provides a marked simplification of the reservoir scheme for hardware implementations, and has been realized in hardware with optical or electronic oscillators assembled from several components^{23–26}.

Spoken-digit recognition. For this task, the inputs are taken from the NIST TI-46 data corpus²⁷. The input consists of isolated spoken digits said by five different female speakers. Each speaker pronounces each digit ten times. The 500 audio waveforms are sampled at a rate of 12.5 kHz and have variable time lengths.

We used two different filtering methods: spectrogram and cochlear models. Both filters break the word into several time intervals N_r of duration τ and analyse the frequency content in each interval τ through either a Fourier transform (spectrogram model; 65 channels, $N_r \in \{24, \dots, 67\}$; Fig. 2b) or a more complicated nonlinear approach (cochlear model; 78 channels, $N_r \in \{14, \dots, 41\}$). The input for each word is composed of an amplitude for each of the $N_f = 65$ or $N_f = 78$ frequency channels times N_r time intervals. This input is pre-processed by multiplying the

frequency content for each time interval by a mask matrix containing $N_f \times N_\theta$ random binary values, giving a total of $N_r \times N_\theta$ values as input to the oscillator (Fig. 2c). Here, we are modelling $N_\theta = 400$ input neurons, each of which is connected to all of the frequency channels for each time interval.

Each preprocessed input value is consecutively applied to the oscillator as a constant current for a time interval of $\theta \approx 100$ ns, which is about five times shorter than the relaxation time of the oscillator, as recommended in ref. 24. This time is short enough to guarantee that the oscillator is maintained in its transient regime so the emulated neurons are connected to each other, but is long enough to let the oscillator respond to the input excitation. The amplitude of the a.c. voltage across the oscillator is recorded for offline post-processing (Fig. 2d).

The post-processing of the output consists of two distinct steps. The first is called the training (or learning) process and the second is called the classification (or recognition) process. The goal of training is to determine a set of weights $w_{i,\theta}$, where i indexes the desired digit. These weights are used to multiply the output voltages to give $10N_r$ output values, which are then averaged over the N_r time intervals to give 10 output values y_θ , which should ideally be equal to the target values $\tilde{y}_i = 1.0$ for the appropriate digit and 0.0 for the rest. In the training process, a fraction of the utterances are used to train these weights; the rest of the utterances are used in the classification process to test the results.

The optimum weights are found by minimizing the difference between \tilde{y}_i and y_i for all of the words used in the training. In practice, optimal values are determined by using techniques for extracting meaningful eigenvalues from singular matrices such as the linear Moore–Penrose pseudo-inverse operator (denoted by a dagger symbol †). If we consider the target matrix \tilde{Y} , which contains the targets \tilde{y}_i for all of the time steps τ used for the training, and the response matrix S , which contains all neuron responses for all of the time steps τ used for the training, then the matrix W , which contains the optimal weights, is given by $W = \tilde{Y}S^\dagger$. This step is performed on a computer and takes several seconds. In the future, real-time processing on a nanosecond timescale could be realized using fully parallel networks of interacting nano-oscillators.

During the classification phase, the ten reconstructed outputs corresponding to one digit are averaged over all of the time steps τ of the signal, and the digit is identified by taking the maximum value of the ten averaged reconstructed outputs. The averaged reconstructed output that corresponds to the digit in question should be close to 1 and the others should be close to 0. The efficiency of the recognition is evaluated by the word success rate, which is the rate of digits that are correctly identified. The training can be done using more or fewer data (here ‘utterances’). We always trained the system using the ten digits spoken by the five speakers. The only parameter that we changed is the number of utterances used for the training. If we use N utterances for training, then we use the remaining $10 - N$ utterances for testing. However, some utterances are very well pronounced whereas others are hardly distinguishable. As a consequence, the resulting recognition rate depends on which N utterances are picked for training in the set of ten (for example, if $N = 2$, then the utterances picked for training could be the first and second, but also the second and third, or the sixth and tenth, or any other of the $10!/(8!2!)$ combinations of 2 picked out of 10). To avoid this bias, the recognition rates that we present here are the average of the results over all possible combinations. The error bars corresponds to the standard deviation of the word recognition rate. The raw spectrogram is not complex enough to allow a correct reconstruction of the target during the training. Adding the oscillator brings complexity and suppresses this phenomenon.

Sine- and square-wave classification. For this classification task, the input is a random sequence of 160 sines and squares with the same period—the first half of the sequence for training and the second half for classification. Each period is discretized into eight points separated by a time step τ . The pre-processing consists of multiplying the value of each point by the same binary sequence that is generated by a random distribution of +1 and –1 values. In contrast to spoken-digit recognition, the mask is a binary vector (instead of a binary matrix). The fast binary sequence contains 24 values, so 24 neurons \tilde{V}_i are emulated during each time step τ .

The target \tilde{y} for the network output y is 0 for all of the trajectories in response to a sine and 1 for all of the trajectories in response to a square. The best weights are found by linear regression, as explained above for the spoken-digit recognition task. For sine/square recognition, we record five points instead of one for each neuron when we measure the output of the oscillator. During post-processing, we use these additional states between \tilde{V}_i and \tilde{V}_{i+1} to increase the number of coefficients available for solving the problem, and thus increase classification accuracy. In addition, the best performance does not necessarily correspond to a target in exact phase with the oscillator’s output. The standard deviation of the root-mean-square value of $V_{\text{output}} - V_{\text{target}}$, obtained with ten repetitions, is around 1%.

Data availability. The datasets generated and analysed during this study are available from the corresponding author on reasonable request.