# What the Spiking Heidelberg Digits look like

exp059 · 11 July 2026 · Draft

## Abstract

The Spiking Heidelberg Digits (SHD) [1] benchmark delivers spoken digits as cochlear spikes. Unlike the static MNIST images the gamma-gated-sparsity collection Poisson-encodes into spike trains before its conductance-based spiking network (COBANet) ever sees them, SHD arrives *already* as events, with no image and no dense array anywhere: each sample is a spoken digit passed through a model of the inner ear, so it is a list of events, each an event time and the cochlear channel that fired. Before training a network on it, this entry looks at the raw data: one utterance per class, then several utterances of a single digit. Different classes paint visibly distinct time–frequency signatures and repeated utterances of one digit share a family resemblance, so the data is well-behaved and separable enough to be worth training on.

## Methods

Nothing here runs the network: every raster is drawn straight from the raw SHD event lists, with no binning and no model.

1. Load the SHD train split as raw events: $N = 8156$ training utterances (a held-out test split adds more), each a spoken digit recorded from several speakers and converted to spikes by a Cramer et al. [1] cochlea model, a computational model of the inner ear that maps sound onto firing across an array of frequency-tuned channels.
2. Draw the class gallery: the first utterance of each of the 20 classes (the digits 0–9 spoken in German, labels 0–9, then English, labels 10–19), each as a spike raster.
3. Draw the within-class spread: the first 4 utterances of class 0, *null*, side by side.

Every utterance is a set of events $\{(t_k, u_k)\}$, where:

- $t_k$, the **spike time** of event $k$, in seconds (utterances run ≈ 0.2–1.4 s, median ≈ 0.7 s);
- $u_k \in \{0, ..., 699\}$, the **cochlear channel** that fired: a place code for frequency, where a low channel ≈ low pitch and a high channel ≈ high pitch.

So an utterance carries $C = 700$ input channels and a median of ≈ 7605 events per utterance (range ≈ 2410–14917).

## Results

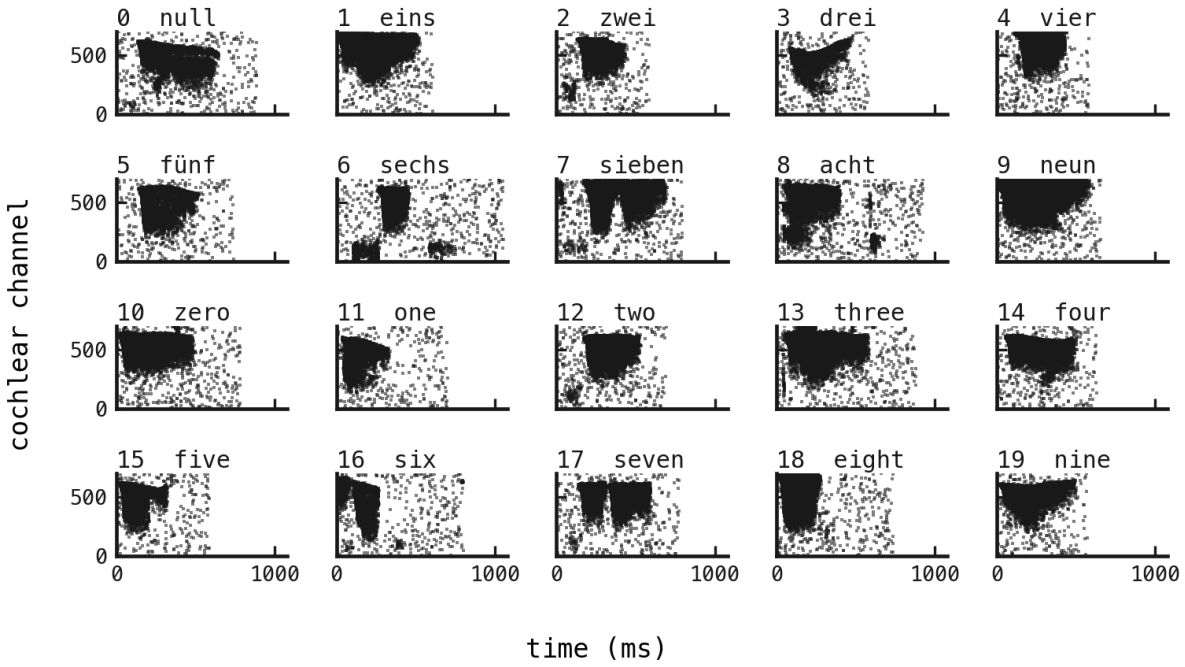


Figure 1: One utterance per class (German 0–9, then English 0–9), each a raster of time (ms) against cochlear channel. **What we expect.** If SHD is learnable, different words should paint different time–frequency signatures. **What we see.** They do: a compact high-channel onset for *zwei*, a long two-lobe sweep for *sieben*/*seven*, a low-channel tail on *sechs*. A diffuse haze of isolated events sits under every panel: the cochlea model's spontaneous background firing, which the classifier has to see past.

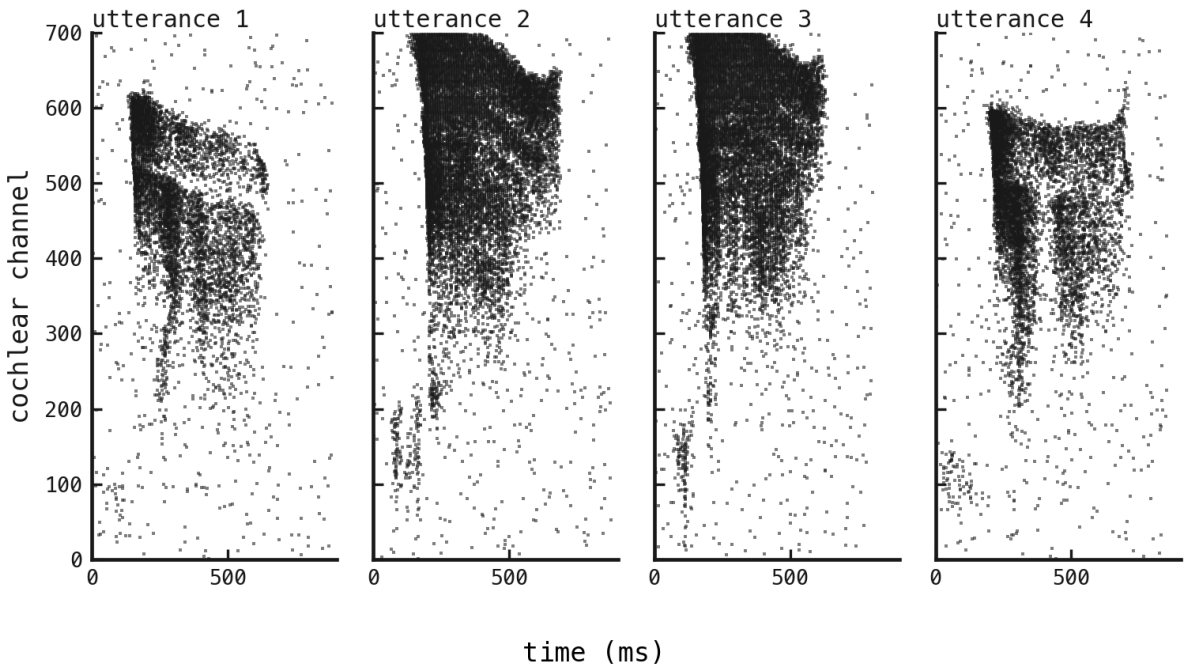


Figure 2: 4 different utterances of class 0, *null*, each a raster of time (ms) against cochlear channel: the within-class spread. **What we expect.** Different speakers saying the same word should share a family resemblance while differing in the particulars. **What we see.** Exactly that: all 4 carry the same broad high-channel onset falling into a mid-channel body, but they differ in duration, event density, and the fine structure of the low-channel tail. This speaker variability is what a classifier trained on SHD must generalise over.

## Next steps

The data is well-behaved and visibly class-separable, so it is worth training on. The trainer accepts SHD directly: it bins these events onto the model's integration grid (a spike tensor of shape $T_{\text{steps}} \times 700$ at the integration timestep $\Delta t$) and feeds them to the PING (pyramidal–interneuron gamma) network, the excitatory and inhibitory populations of the COBANet, where:

- $T_{\text{steps}}$, the number of time bins on the integration grid;
- $\Delta t$, the integration timestep, in seconds.

The next entries take that path: a first PING network trained on SHD, then the firing-rate regulariser (the upper firing-rate bound from Cramer et al. [**1**], with target $\theta_u = 100$ and weight $s_u = 0.06$) that event-based benchmarks need to keep the hidden-layer rates in check, where:

- $\theta_u$, the target upper bound on a unit's firing rate;
- $s_u$, the strength of the penalty applied above that bound.

## References

1. Cramer, Stradmann, Schemmel & Zenke — *The Heidelberg Spiking Data Sets for the Systematic Evaluation of Spiking Neural Networks*. IEEE Transactions on Neural Networks and Learning Systems, 2020. doi:10.1109/TNNLS.2020.3044364