S1 Appendix

Acoustic Space

The general approach for defining acoustic space is as follows. The elements of acoustic space are pure tones of varying frequencies and sound pressures and may be expressed as a product of frequency space $S^F$ and pressure space $S^P$: $S^A = S^F \times S^P$.

The frequency space ($S^F$) is defined as a half-open interval on the positive real line: $S^F = [f_{\text{min}}, f_{\text{max}}) \subset \mathbb{R}^+$, where $f_{\text{min}}, f_{\text{max}}$ are the minimum and maximum audible frequencies, respectively. To account for the fact that discriminability is limited (i.e. tones of similar frequencies are perceived as the same), $S^F$ is partitioned into $N_f$ half-open intervals $h^F = [f_\alpha, f_\beta) = [f_\alpha, f_\alpha + \Delta f)$ where $\Delta f = \frac{1}{N_f}$ is the length of the interval. The use of half-open intervals ensures that there is no ‘gap’ between adjacent frequency intervals. The set of all half-open intervals is:

$$\mathcal{H}^F = \{h^F_i \mid h^F_i = [f_{\text{min}} + i\Delta f, f_{\text{min}} + (i+1)\Delta f), f \in S^F, i \in \mathbb{N}_0, i < N_f\}$$  \hspace{1cm} (1)$$

The set of intervals is a partition where each interval is an equivalence class $[f_j]$; for $f_\alpha, f_\beta \in S^F$, $f_\alpha \sim f_\beta$ if $f_\alpha, f_\beta \in [f_i, f_i + \Delta f)$. This may be extended so that $f_\alpha \sim f_\beta$ if $f_\alpha, f_\beta > f_{\text{max}}$ and $f_\alpha \sim f_\beta$ if $f_\alpha, f_\beta \leq f_{\text{min}}$. Frequencies $f_\alpha, f_\beta$ are ‘assigned’ to $f_i$ if they are between $f_i$ and $f_i + \Delta f$. Thus, $f_\alpha, f_\beta$ are indistinguishable from each other since both are classified as $f_i$. Each frequency interval is uniquely identified by the minimum point at the closed end. The set of these points is:

$$\mathcal{F} = \{f_i \mid f_i = f_{\text{min}} + i\Delta f, i \in \mathbb{N}_0, i < N_f\} \subset S^F$$

$$= \{f_0, f_1, \ldots, f_{N_f-1}\}$$  \hspace{1cm} (2)$$

$\mathcal{F}$ is a discretization of $S^F$ and may be thought of as a set of audible frequencies.
The sound pressure space is constructed similarly to yield:

\[ \mathcal{H}^p = \{ h_i^p | h_i^p = [p_{min} + i\Delta p, p_{min} + (i+1)\Delta p], p \in S^p, i \in \mathbb{N}_0, i < N_p \} \quad (3) \]

\[ \mathcal{P} = \{ p_0, p_1, \ldots, p_{N_p-1} \} \subset S^p \text{ (set of physiological pressures)} \]

If only the sets of audible frequencies and pressures are considered, then the discretized acoustic space is:

\[ \mathcal{A} = \mathcal{F} \times \mathcal{P} \quad (4) \]

**Neural Space**

The neural space \( S^X \) is constructed in a similar fashion as \( \mathcal{F} \) and \( \mathcal{P} \). The idea is to partition \( S^X \) into subintervals that represent neurons and then to combine the subintervals contiguously to form synaptic intervals. To reflect these properties, the neural space will have a topology that is coarser than the usual topology.

The neural space \( S^X \subset \mathbb{R}^+ \) is a half-open interval representing the one dimensional tonotopic axis: \( S^X = [x_{min}, x_{max}) \). \( S^X \) is partitioned into \( N_h \) half-open intervals, each of width \( \Delta x \) and expressed as \( h = [x_\alpha, x_\beta), x_\beta > x_\alpha \) or \( [x_\alpha, x_\alpha + \Delta x) \). Each interval represents the length \( (\Delta x) \) of a single cell on the tonotopic axis. The set of intervals is expressed as:

\[ \mathcal{H}^{X,1} = \{ h_i | h_i = [x_{min} + i\Delta x, x_{min} + (i+1)\Delta x), i \in \mathbb{N}_0, i < N_h \} \quad (5) \]

The set \( \mathcal{H}^{X,1} \) is a partition, which effectively discretizes \( S^X \) (see section above for a similar treatment). Each interval is uniquely identified by the point at the closed end, which also gives it’s location along the tonotopic axis. The set \( X \) containing these points is:

\[ X = \{ x_i | x_i = x_{min} + i\Delta x \mid x_i \in S^X, i \in \mathbb{N}_0, i < N_h \} \]

\[ = \{ x_0, x_1, \ldots, x_{N_h-1} \} \quad (6) \]
A synaptic interval is also a half-open interval consisting of a contiguous set of cells. It’s length \( \lambda \) is an integral number \( (n_\lambda) \) of cells:

\[
h_{x,\lambda} = [x, x + \lambda), \ x \in \mathcal{X}_\lambda, \ 0 < \lambda \leq \lambda_{\text{max}}
\]

\[
= [x, x + n_\lambda \Delta x), \ n_\lambda \in \mathbb{N}^+, \ n_\lambda \leq n_{\lambda,\text{max}} \leq N_h
\]

where \( n_{\lambda,\text{max}} \) is the maximum number of cells allowed in each synaptic interval. To ensure that all synaptic intervals are within \( S^\mathcal{X} \), the starting points \( x \)'s are confined to \( X^\lambda \subset \mathcal{X} \), which has a ‘buffer’ that takes into account the maximum synaptic interval length:

\[
\mathcal{X}^\lambda = \{x_i | i \in \mathbb{N}_0, i \leq (N_h - n_{\lambda,\text{max}})\} \subset \mathcal{X}
\]

The set containing the possible interval lengths is:

\[
\Lambda = \{\lambda_{n_\lambda} | \lambda_{n_\lambda} = n_\lambda \Delta x, n_\lambda \in \mathbb{N}^+, \ n_\lambda \leq n_{\lambda,\text{max}} \leq N_h\} \subset S^\Lambda
\]

\[
= \{\lambda_1, \lambda_2, \ldots, \lambda_{\text{max}}\}
\]

The set of intervals is expressed as:

\[
\mathcal{H}^{\mathcal{X},\Lambda} = \{h_{x,\lambda} | h_{x,\lambda} = [x, x + \lambda), \ x \in \mathcal{X}^\lambda, \ \lambda \in \Lambda\}
\]

A topology \( \tau^\mathcal{X} \) on \( S^\mathcal{X} \) is defined as the collection of open sets generated with \( \mathcal{H}^{\mathcal{X},1} \) (or \( \mathcal{H}^{\mathcal{X},\lambda} \)) as a basis. That is, an element (open set) of \( \tau^\mathcal{X} \) is any set that is constructed from the union of the \( \Delta x \) intervals. For example, an open set can consist of a single interval made up of a contiguous set of cells (eq. 7) or the union of several such intervals. The topology will be used in developing algebraic operations on intervals (see below).
**Subcellular resolution**

To facilitate explanation of the model, the $\Delta x$ was taken to be equal to the width of a single cell. This is true if the divisions of the tonotopic axis are determined by the projections of a single layer of neurons (Fig 2 of main text). Subcellular resolution with a purely place coding scheme may be achieved if the divisions of the tonotopic axis is determined by the projections of multiple layers of neurons (Fig A). If neurons in the different layers are staggered, then there will be more points, with effectively smaller $\Delta x$, along the tonotopic axis that can be used to encode frequency. In this scheme, $\Delta x$, and hence $\Delta f$, can be very small, depending on the neuron density and number of layers.

![Fig A: Attaining subcellular spatial resolution](image)

*Fig A: Attaining subcellular spatial resolution  a, multiple layers of staggered neurons. b, Projections of individual neurons partition the tonotopic axis, with $\Delta x$ corresponding to the distance between the nearest neighbors on tonotopic axis. Cyan bar depicts synaptic interval generated by afferents in a.*

**Mapping single elements in acoustic space to neural space**

The aim is to construct a mapping between the acoustic and neural spaces in such a way that allows for both encoding and decoding *single* tones in neural space.

Note that a pure tone $a_{f,p}$ corresponds to the ordered pair $(f,p) \in S^F \times S^P$ and an interval
\[ h_{x,\lambda} = [x, x + \lambda] \] corresponds to the ordered pair \((x, \lambda) \in S^X \times S^\Lambda \). The mapping \(\psi\) is obtained by first defining mappings from frequency space to neural space and from pressure space to interval length space. After partitioning, the spaces are effectively discretized into \(\mathcal{F} \subset S^F, \mathcal{P} \subset S^P, X^\lambda \subset S^X, \Lambda \subset S^\Lambda\). These are the functionally and biological relevant sets as they contain the audible frequencies and sound pressures that can be mapped into the compartmentalized neural space. Therefore, the mapping is:

\[
\begin{align*}
\psi_f : \mathcal{F} &\rightarrow X^\lambda, \psi_f(f) = x; \\
\psi_p : \mathcal{P} &\rightarrow \Lambda, \psi_p(p) = \lambda; \\
\psi_{f,p} : \mathcal{F} \times \mathcal{P} &\rightarrow X^\lambda \times \Lambda, \\
\psi_{f,p}(f, p) &= (\psi_f(f), \psi_p(p)) = (x, \lambda)
\end{align*}
\]

The maps are bijective if \(|\mathcal{F}| = |X^\lambda|\) and \(|\mathcal{P}| = |\Lambda|\) and continuous because the discretized acoustic space has the discrete topology.

### Algebraic Operations in Neural Space

In this section, the algebraic operations for combining intervals in neural space will be defined. To make the explanations more intuitive, the operations will be developed specifically with open sets in \(\tau^X\) with single intervals, though they apply generally to any open set in \(\tau^X\).

#### Addition

Addition ("\(+\)") is defined as the union. Given two intervals \(h_{x,\lambda_\alpha}, h_{x,\lambda_\beta} \in \tau^X\):

\[
+ : \tau^X \times \tau^X \rightarrow \tau^X
\]

\[
h_{x,\lambda_\alpha} + h_{x,\lambda_\beta} \overset{\text{def}}{=} h_{x,\lambda_\alpha} \cup h_{x,\lambda_\beta}
\]

The addition operation yields a single interval if the intervals overlap and the same two intervals if the intervals are disjoint.
\[ h_{x\alpha\lambda} + h_{x\beta\lambda} = \]
\[ \begin{cases} 
  h_{x\lambda}, x = \min(x_\alpha, x_\beta), \lambda = \left| h_{x\lambda} \right| + \left| h_{x\lambda_\beta} \right| - \left| h_{x\lambda} \cap h_{x\lambda_\beta} \right|, & h_{x\lambda} \cap h_{x\lambda_\beta} \neq \emptyset \\
  \{ h_{x\alpha\lambda}, h_{x\beta\lambda} \} & h_{x\lambda} \cap h_{x\lambda_\beta} = \emptyset
\end{cases} \tag{13} \]

For more than two intervals, the total length \( \lambda \) is calculated using the Inclusion-Exclusion principle.

The algebraic structure with the addition operation is a commutative monoid since it inherits the algebraic properties of the union operation. In particular, addition is closed because the result is either a longer, half-open interval or two disjoint half-open intervals, either of which is an element of \( \tau^X \). Addition is commutative, associative, has the empty set as the identity element, and does not have an inverse.

### Multiplication

Multiplication “\( \cdot \)" of synaptic intervals is defined as the set minus operation (\( \setminus \)). Given two intervals \( h_{x\alpha\lambda}, h_{x\beta\lambda} \in \tau^X \), where \( h_{x\alpha\lambda} \) is the multiplier and \( h_{x\beta\lambda} \) is the multiplicand,

\[ \cdot : \tau^X \times \tau^X \rightarrow \tau^X; \]

\[ h_{x\alpha\lambda} \cdot h_{x\beta\lambda} \overset{\text{def}}{=} h_{x\beta\lambda} \setminus h_{x\alpha\lambda} \tag{14} \]

Multiplication can yield several results.
The algebraic structure with the multiplication operation is a magma, with properties inherited from the set minus operation. Multiplication is closed because the set minus operation yields an interval that is either the empty set or a half-open interval(s) with a minimum length(s) of $\Delta x$ and maximum length of $\lambda_{max}$. In either case, the resulting interval is in $\tau^X$. While the $\emptyset$ is a left identity element, there is no unique inverse element since the result is the $\emptyset$ whenever the length of multiplier is greater than or equal to that of the multiplicand. It is also neither commutative nor associative.

**Proposition 0.0.1.** The multiplication operation is left but not right distributive over addition.

**Proof.** Left distributive: Let $h_{x_\alpha, \lambda_\alpha}, h_{x_\beta, \lambda_\beta}, h_{x_\gamma, \lambda_\gamma} \in \tau^X$

\[
\begin{align*}
&= h_{x_\gamma, \lambda_\gamma} \cdot (h_{x_\alpha, \lambda_\alpha} + h_{x_\beta, \lambda_\beta}) \\
&= (h_{x_\alpha, \lambda_\alpha} \cup h_{x_\beta, \lambda_\beta}) \setminus h_{x_\gamma, \lambda_\gamma} \\
&= h_{x_\gamma, \lambda_\gamma} \cap (h_{x_\alpha, \lambda_\alpha} \cup h_{x_\beta, \lambda_\beta}), \text{‘c’ is complement} \\
&= (h_{x_\gamma, \lambda_\gamma} \cap h_{x_\alpha, \lambda_\alpha}) \cup (h_{x_\gamma, \lambda_\gamma} \cap h_{x_\beta, \lambda_\beta}) \\
&= h_{x_\alpha, \lambda_\alpha} \setminus h_{x_\gamma, \lambda_\gamma} \cup h_{x_\beta, \lambda_\beta} \setminus h_{x_\gamma, \lambda_\gamma} \\
&= h_{x_\gamma, \lambda_\gamma} \cdot h_{x_\alpha, \lambda_\alpha} + h_{x_\gamma, \lambda_\gamma} \cdot h_{x_\beta, \lambda_\beta}
\end{align*}
\]
Not Right distributive:

\[(h_{x_\alpha,\lambda_\alpha} + h_{x_\beta,\lambda_\beta}) \cdot h_{x_\gamma,\lambda_\gamma} = h_{x_\gamma,\lambda_\gamma} \setminus (h_{x_\alpha,\lambda_\alpha} \cup h_{x_\beta,\lambda_\beta}) = (h_{x_\gamma,\lambda_\gamma} \setminus h_{x_\alpha,\lambda_\alpha}) \cap (h_{x_\gamma,\lambda_\gamma} \setminus h_{x_\beta,\lambda_\beta}) \neq h_{x_\alpha,\lambda_\alpha} \cdot h_{x_\gamma,\lambda_\gamma} + h_{x_\beta,\lambda_\beta} \cdot h_{x_\gamma,\lambda_\gamma}\]

Comparison with columnar organization

In the mathematical model, the ‘functional’ unit is an interval that can vary in size, depending on the sound intensity. The indistinct borders are consistent with topographical organization of afferents [1]. The expanding border facilitates high resolution representation of both frequency and pressure. Assuming a bijective mapping between acoustic and neural spaces, the number of pressure levels that can be represented is \(|P| = |\Lambda| = n_{\lambda,max}\) and the number of frequency levels is given by: \(|F| = |X^\lambda| = N_h - n_{\lambda,max} + 1\). The relation between the number of frequency and pressure levels is then: \(|F| = N_h - |P| + 1\). The total number of neurons needed to achieve an equal number of frequency and pressure levels \((|F| = |P| = n_{\text{levels}})\), is \(N_h = |F| + |P| - 1 = 2 \times n_{\text{levels}} - 1\).

In contrast, consider an alternative neural space that consists of classical columns [2]. The definition of a column has evolved [3] but for illustrative purposes, columns are defined here as having distinct borders; the enclosed neurons have identical receptive fields. In this scheme, the location of the active column encodes frequency and sound pressure is encoded as the number of neurons within each column that become active during the stimulus (assuming a brief stimulus so that neurons are effectively binary). For example, the probability of neurons firing within a column may increase with sound pressure (e.g. population rate code).

The neural space consists of a set of adjacent columns, each of which is a half-open interval \([x, x + n_w\Delta x]\) where \(n_w\) is the number of cells along the tonotopic axis. The number of frequency levels that can be represented is equal to the number of columns that fit along the tonotopic axis: \(|F| = \frac{N_h}{n_w}\). The number of pressure levels that can be represented is equal to the number
of cells in each column: $|\mathcal{P}| = n_w$. Hence, the relationship between the number of frequency and pressure levels is: $|\mathcal{P}| = \frac{N_{\text{total}}}{|\mathcal{F}|}$ where $N_{\text{total}}$ is the total number of neurons and is equal to $N_h$ for a monolayer column and equal to $N_h \times n_z$ if each column is 2 dimensional and has a height of $n_z$ cells. The total number of neurons needed to achieve an equal number of frequency and pressure levels ($|\mathcal{F}| = |\mathcal{P}| = n_{\text{levels}}$), is $N_h = n_{\text{levels}}^2$ (or $N_h \times n_z = n_{\text{levels}}^2$ for a column with height $n_z$).

Algebra of loudness summation

A multi-frequency stimulus is a set of tones with increasing frequencies $F_m = \{f_i \mid i \in I\}$ where $I = \{j, \ldots, n\}, j \geq 0, n \leq N_f$. In neural space, the stimulus generates a set of excitatory intervals $H_m$ that is the union of the individual intervals generated by each tonal component: $H_m = \bigcup_{i \in I} (h_i^E = [x_i, x_i + \lambda])$, where $\lambda$ is the length of each interval.

Definition 0.0.1. An inhibitory (half open) interval $h^I$ is in the lateral inhibitory configuration with an excitatory (half open) interval $h^E$, if $h^I \cap h^E = \emptyset$ and there exists a half-open interval $\hat{h} \supset (h^E \cup h^I)$ so that if $x \in \hat{h}$, then $x \in h^E \cup h^I$. This ensures that the excitatory and inhibitory intervals abut each other with no gap. Note that this definition applies whether $h^I$ is to one side of $h^E$ or is split into two that occupies either side (as in the main text and Proposition 0.0.2 above) $h^I = h^I_L \cup h^I_R$.

Definition 0.0.2. A tone in multi-frequency stimulus is ‘dominant’ if it results in an excitatory interval $h^E_d$ and an inhibitory interval $h^I$ that is much larger than that generated by the other tones.

Physiologically, the dominant frequency $f_d$ could be the center frequency that identifies a critical band (CB) [4] or the lowest frequency tone of a complex stimulus [5]. It is assumed that every multi-frequency stimulus has a dominant frequency, which generates the synaptic interval $h^E_d \subset H_m$. The non-dominant frequencies may also generate associated inhibitory intervals but in the following analyses, they are assumed to be much weaker than that generated by the dominant frequency and can be ignored.

The net excitatory interval in the presence of inhibition is given by the algebraic operation: $h_l = h^I \cdot H_m$. The length $|h_l|$ is taken to be the proxy for loudness perception.
Proposition 0.0.2. In a lateral inhibition configuration consisting of the dominant excitatory interval \( h_d^E \) and an inhibitory interval \( h_I \), \(|h_I|\) does not change when the size (number of element and/or length of each element) of \( H_m \) is increased as long as \( H_m \) remains within (is a subset of) a ‘critical interval’ \( H_{CI} = h_d^E + h_I \defeq h_d^E \cup h_I \) and increases thereafter.

Proof. First note that \( H_m = h_d^E \cup (H_m \setminus h_d^E) = h_d^E + h_d^E \cdot H_m \) since \( h_d^E \in H_m \). Thus,

\[
h_I = h_I \cdot H_m = h_I \cdot (h_d^E + h_d^E \cdot H_m) = h_I \cdot h_d^E + h_I \cdot (h_d^E \cdot H_m)
\]

(16)

Therefore, \(|h_I| = \)

\[
\begin{cases}
|h_d^E|, & h_d^E \cdot H_m \subset h_I \\
|h_d^E|, & h_d^E \cdot H_m \not\subset h_I
\end{cases}
\]

Now, \( h_d^E \cdot H_m \subset h_I \Rightarrow h_d^E + (h_d^E \cdot H_m) \subset h_d^E + h_I \) or \( H_m \subset H_{CI} \)

Thus, \( h_I \) is equal to \( h_d^E \) as long as \( H_m \subset H_{CI} \) and increases when \( H_m \not\subset H_{CI} \).

Definition 0.0.3. A set of tones with frequencies given by \( F_m \), each delivered at some sound pressure level, is represented in neural space by the set of intervals given by \( H_m \). The critical band (CB) is difference between the highest frequency and lowest frequency in \( F_m \) that satisfies \( H_m \subseteq H_{CI} \).

An increase in stimulus intensity (or sound pressure) results in a lengthening of excitatory intervals. Hence, the lengths of \( h_d^E \) and the other excitatory intervals of \( H_m \) will increase.

Proposition 0.0.3. The critical band does not change with stimulus intensity provided that the lateral inhibitory configuration is preserved and the length \( \lambda_I \) of the inhibitory interval \( h_I \) remains constant.

Proof. We need to show that \( H_m \subset H_{CI} \) when the length of each interval of \( H_m \) is increased by an amount \( \Delta h \) due to the increase in sound pressure level of each tonal components of the stimulus. Recall that a set of tones with frequencies in \( F_m \) results in a set of intervals \( H_m \) where one of the intervals \( h_d^E \) is dominant.
The critical interval $H_{CI}$, as defined above, consists of the dominant excitatory interval and an inhibitory interval in the lateral inhibitory configuration: $H_{CI} = h^E_d \cup h^I$. If $H_m \subset H_{CI}$, then $|H_{CI}|$ must be greater than or equal to the distance $\delta = (x_n + \lambda) - x_1$, where $x_1, x_n$ are, respectively, the leftmost points of the first and last interval in $H_m$: $|H_{CI}| = |h^E_d| + |h^I| = \lambda + \lambda_I \geq \delta$.

Increasing the length of each interval (which includes $h^E_d$) by $\Delta h$ extends the last point in $H_m$ so that $\delta = (x_n + \lambda + \Delta h) - x_1$. Under the condition that the lateral inhibitory configuration is preserved ($h^E_d$ and $h^I$ remain disjoint) and the length of the inhibitory interval is constant, the length of $H_{CI}$ increases by the same amount: $|H_{CI}| = (\lambda + \Delta h) + \lambda_I$. Therefore, $H_m \subset H_{CI}$.

The total length ($h_l$) resulting from the operations are shown graphically for band-limited white noise with increasing bandwidth (main text) and for a complex stimulus with 4 tones with increasing spacing between frequencies (Fig B)

![Diagram](image)

**Fig B: Algebra of loudness summation** Predicted interval lengths resulting from the interaction of 4 tones delivered simultaneously (inset). *Boxed inset*, The tones generate 4 excitatory intervals ($H_m$, gray) one of which is dominant ($h^E_d$, blue). The dominant tone also generates an inhibitory synaptic interval (red). Plot shows resultant length ($h_l$) after the operations as the frequency spacing between intervals in $H_m$ is increased (abscissa). Dotted vertical line marks deviation from constant value. Compare with Fig 3 of [6].
References


5. Wegel R, Lane C. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Physical review. 1924;23(2):266.