Modelling vowel sounds with source-filter theory

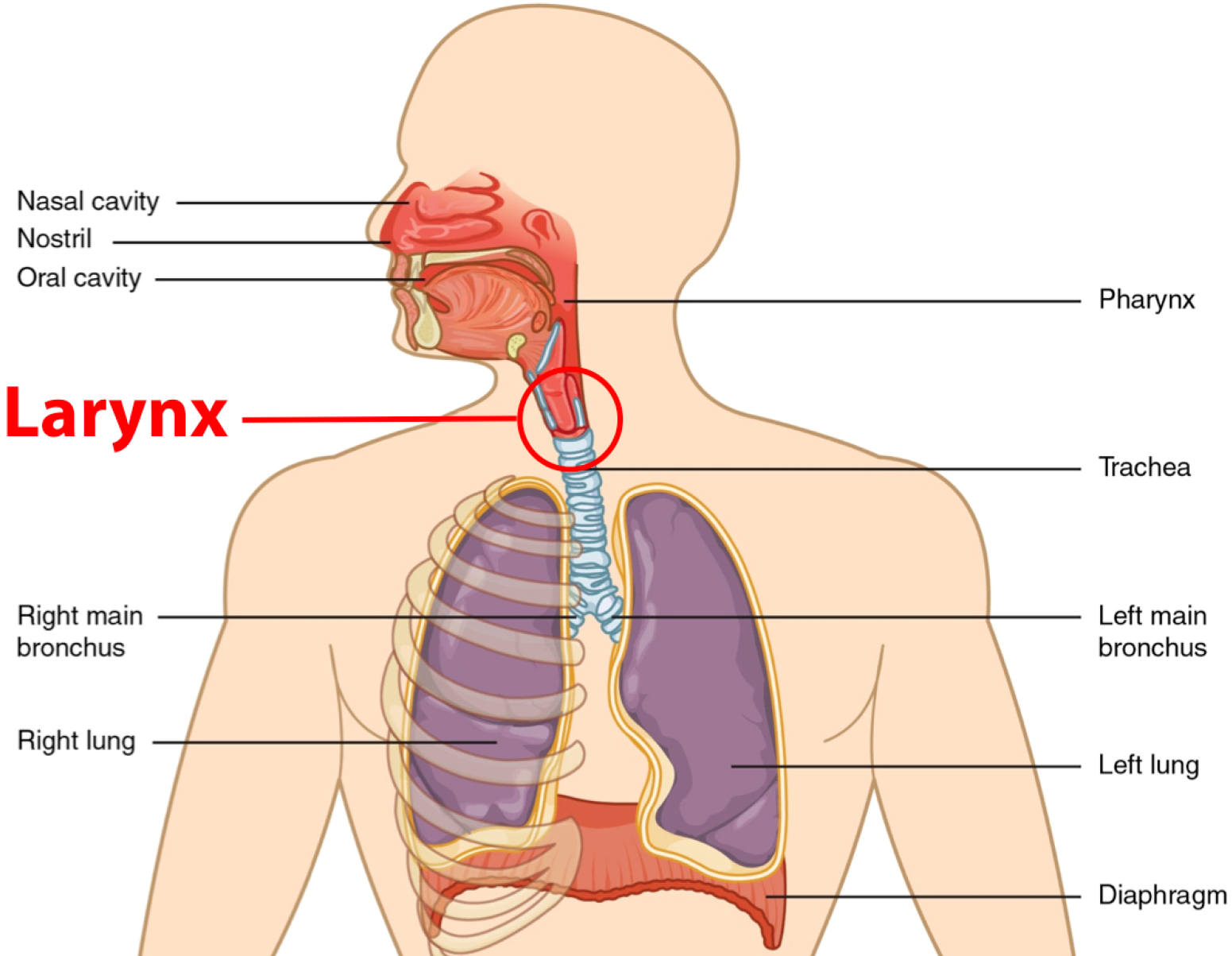
Samuli Siltanen

Department of Mathematics and Statistics
University of Helsinki, Finland
samuli.siltanen@helsinki.fi
www.siltanen-research.net

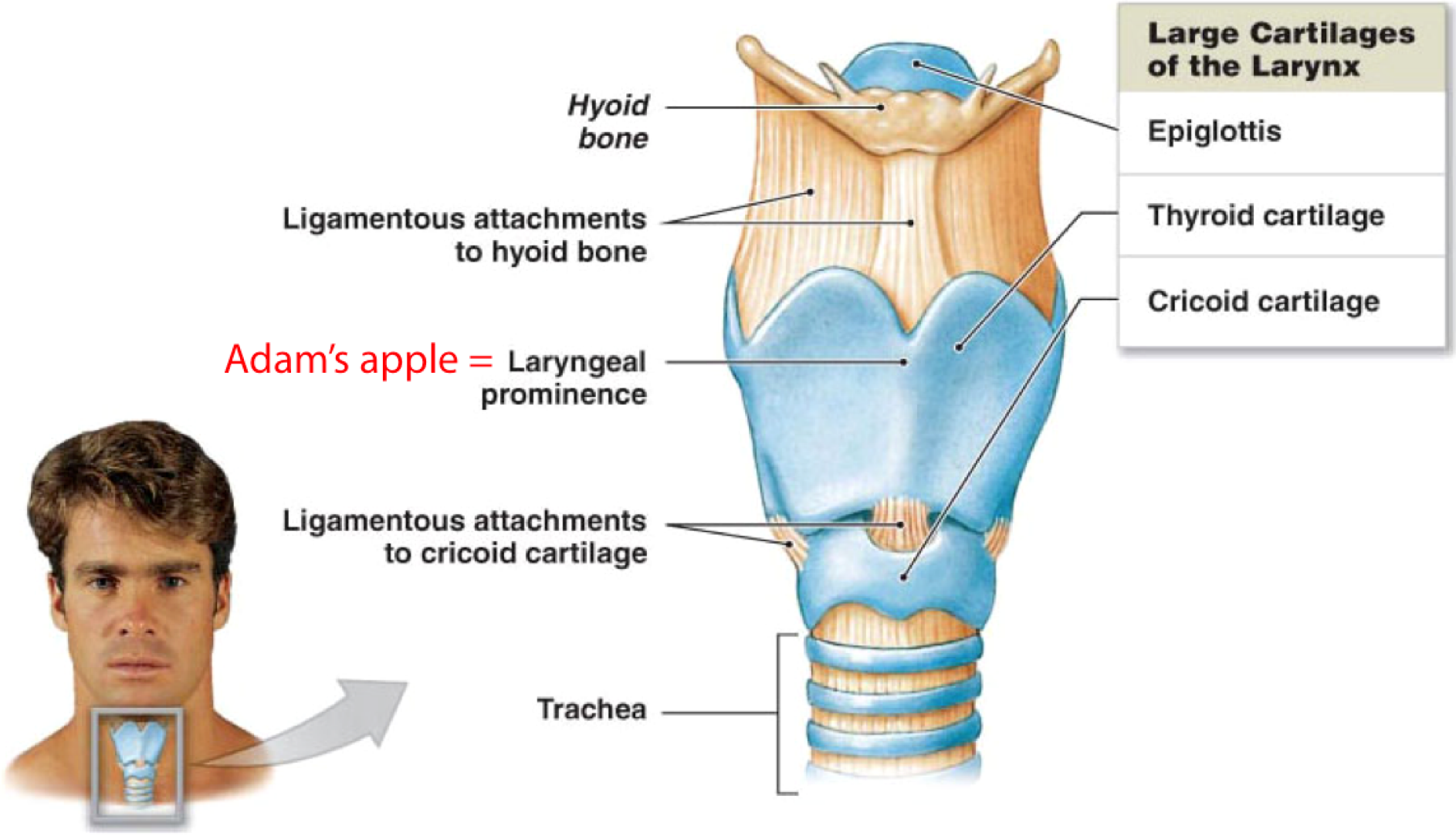
Applications of Matrix Computations 2016

Outline

These parts of the human anatomy are most important for speech production

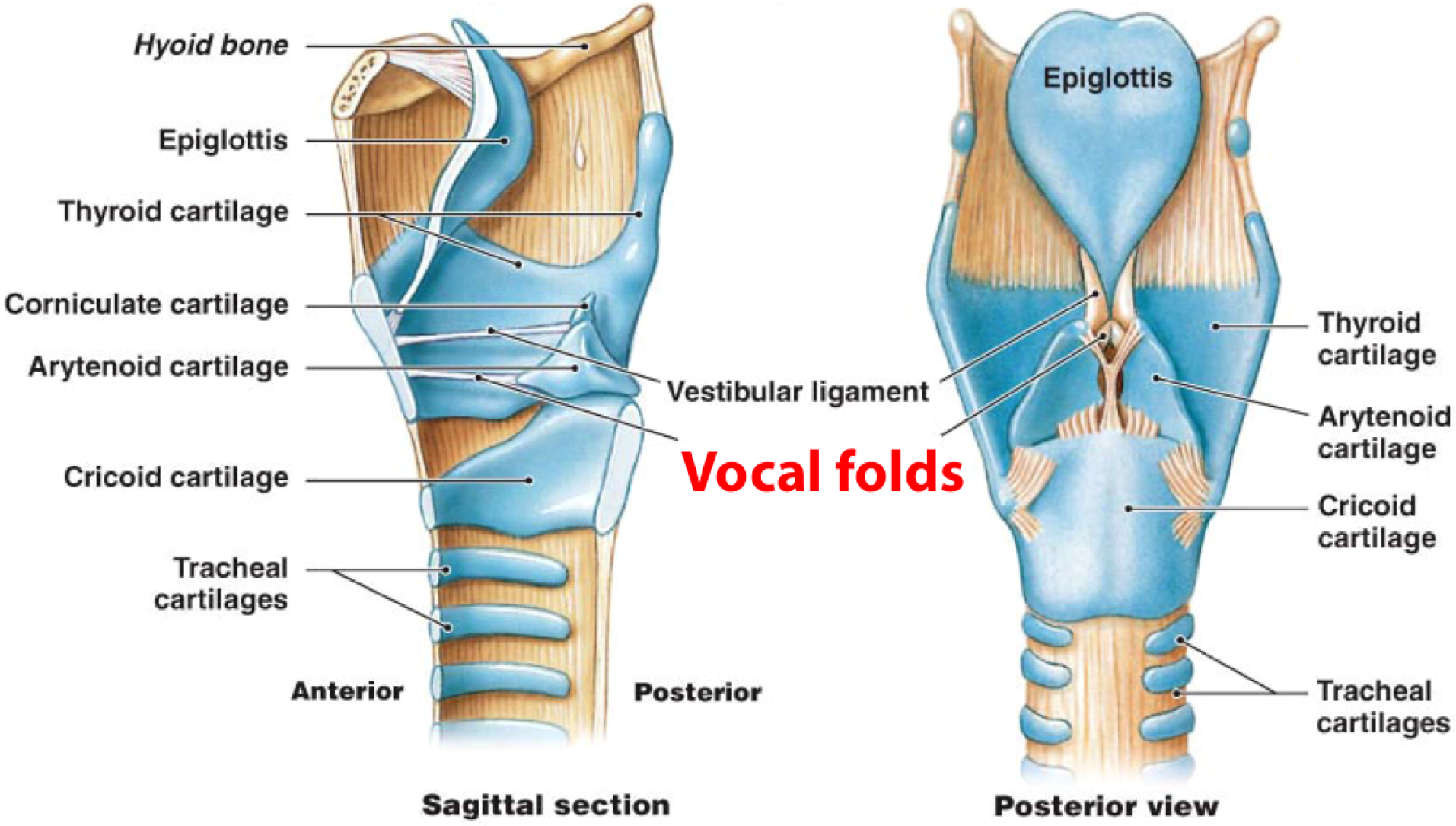


This is frontal view of the larynx



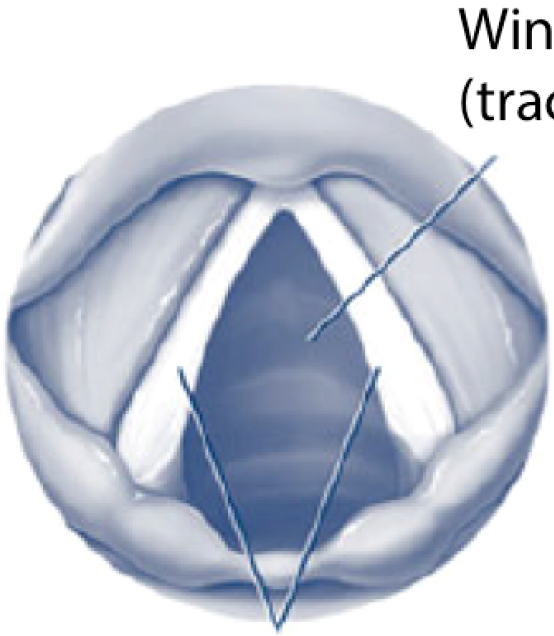
Images: Pearson education

Side and back views of the larynx

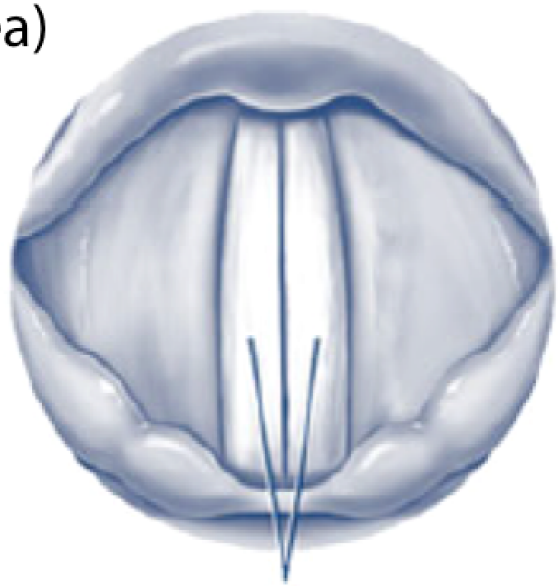


Images: Pearson education

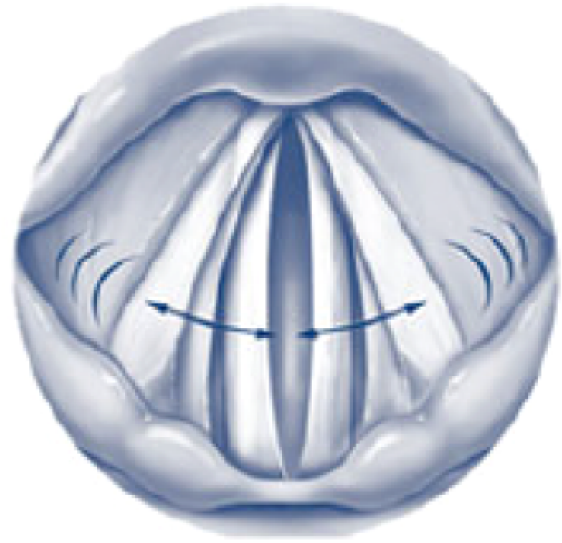
The excitation signal at the vocal folds comes from their periodic flapping against each other



Vocal folds are open when we breathe



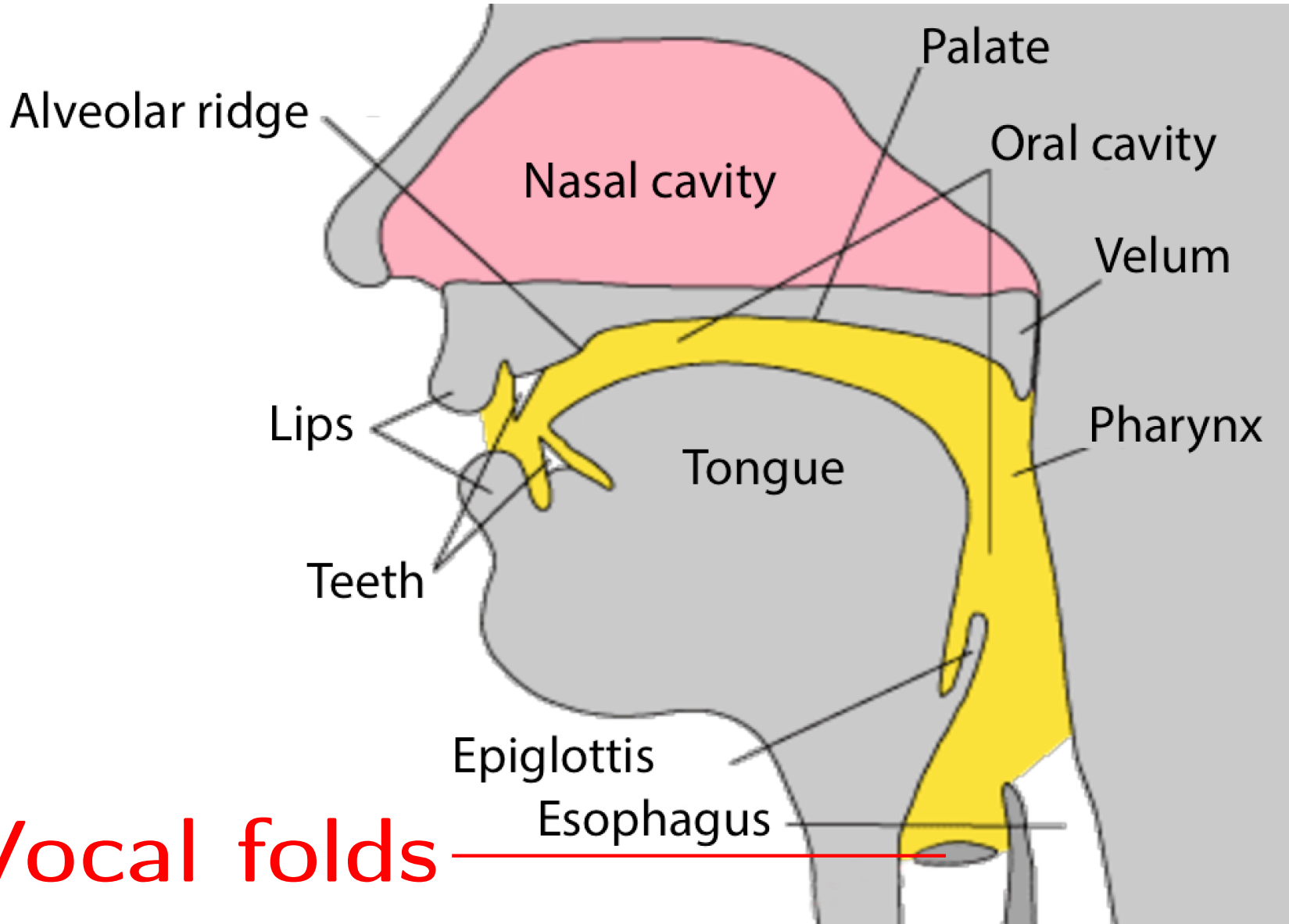
Vocal folds are closed when we swallow or lift something heavy



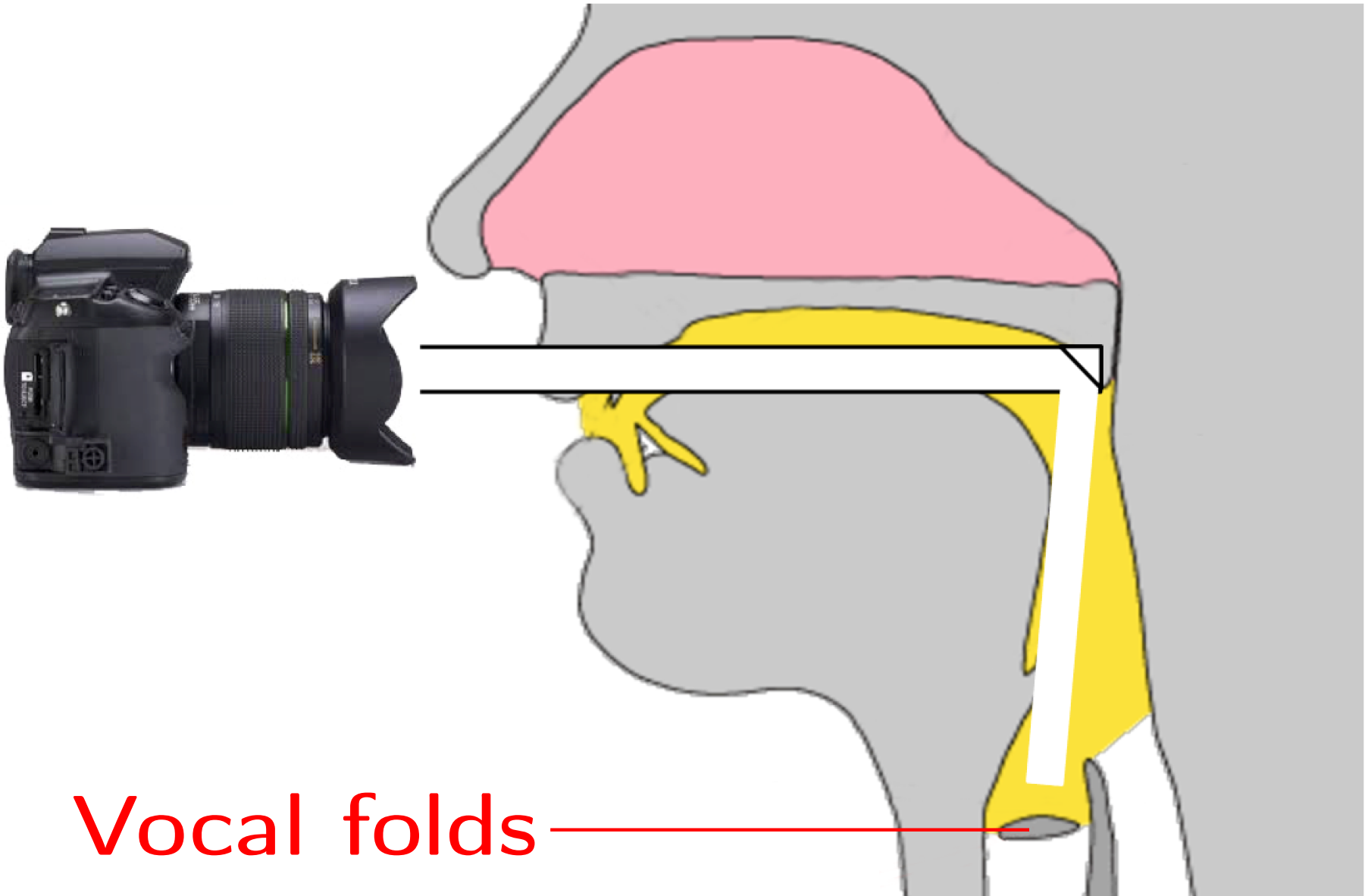
Air causes vocal folds to vibrate between open and closed positions when we talk

Images from www.nidcd.nih.gov

Vocal folds are at the bottom of the vocal tract, which is shown in yellow



If we just had a camera and a mirror,
we could see the vocal folds



Meet phoniatrician Ahmed Geneid, PhD

Inserting the video camera (*laryngoscope*) for imaging vocal folds in action

Now we have a top view of the vocal folds!

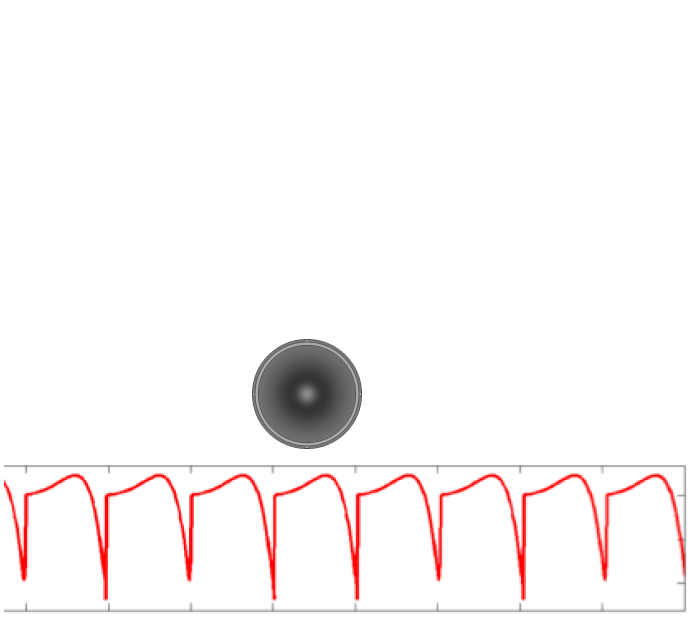
Here you can see the vocal folds moving periodically, creating the excitation signal

Fibroscope allows me to speak during imaging

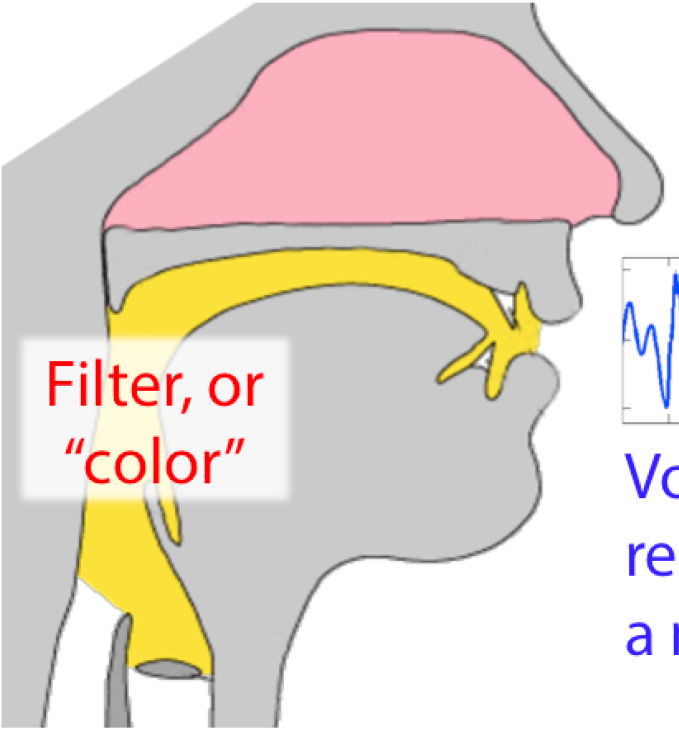
Here you can see the vocal folds moving periodically, creating the excitation signal

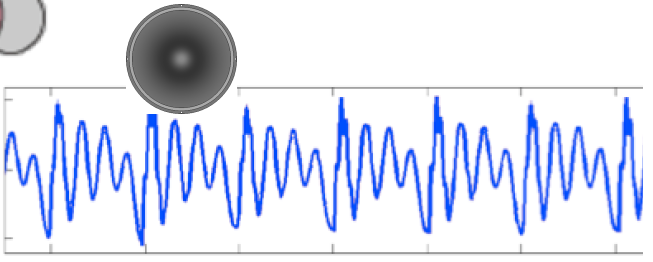
Outline

A vowel sound consists of two structural parts: excitation and vocal tract filter



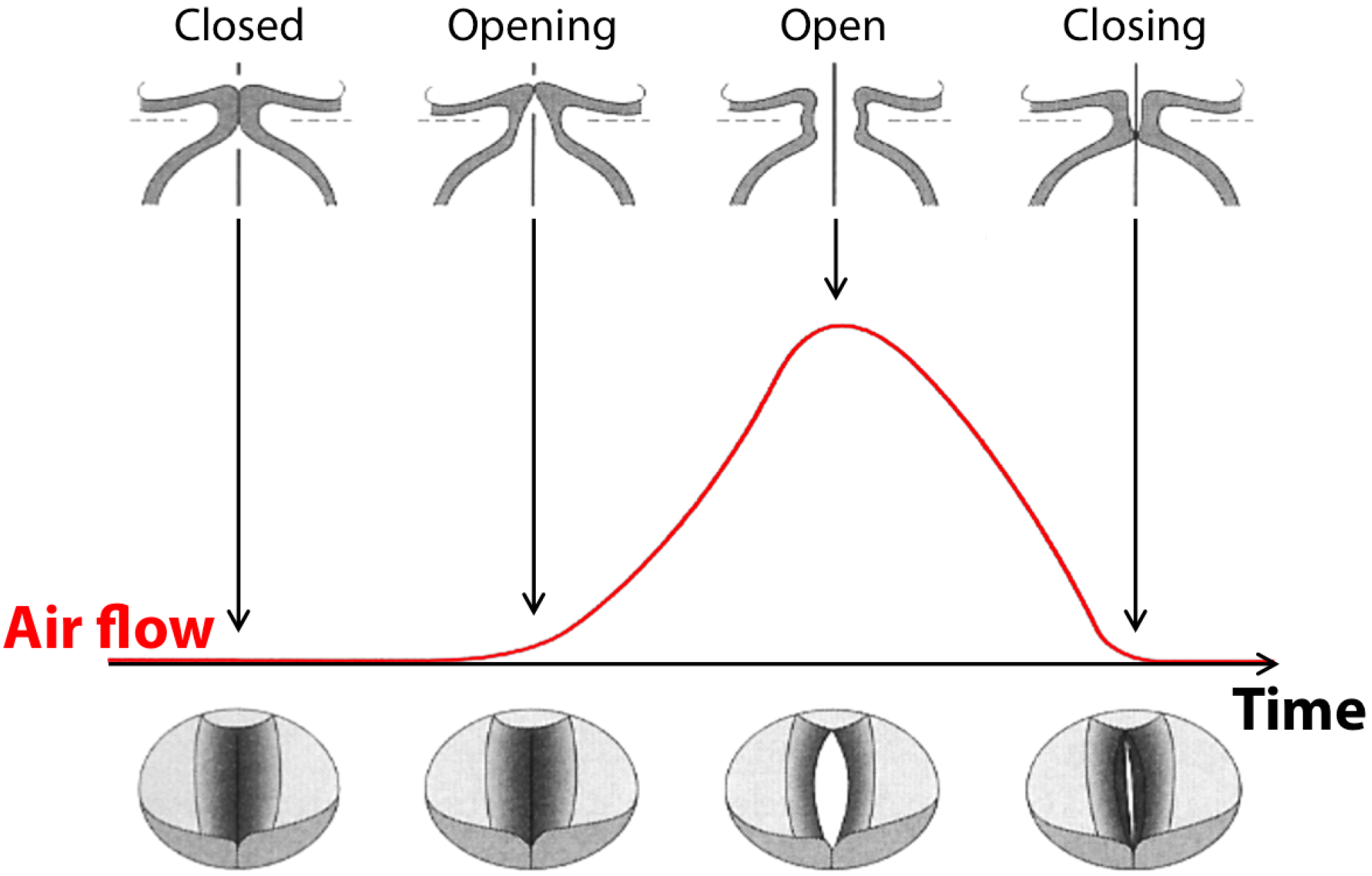
Excitation signal, or
"buzz" from vocal folds



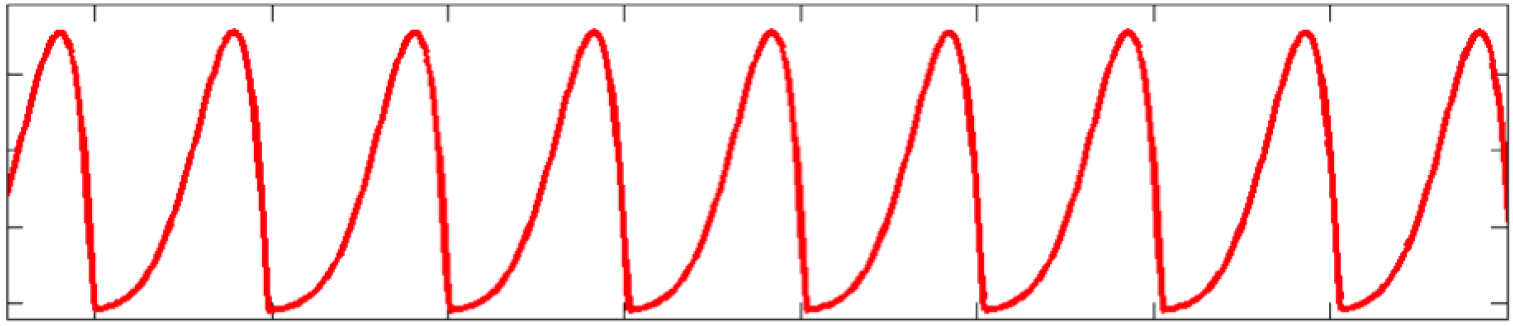


Vowel sound,
recorded with
a microphone

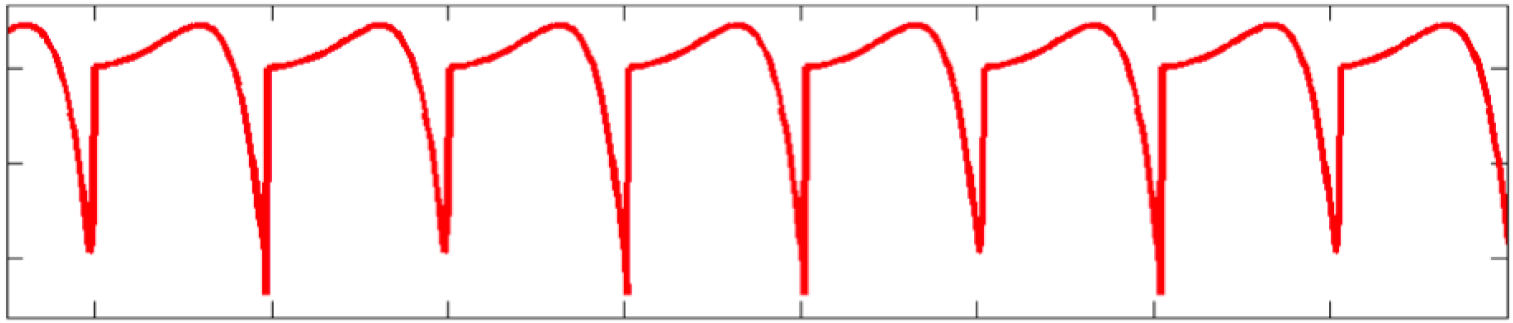
Glottal air flow between the vocal folds



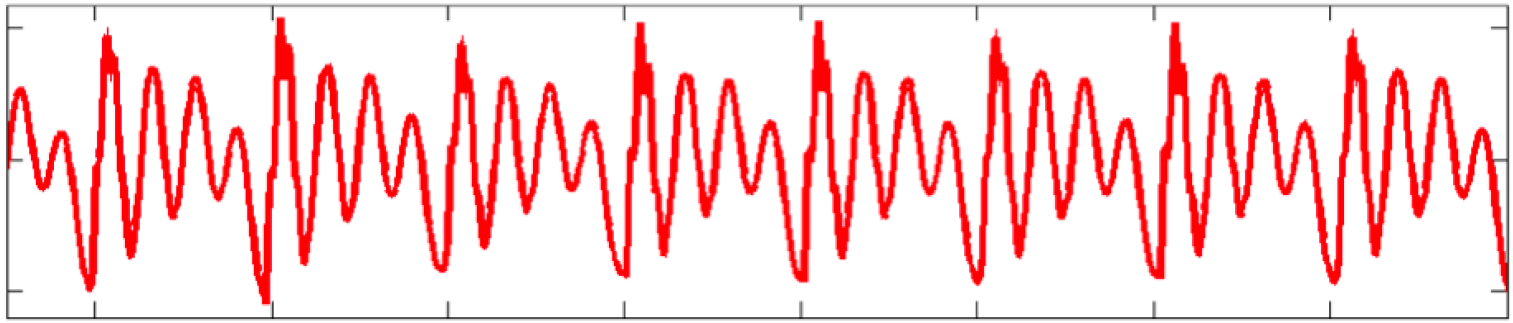
Air flow through the glottis



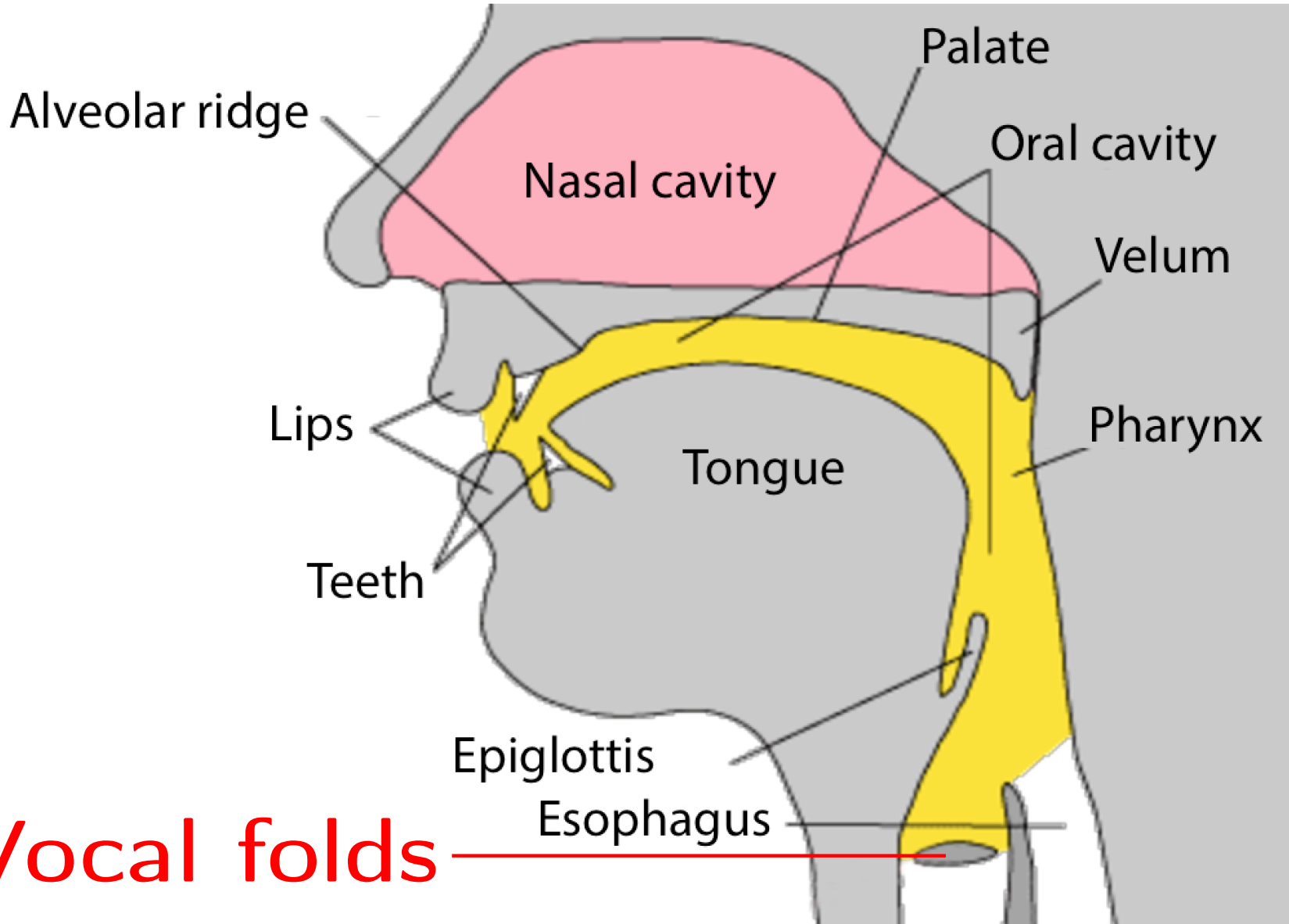
Air pressure at the vocal folds



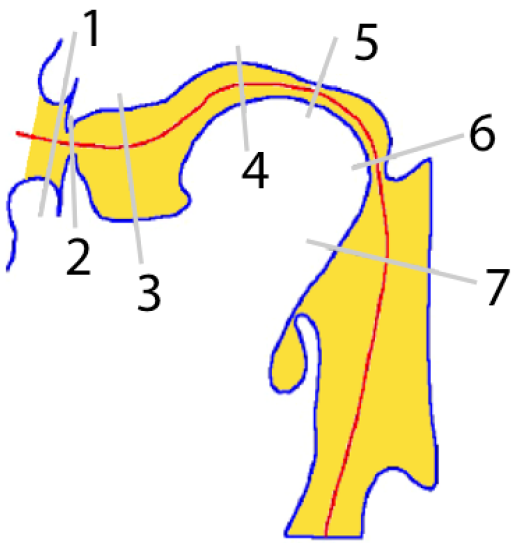
Microphone measures pressure filtered by the vocal tract

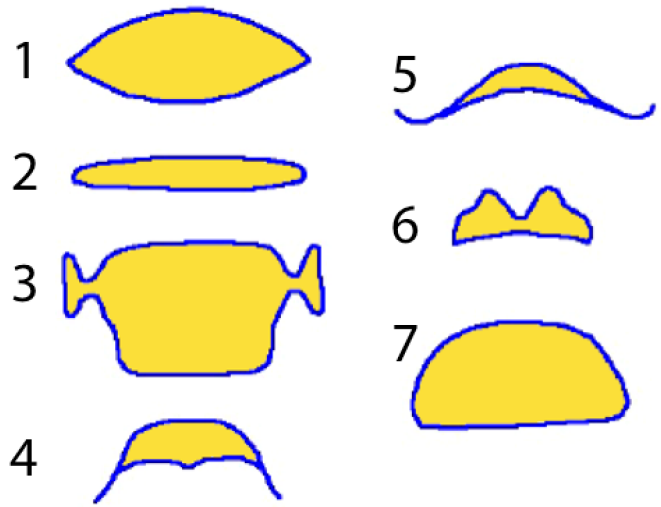


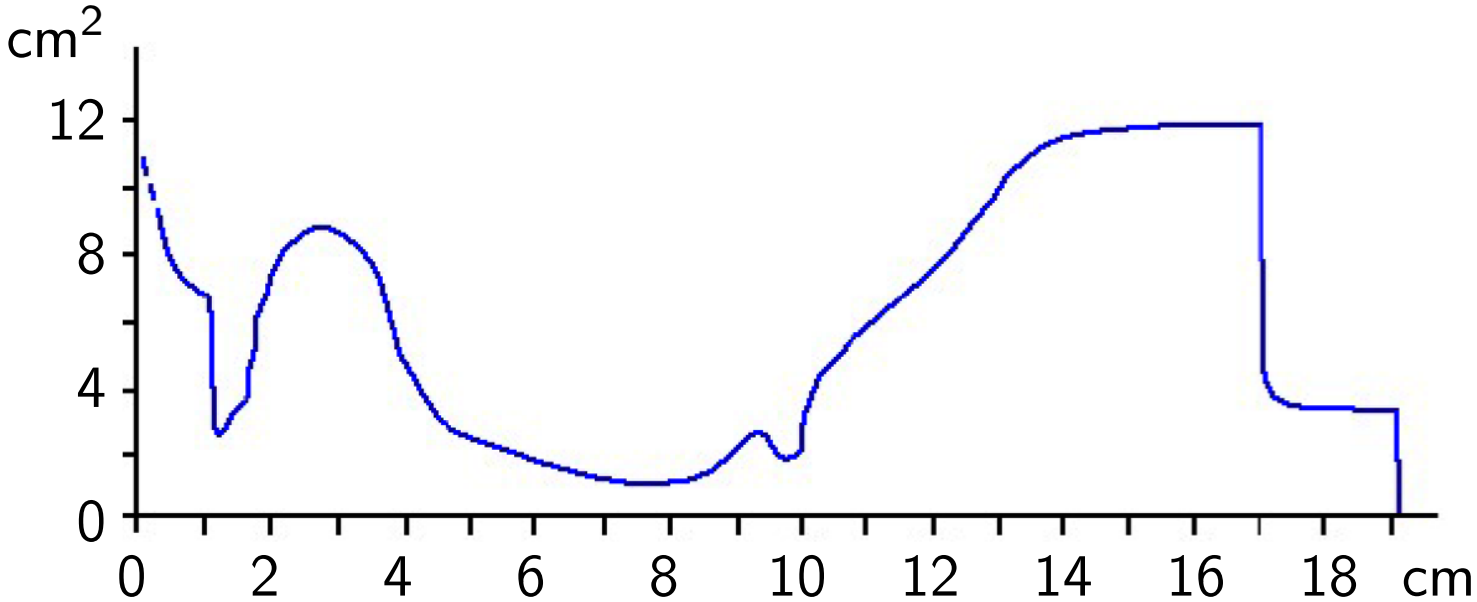
Let's take a closer look at the vocal tract (yellow)



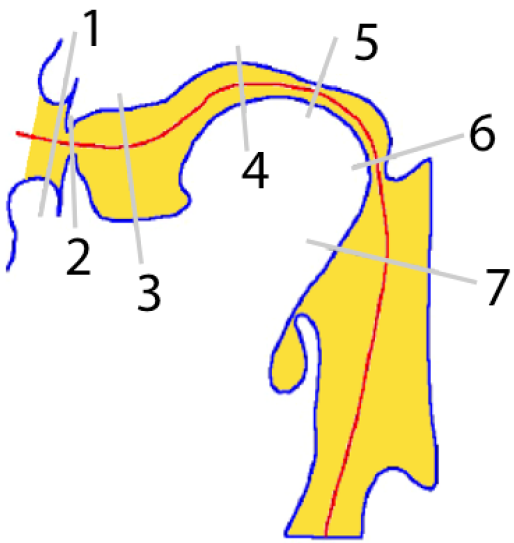
Vocal tract area function shows the size of the tract at different positions along the tube

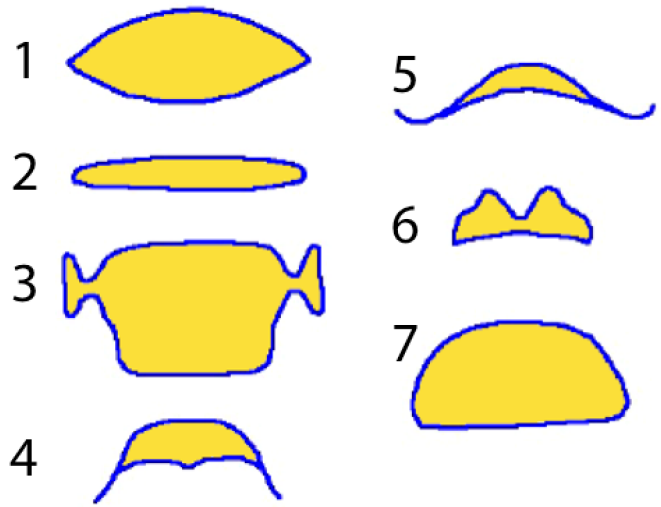


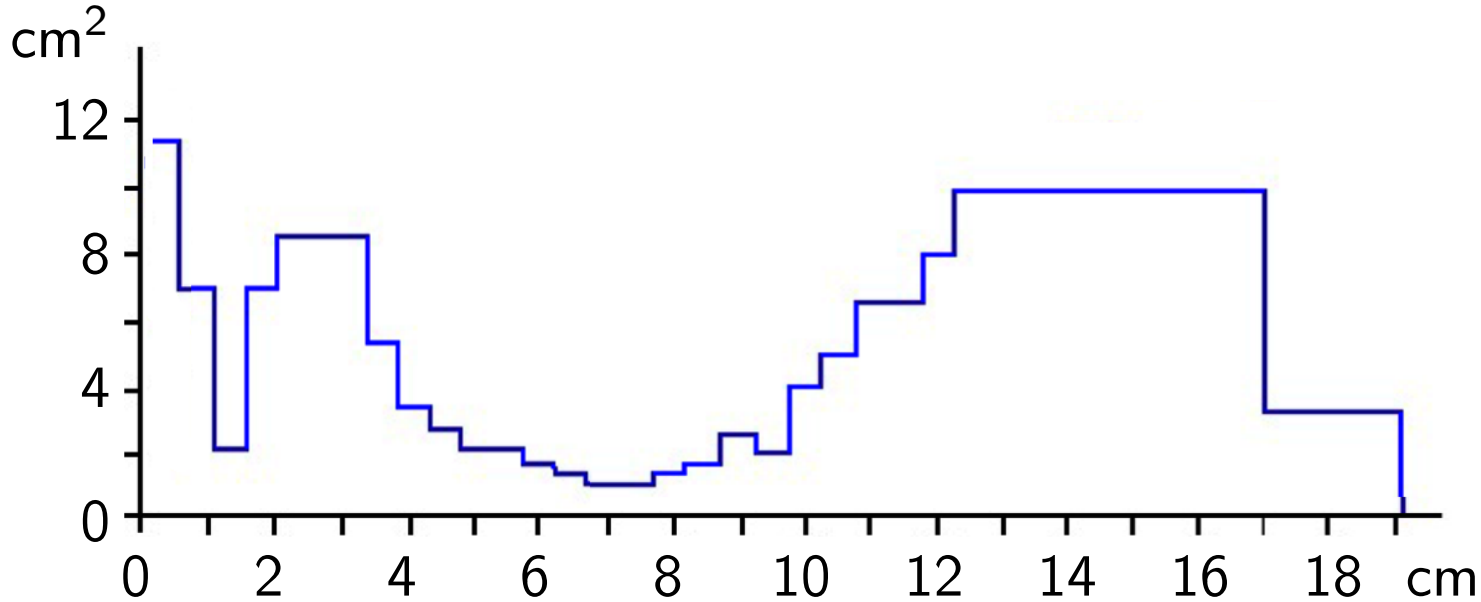




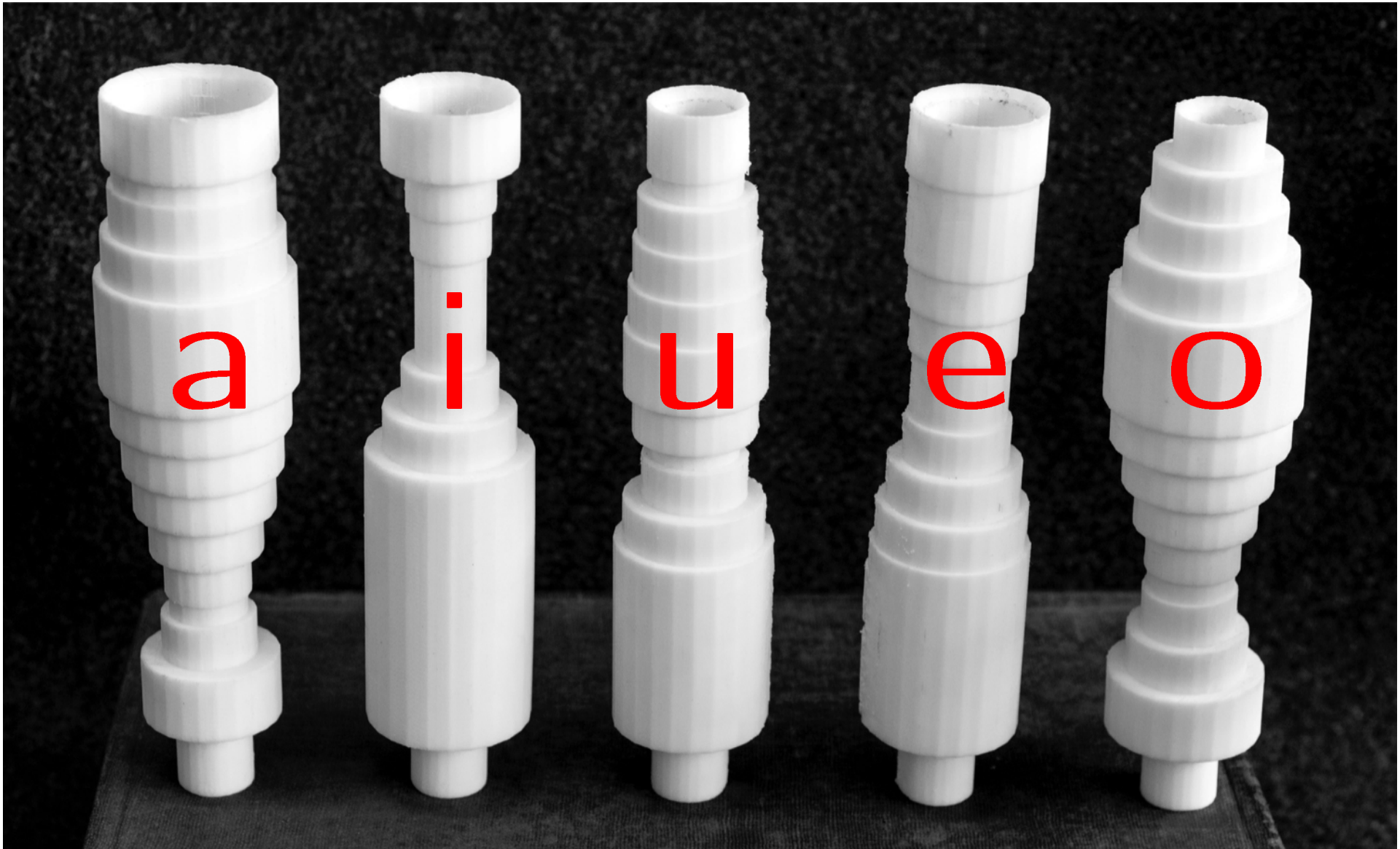
Piecewise constant approximation of the area function gives surprisingly good results



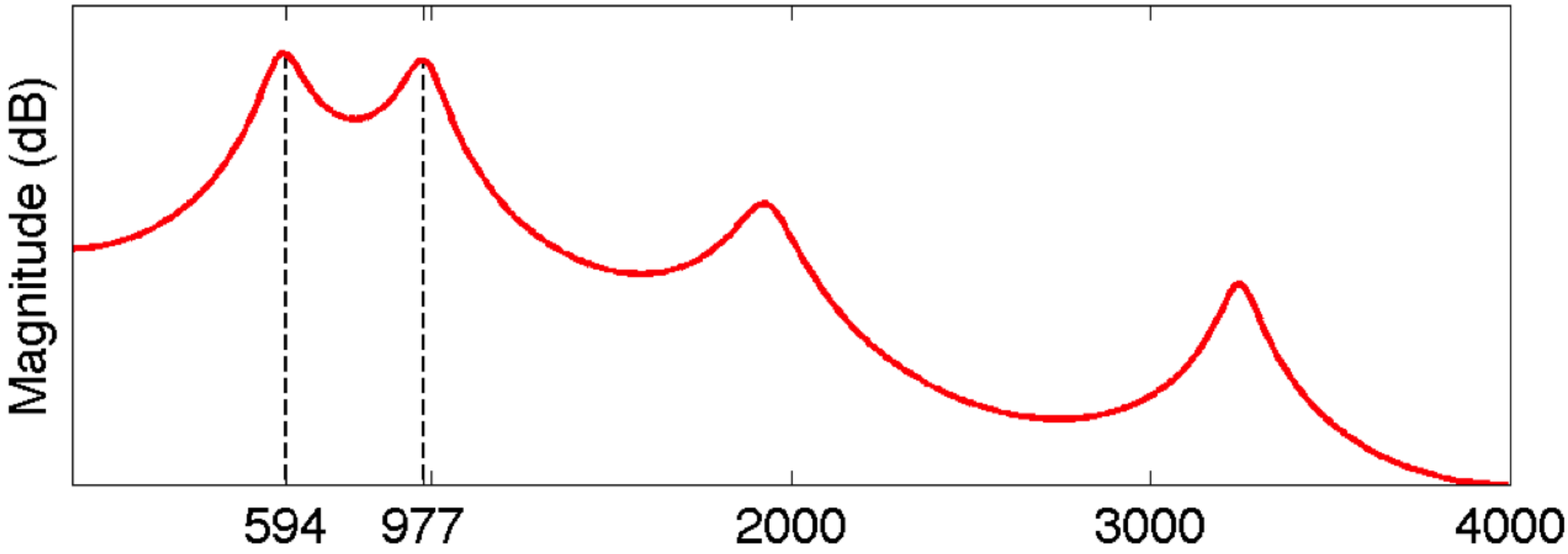




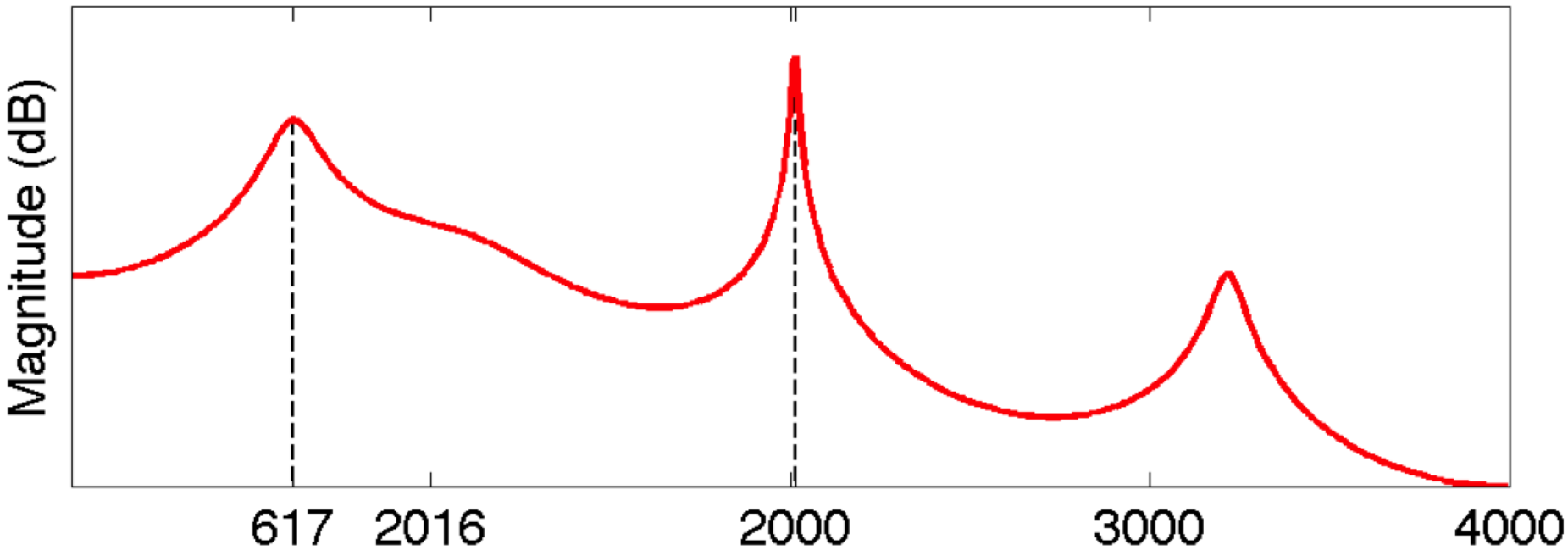
Here are five 3D-printed tube vowel models, adapted from [Arai, Usuki & Murahara 2001]



The vocal tract filter can be described as a frequency response. Here /a/ (as in “car”)



The vocal tract filter can be described as a frequency response. Here /e/ (as in “element”)



Each vowel has two main resonance frequencies called the first and second formant

	First formant (Hz)	Second formant (Hz)
a	850	1610
i	240	2400
u	250	595
e	390	2300
o	360	640

Outline

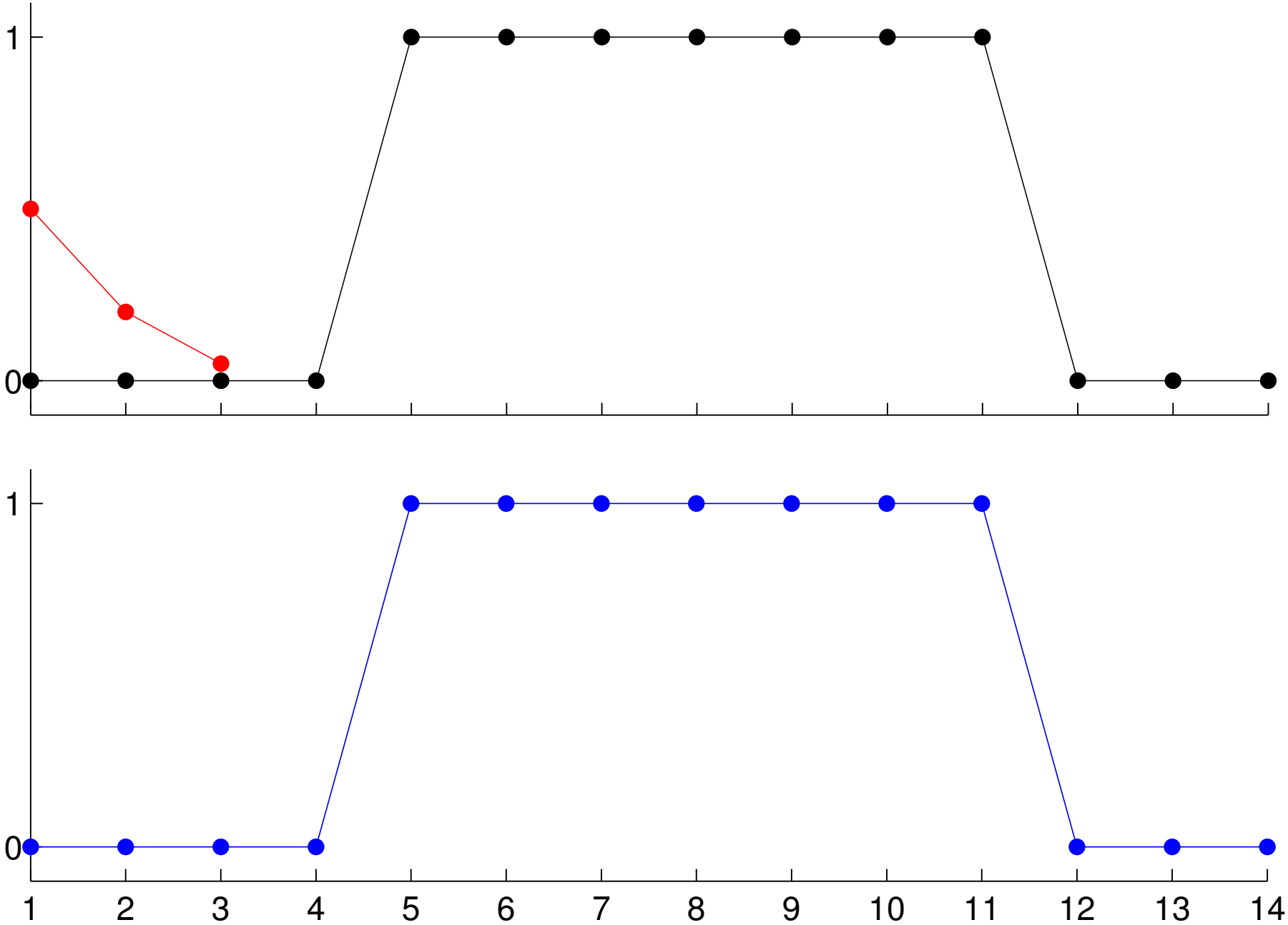
We define discrete convolution using periodic boundary conditions

Let $p \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$. Convolution $p * s \in \mathbb{R}^n$ is defined by the formula

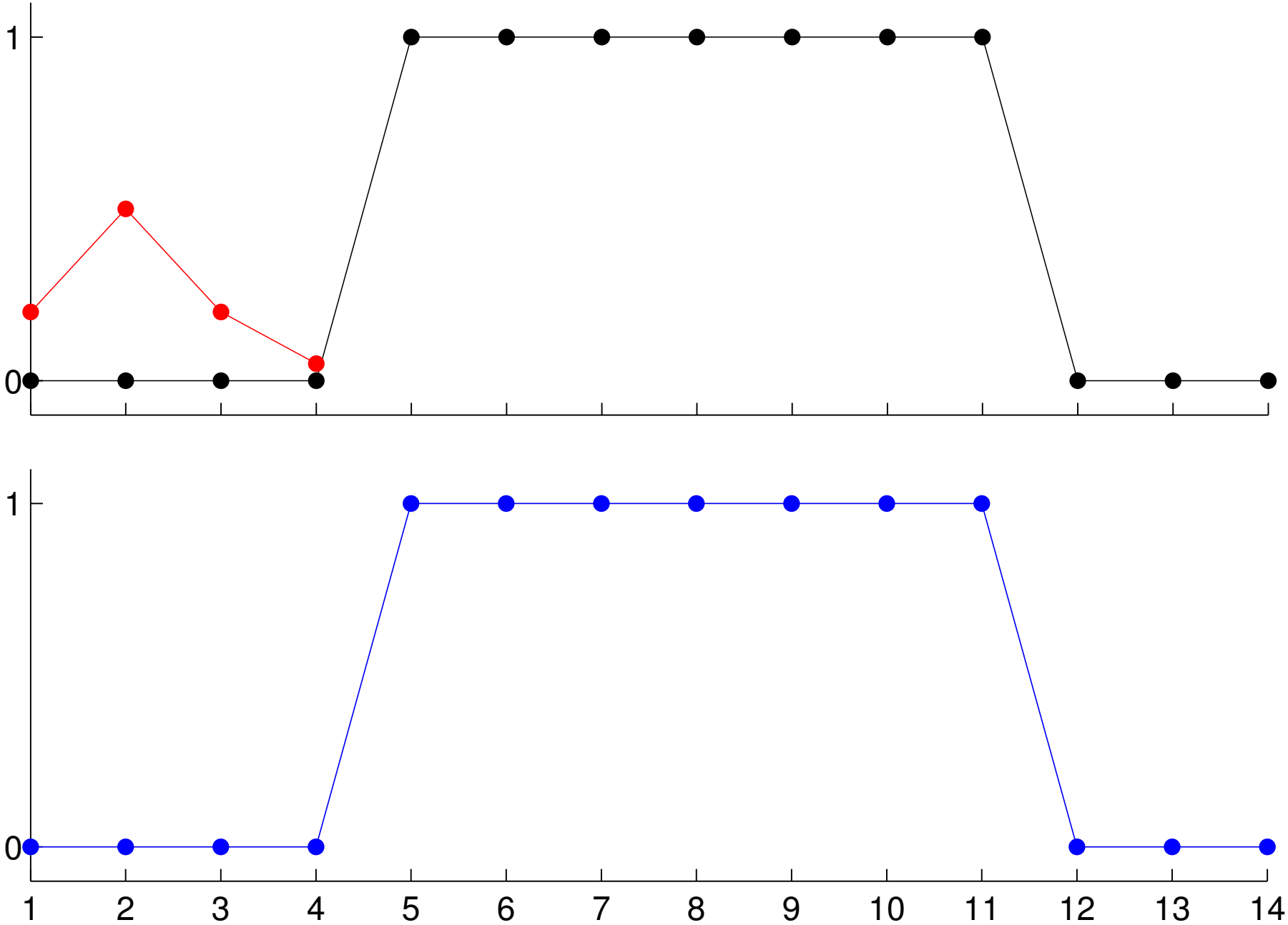
$$(p * s)_j = \sum_{\ell=1}^{n} p_\ell s_{j-\ell},$$

where $s_{j-\ell}$ is defined using periodic boundary conditions for the cases $j - \ell < 1$ and $j - \ell > n$.

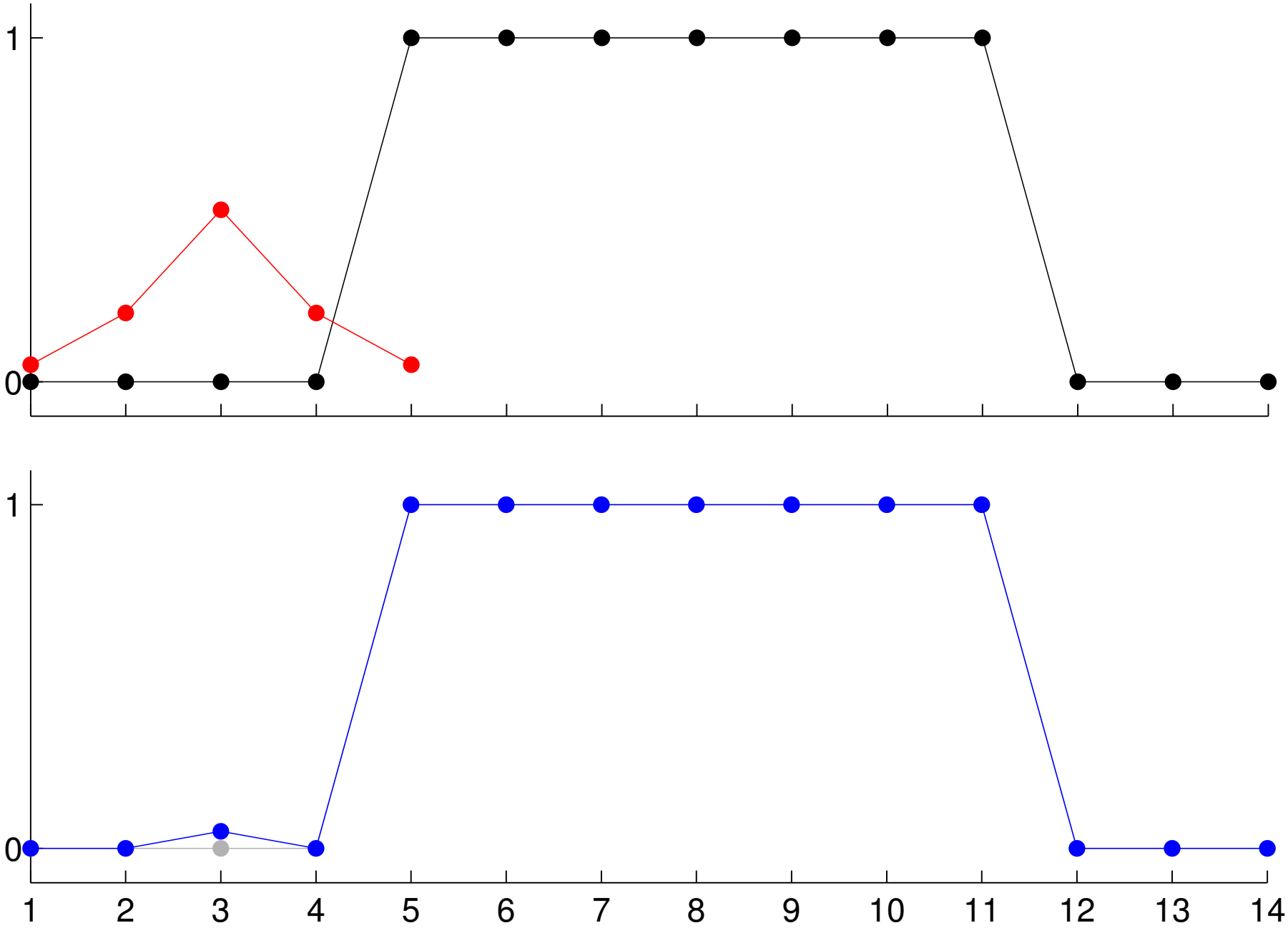
Convolution, position 1



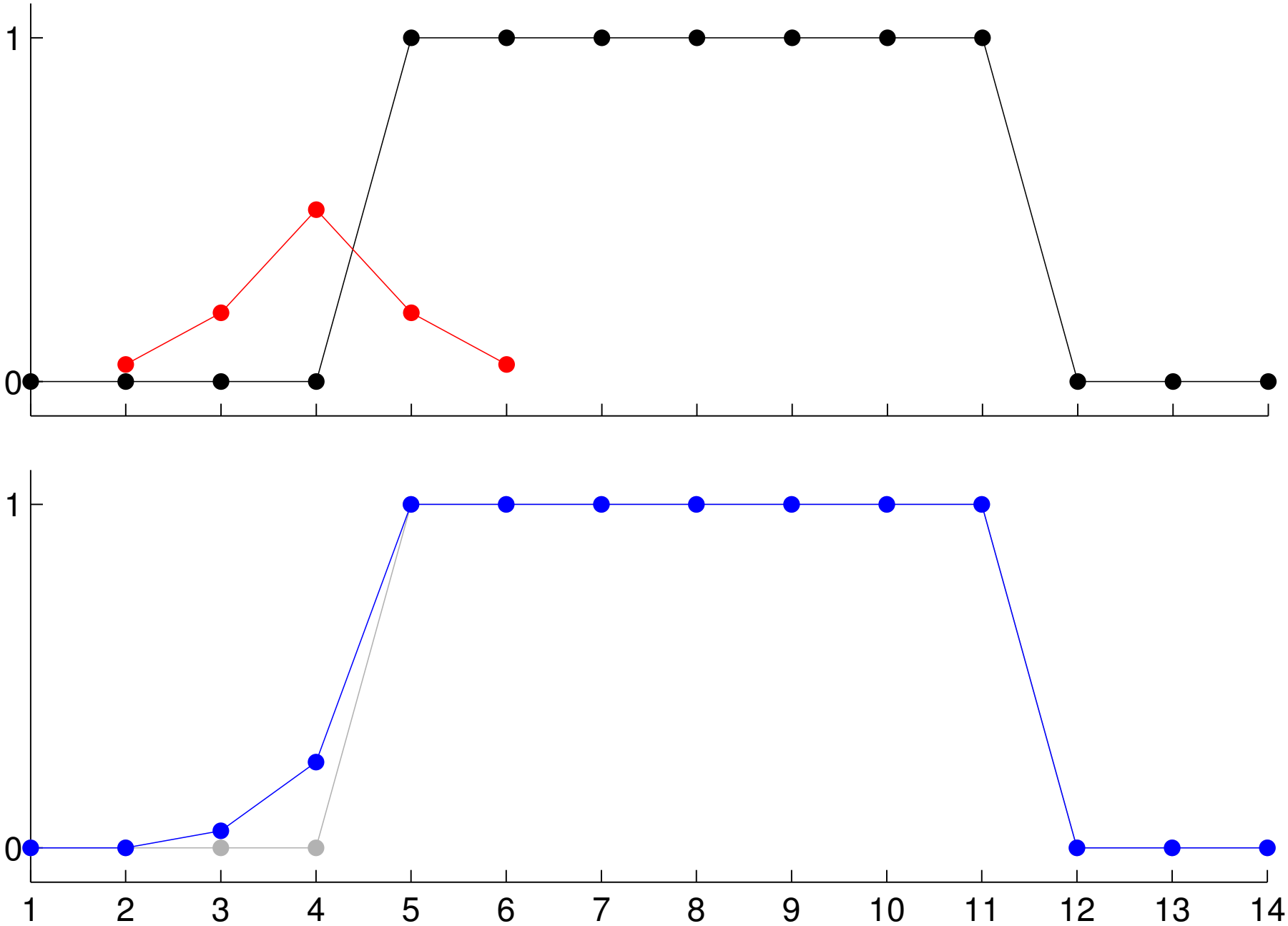
Convolution, position 2



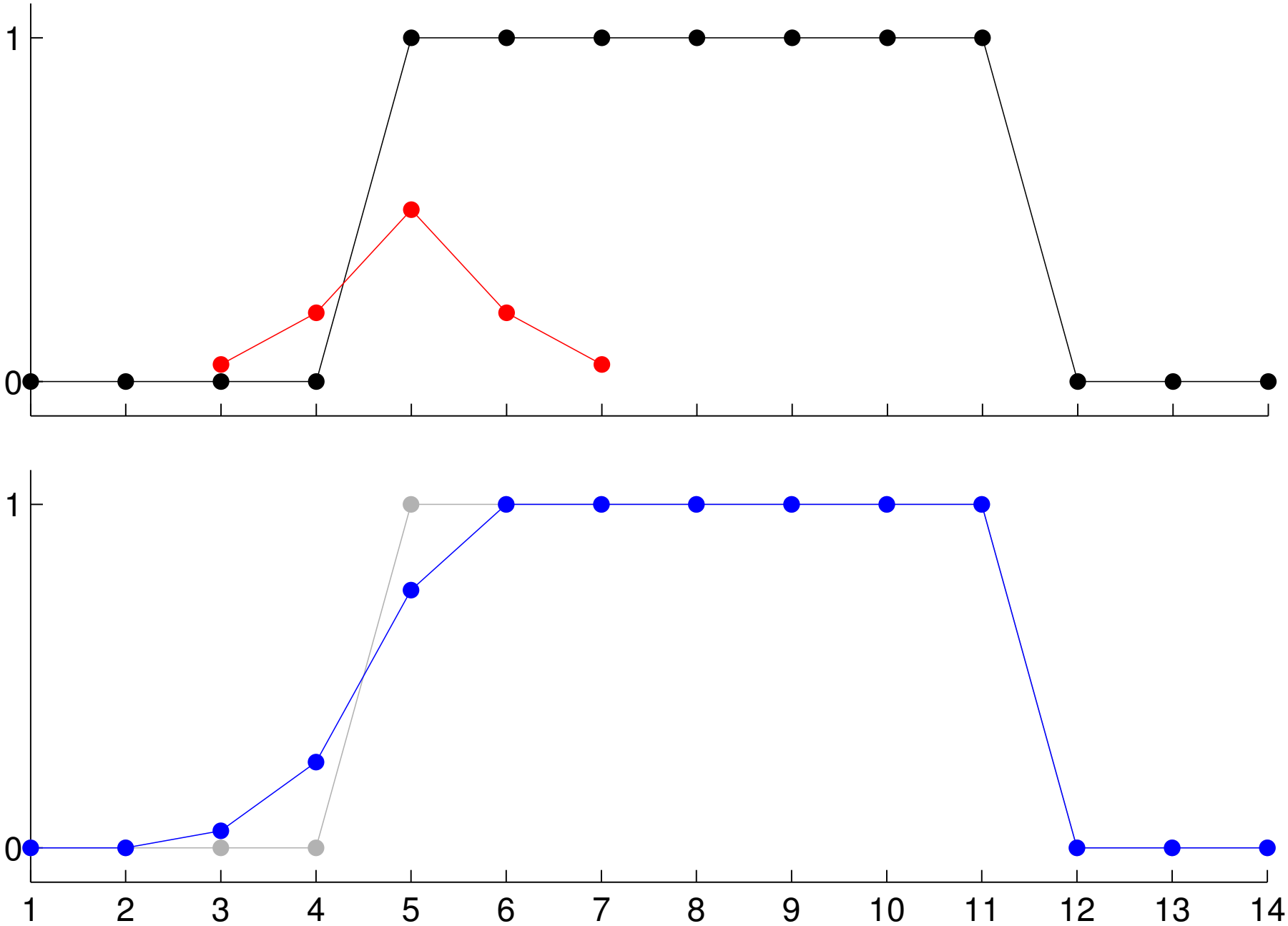
Convolution, position 3



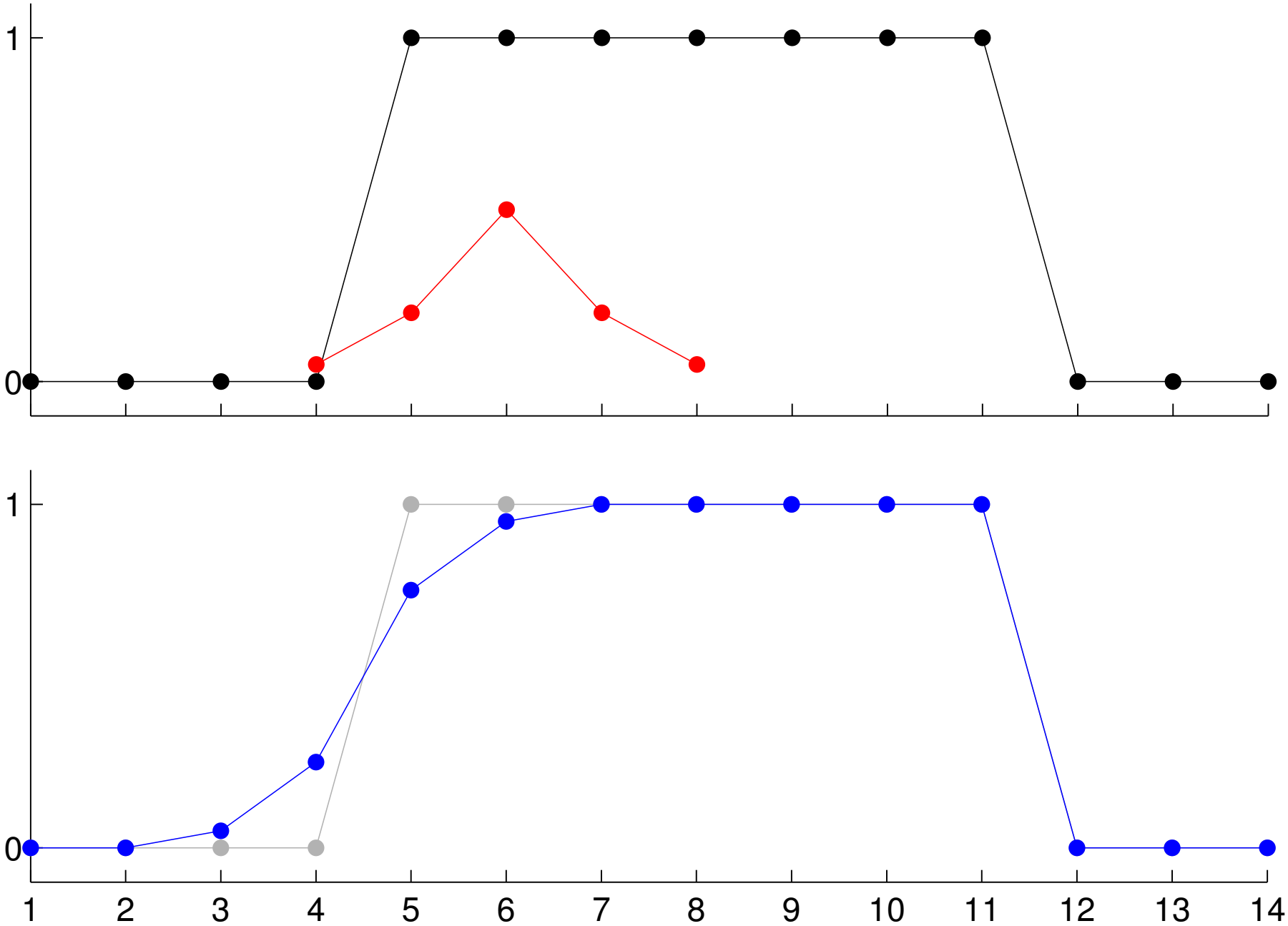
Convolution, position 4



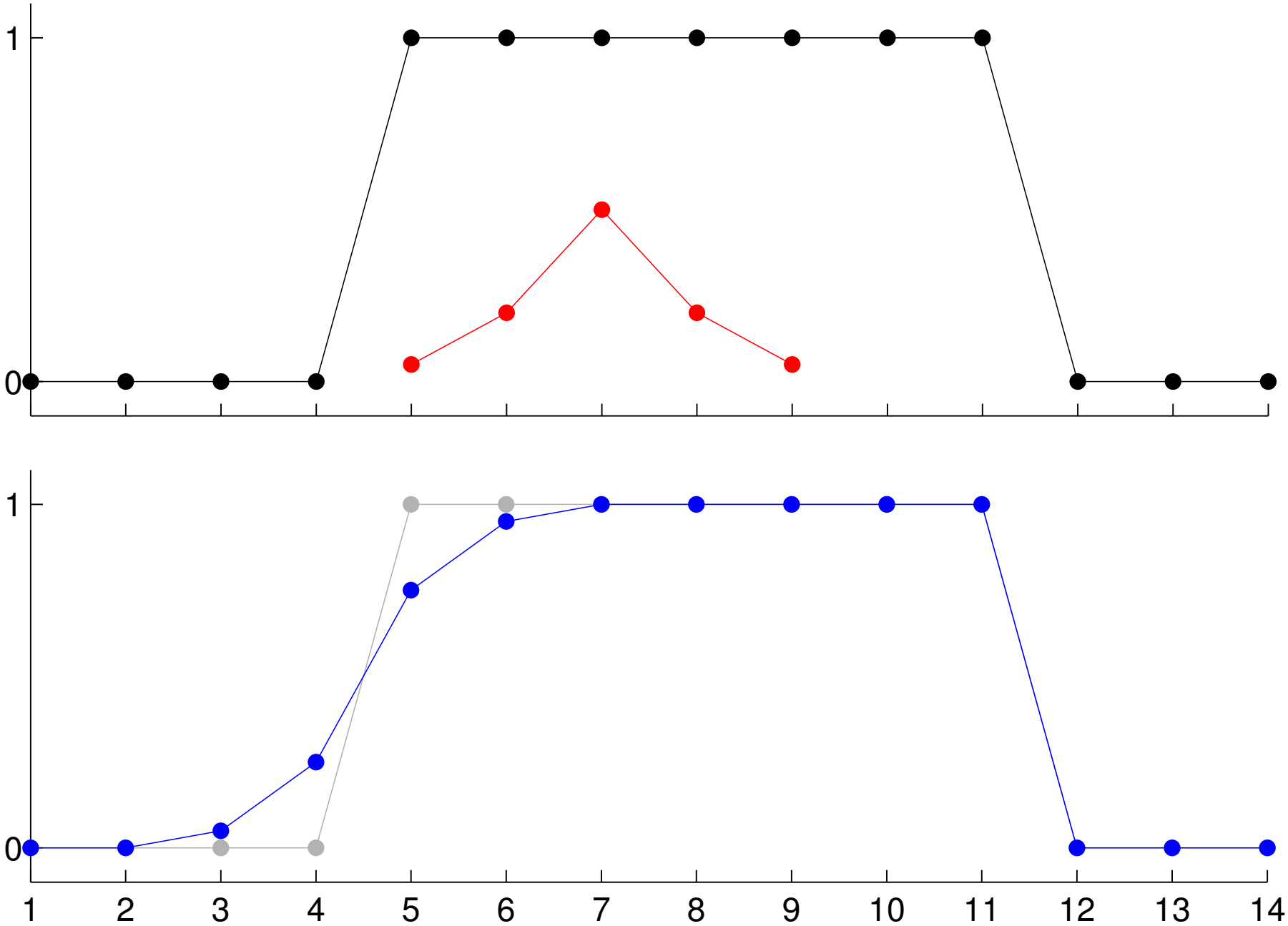
Convolution, position 5



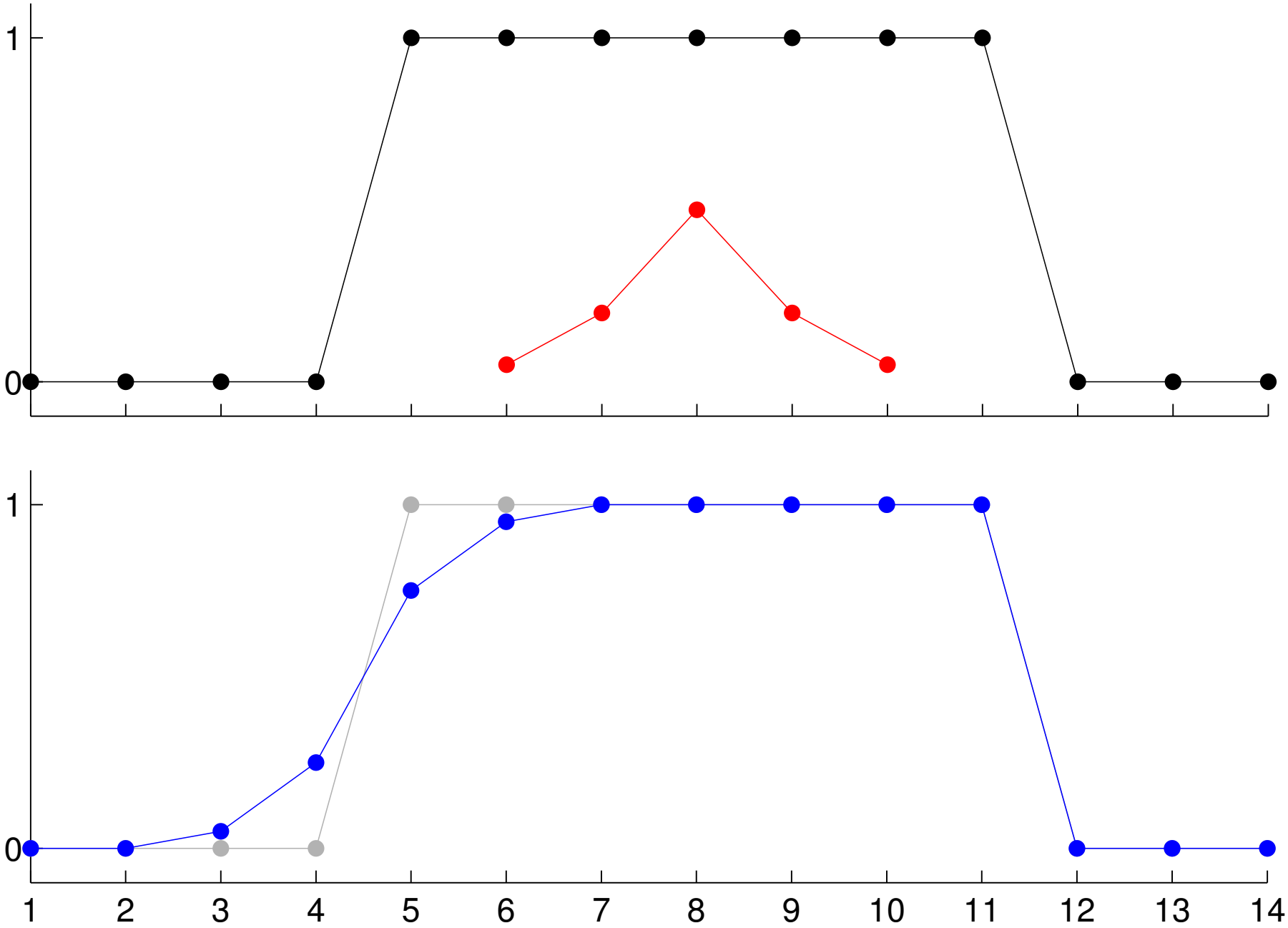
Convolution, position 6



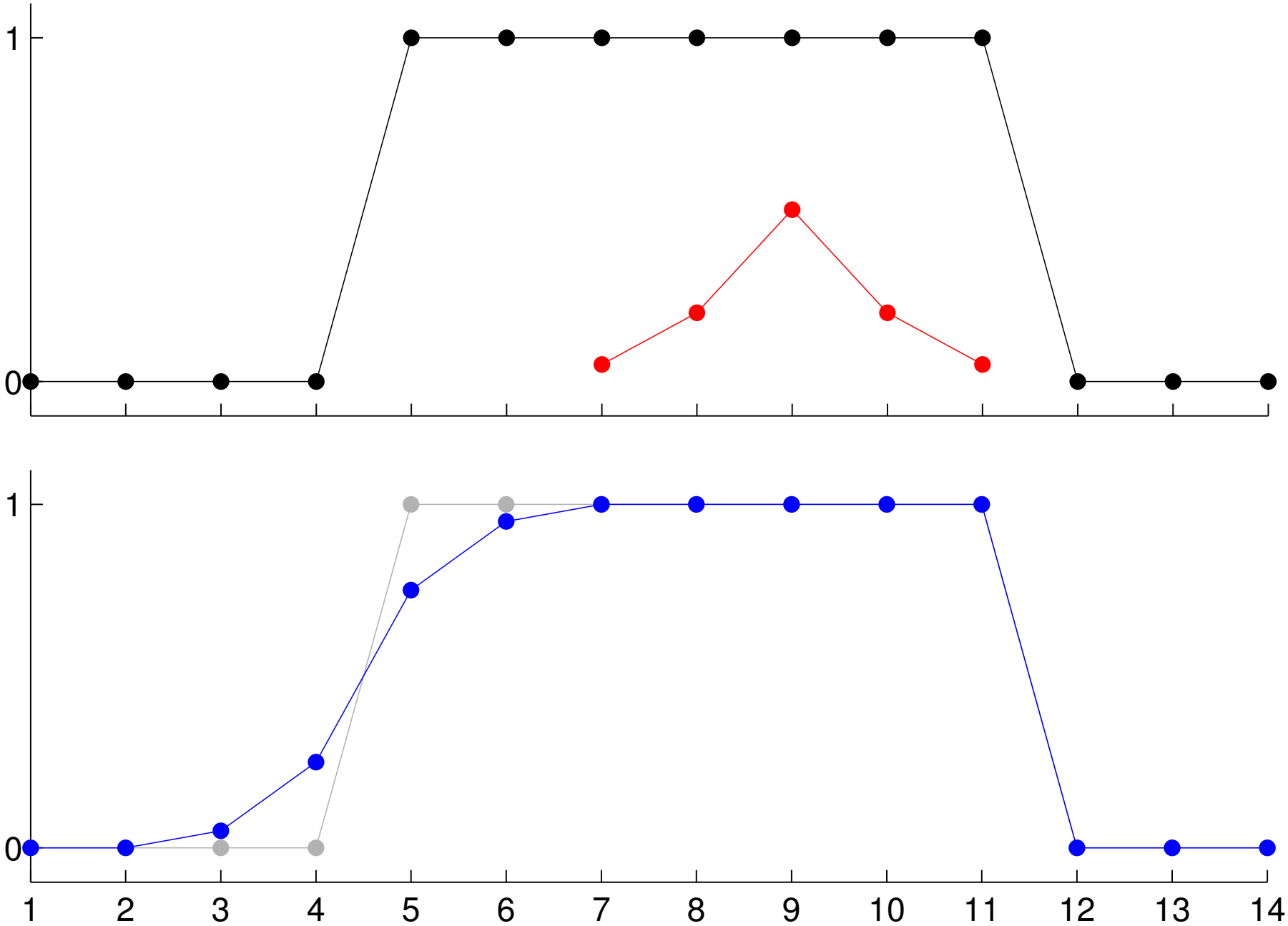
Convolution, position 7



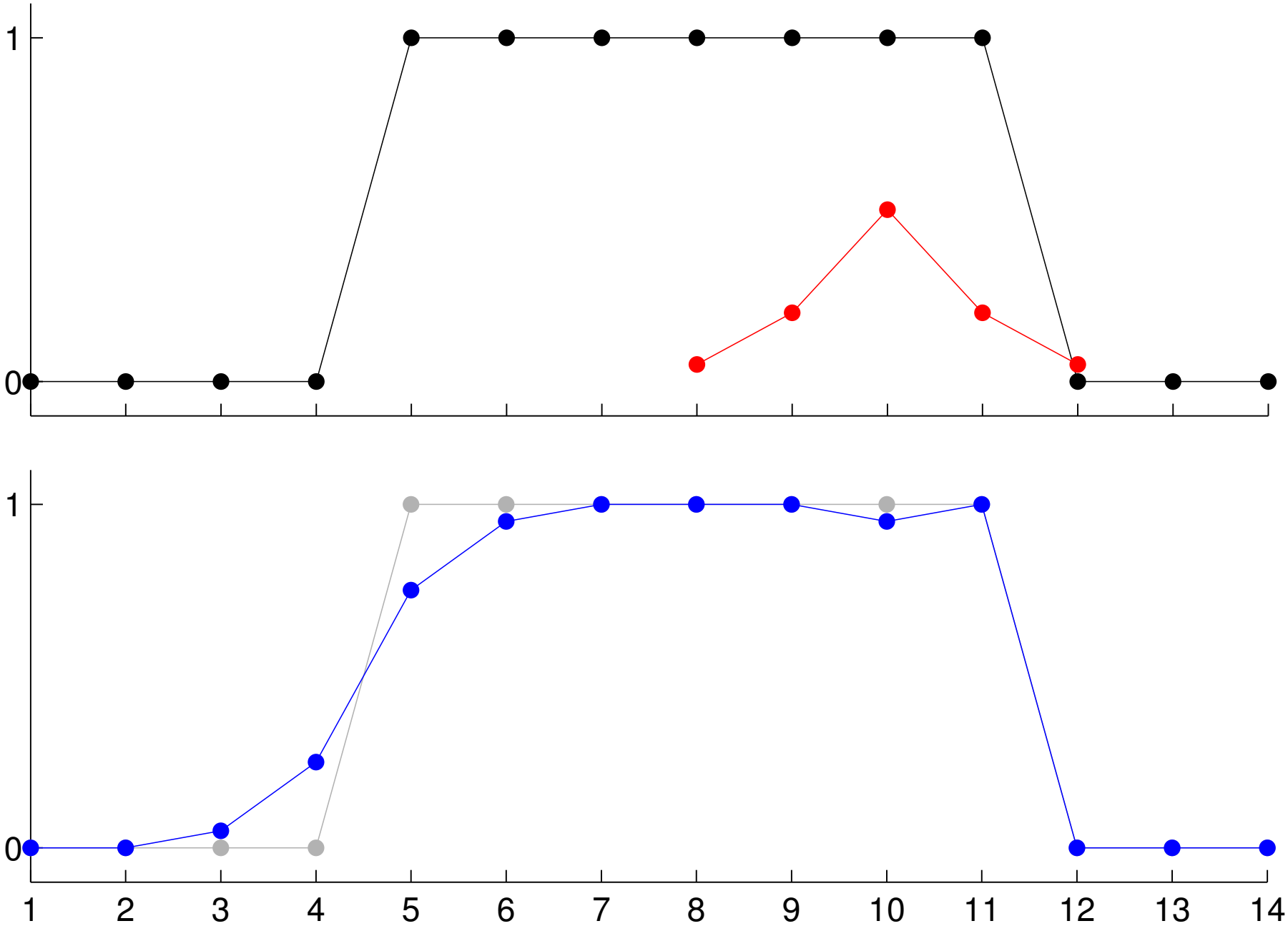
Convolution, position 8



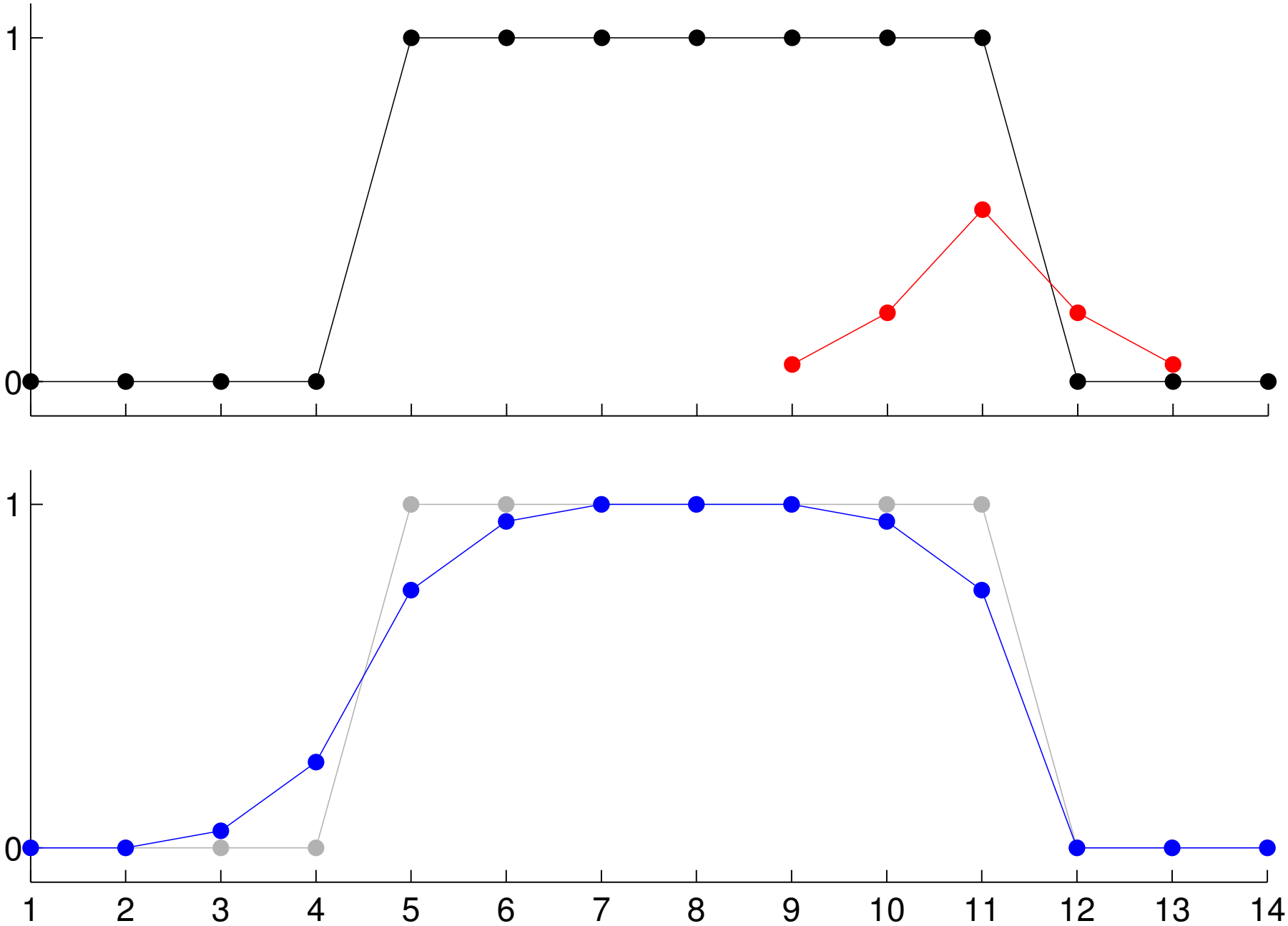
Convolution, position 9



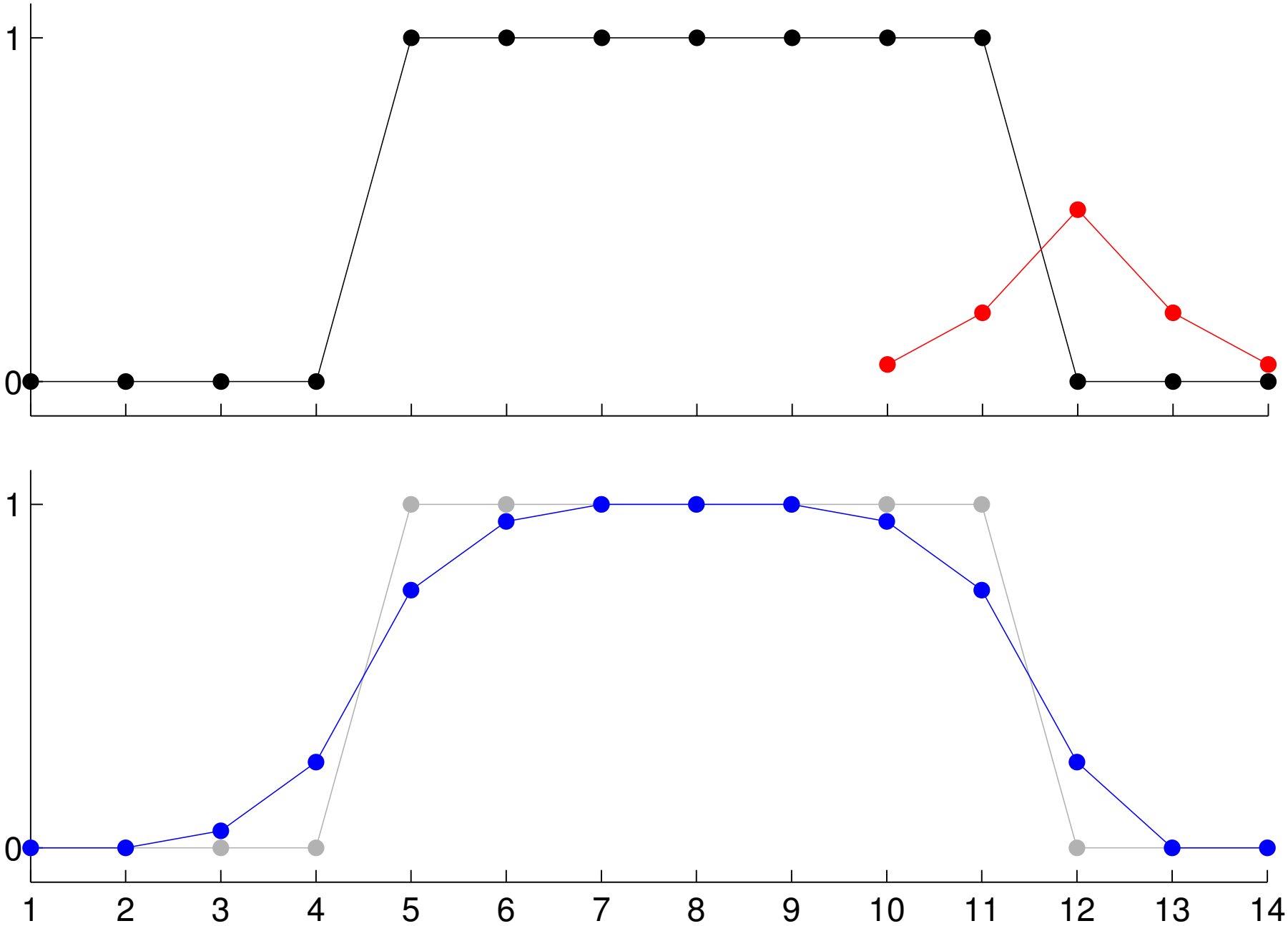
Convolution, position 10



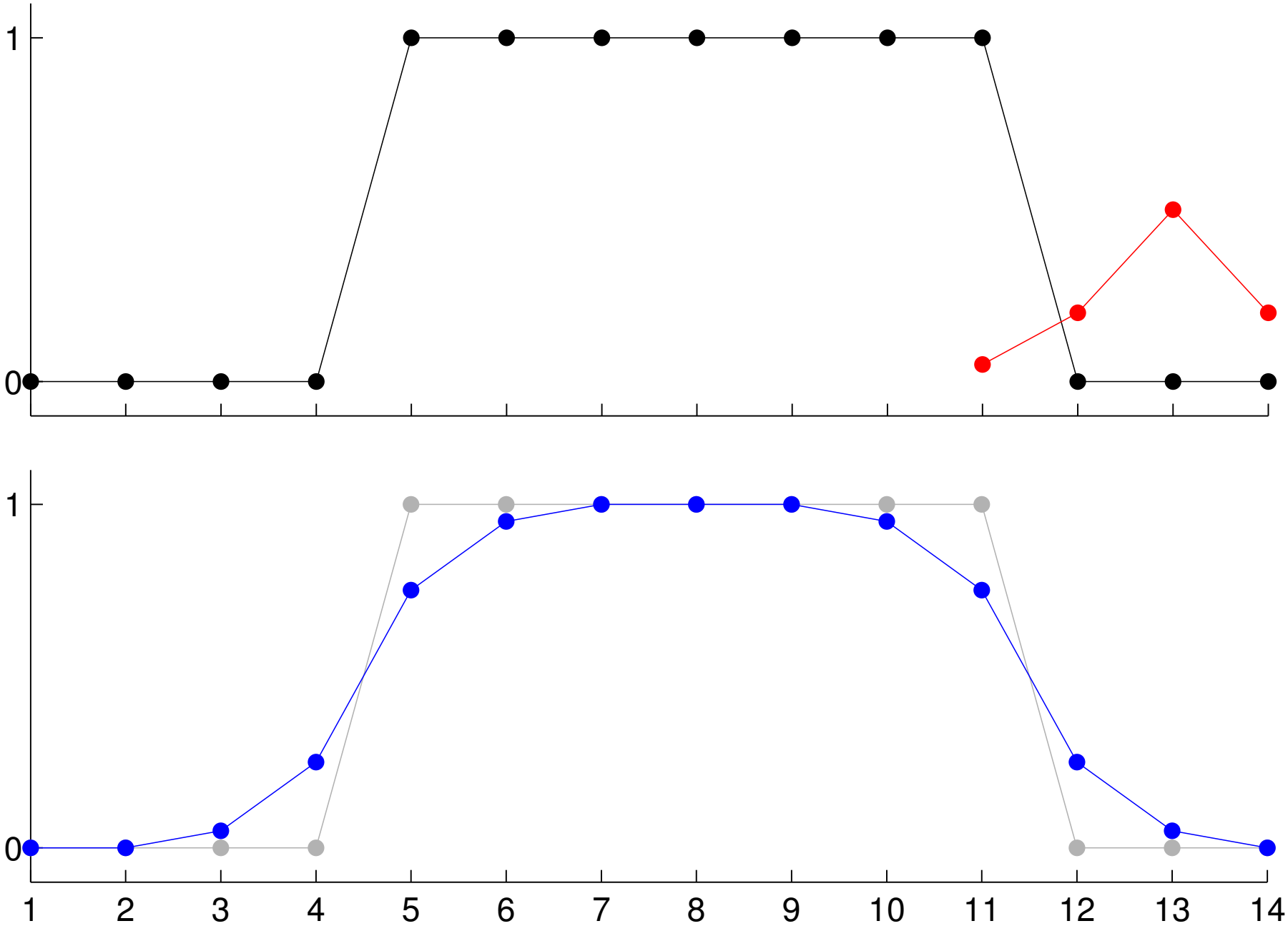
Convolution, position 11



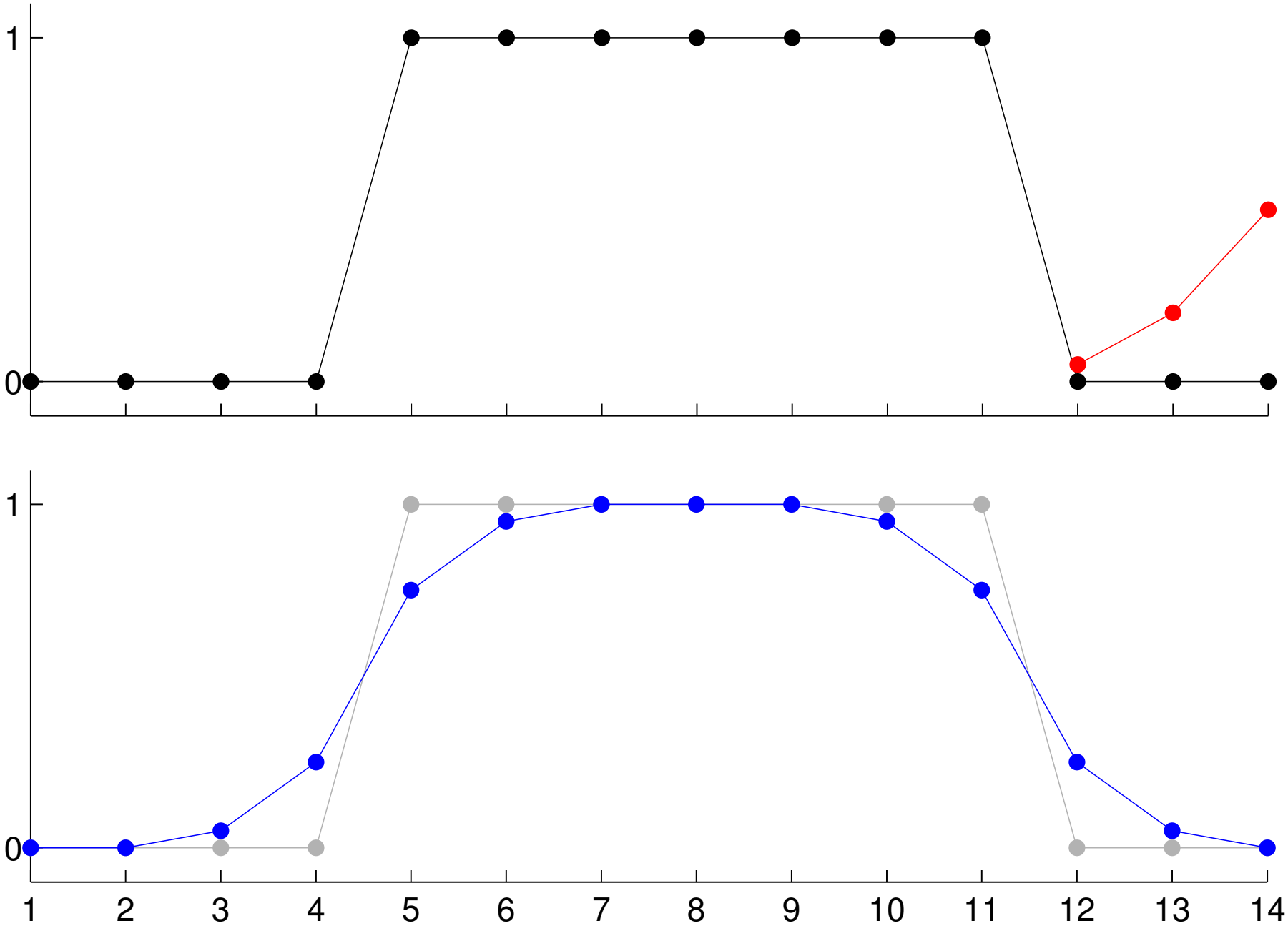
Convolution, position 12



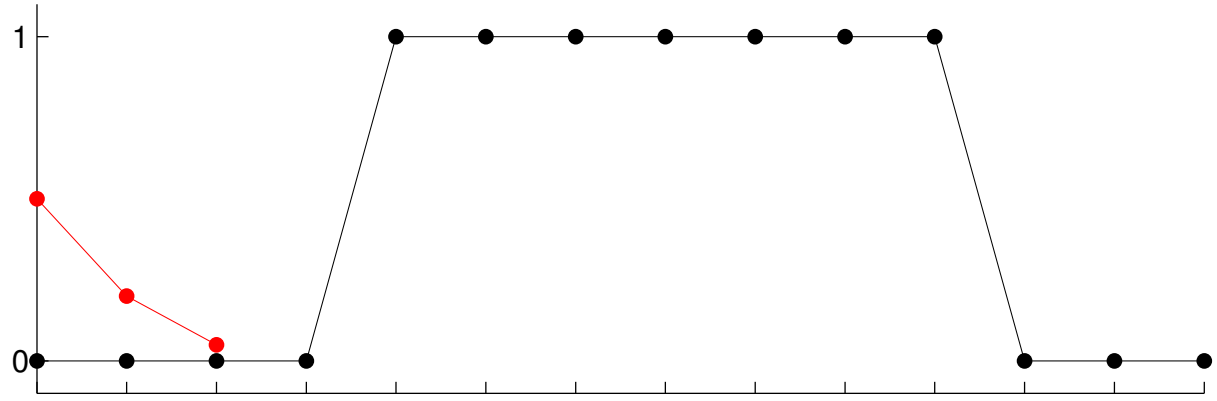
Convolution, position 13



Convolution, position 14

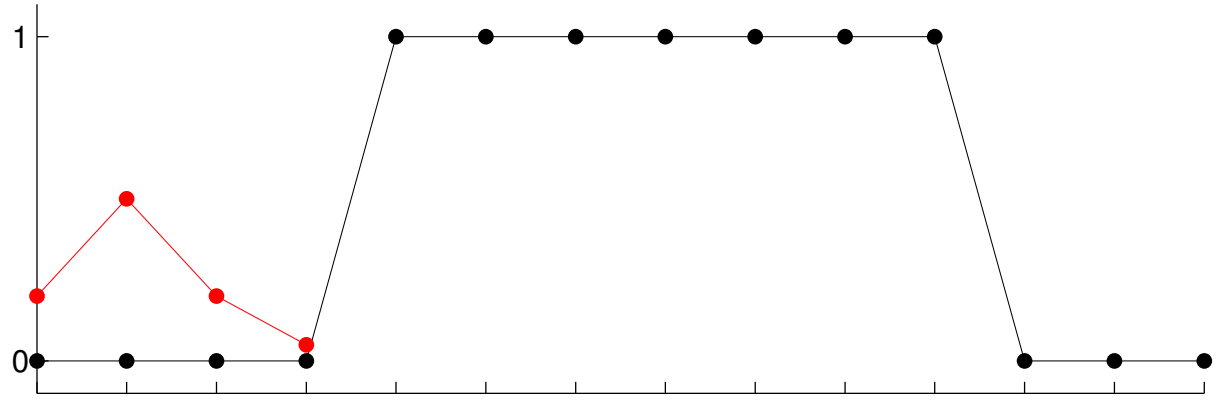


Convolution matrix, positions up to 1



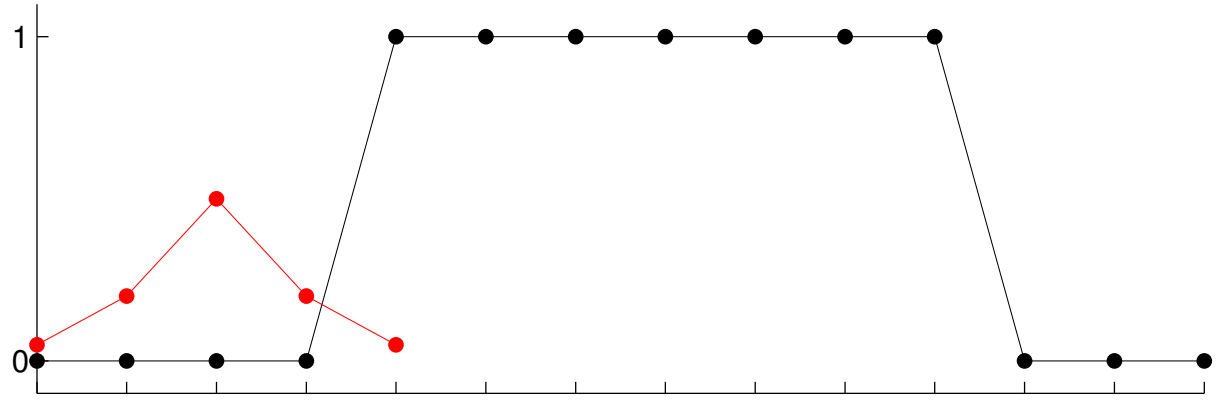
$$\begin{bmatrix} 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & & & & & & & & & & \\ & & & & & & & & & & & & & \\ & & & & & & & & & & & & & \\ & & & & & & & & & & & & & \end{bmatrix}$$

Convolution matrix, positions up to 2



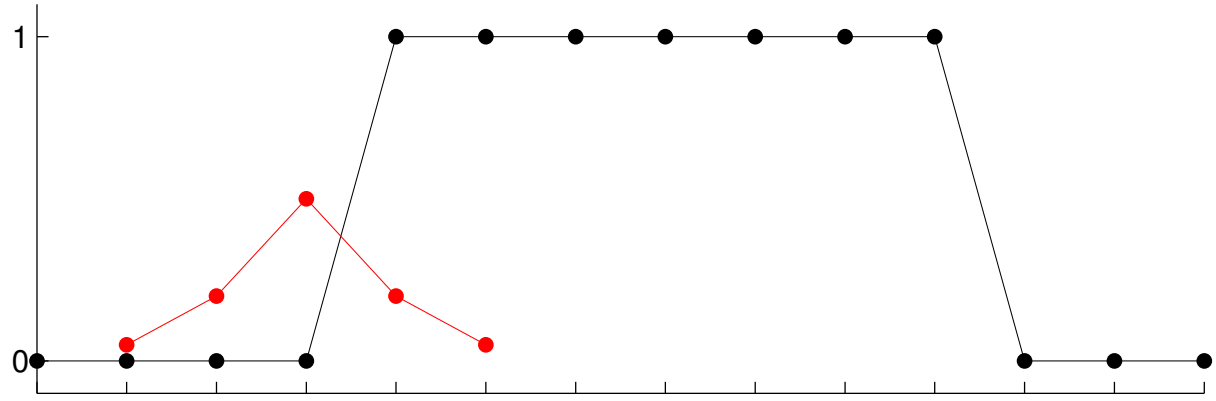
$$\begin{bmatrix} 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Convolution matrix, positions up to 3



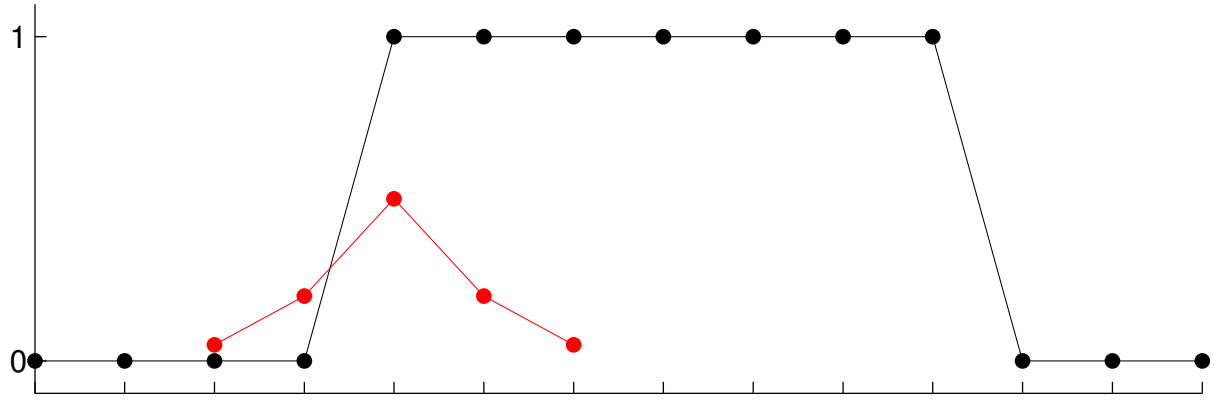
$$\begin{bmatrix} 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Convolution matrix, positions up to 4



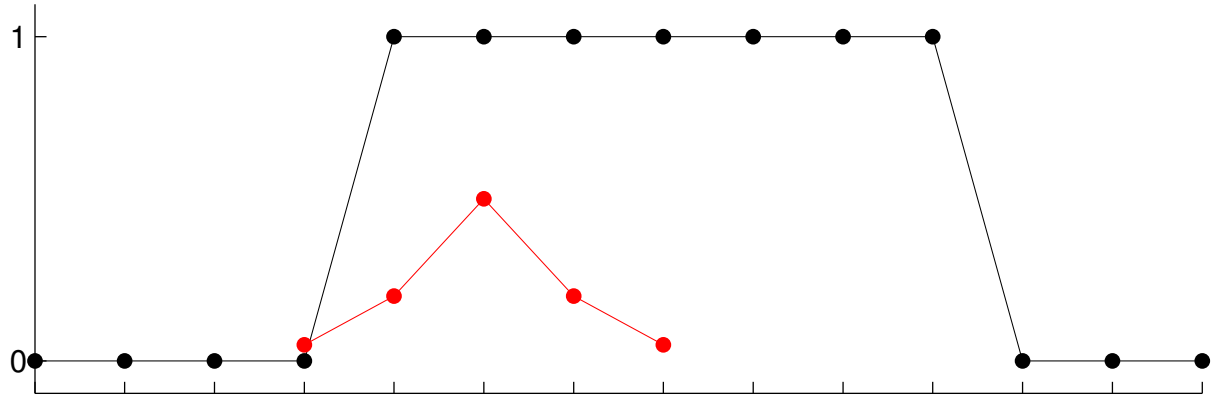
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Convolution matrix, positions up to 5



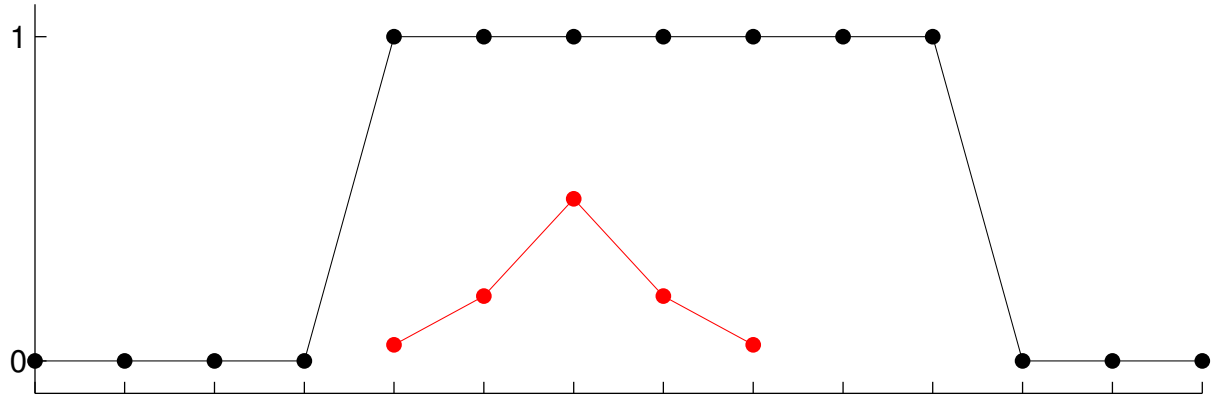
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Convolution matrix, positions up to 6



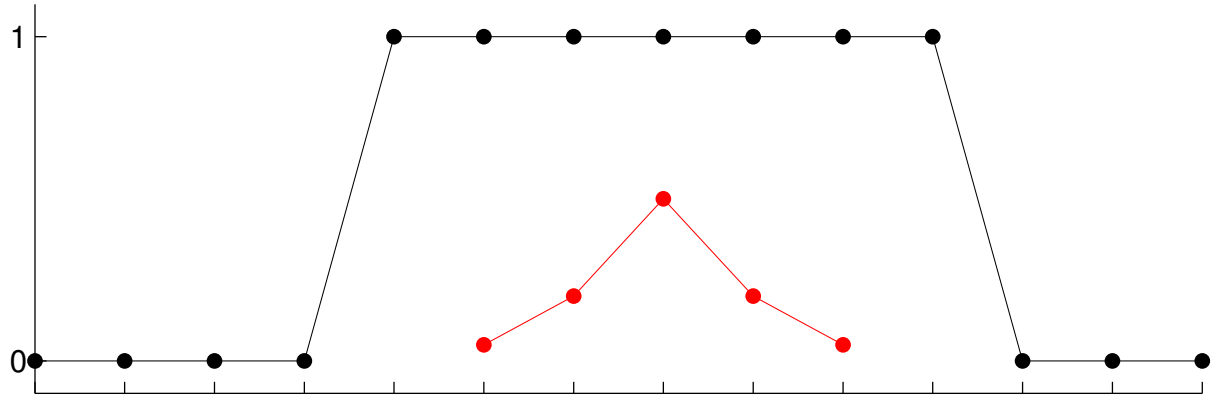
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Convolution matrix, positions up to 7



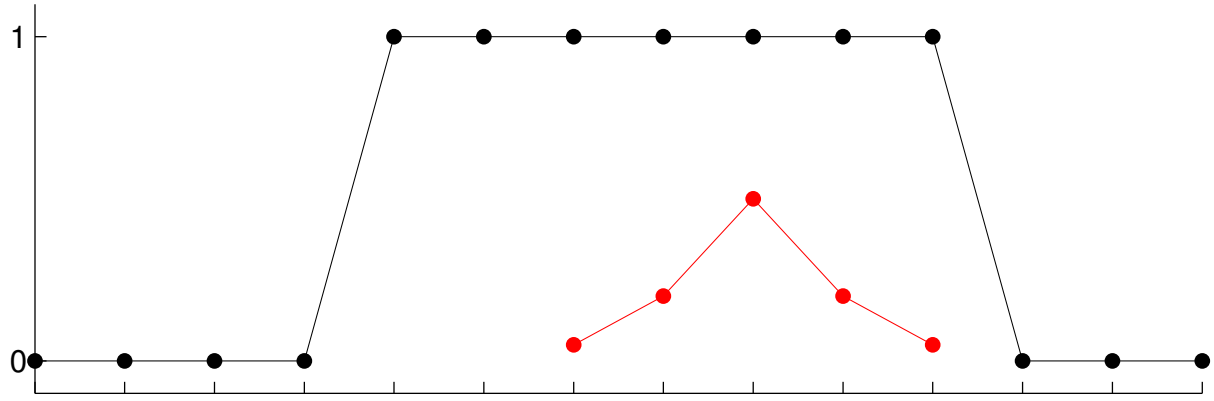
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

Convolution matrix, positions up to 8



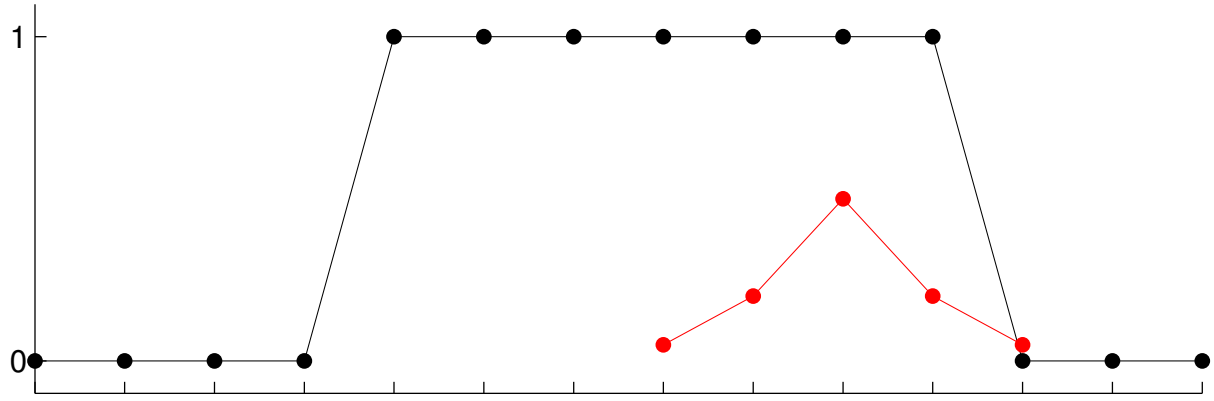
$$
\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Convolution matrix, positions up to 9



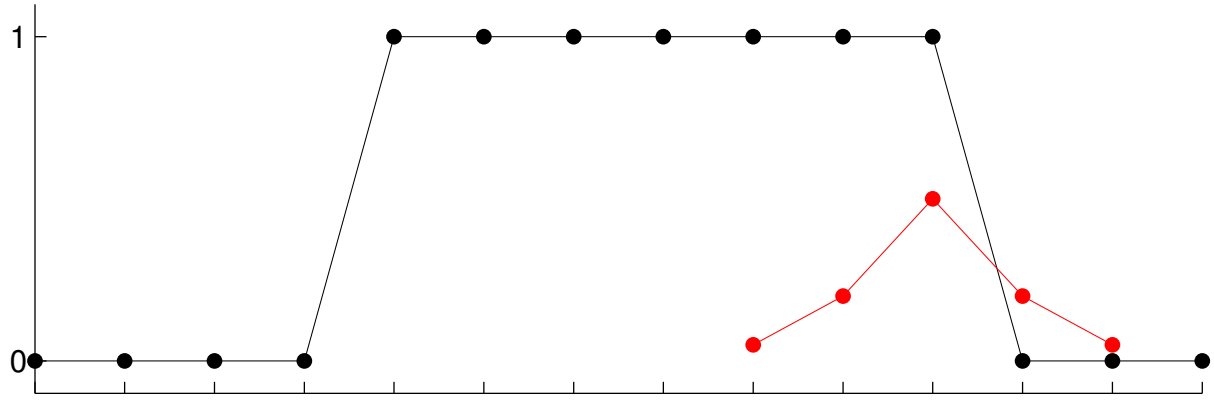
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 \\
\\
\\
\\
\end{bmatrix}$$

Convolution matrix, positions up to 10



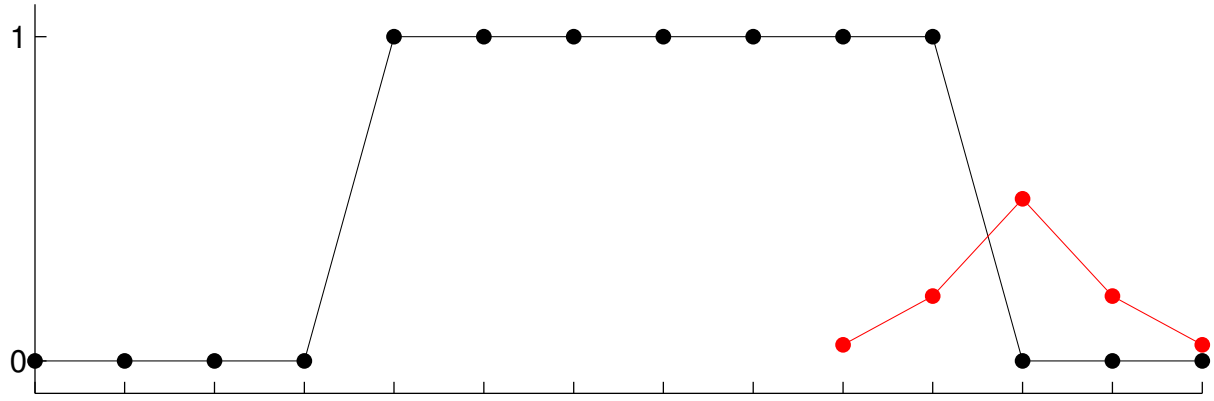
$$
\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0
\end{bmatrix}
$$

Convolution matrix, positions up to 11



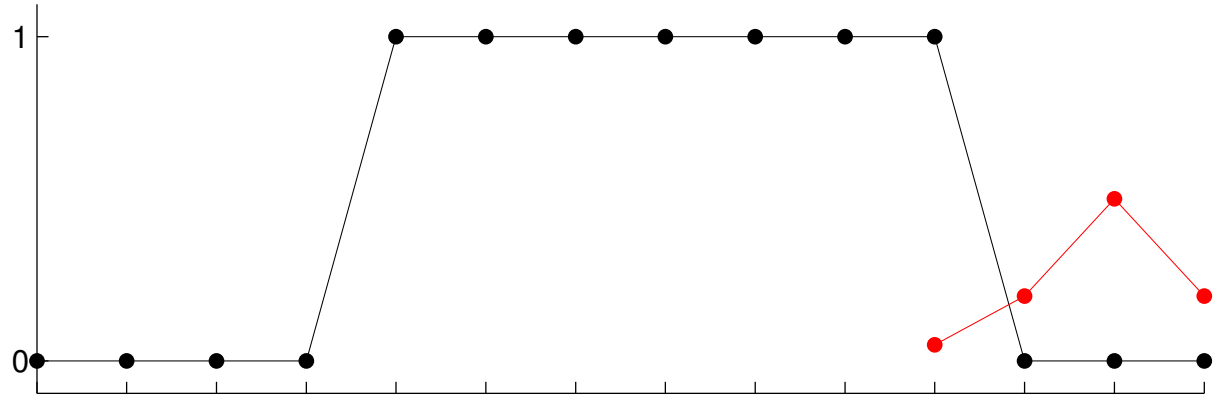
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0
\end{bmatrix}$$

Convolution matrix, positions up to 12



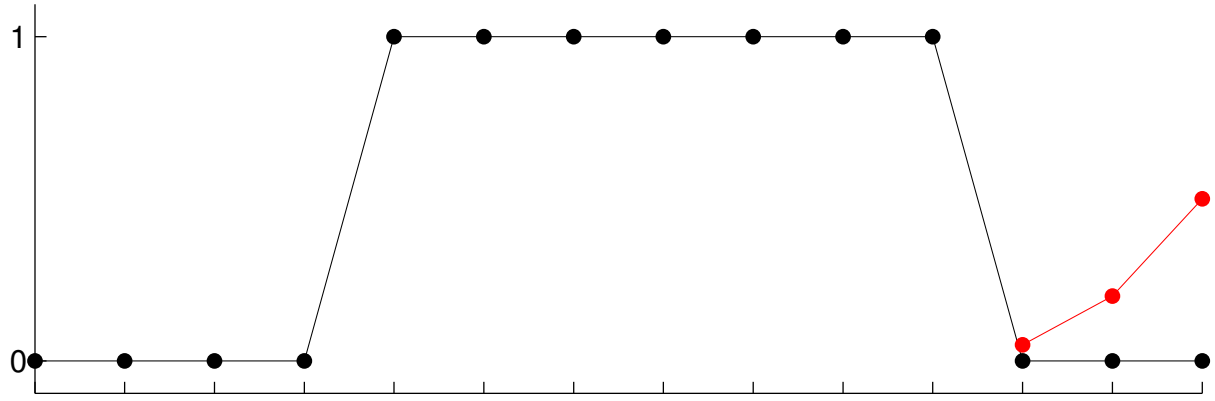
$$\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05
\end{bmatrix}$$

Convolution matrix, positions up to 13



$$
\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20
\end{bmatrix}
$$

Convolution matrix, positions up to 14



$$
\begin{bmatrix}
0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 & 0.05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50 & 0.20 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.05 & 0.20 & 0.50
\end{bmatrix}
$$

These are the direct and inverse problems related to vowel sounds

Direct problem,
from **cause to effect**:

Given the excitation signal $s(t)$ and the filter $p(t)$, determine the resulting vowel sound signal

$$v(t) = (p * s)(t),$$

where $*$ denotes convolution.

This is a well-posed task, easily implemented as multiplication in the Fourier or $\mathcal{Z}$-transform domain.

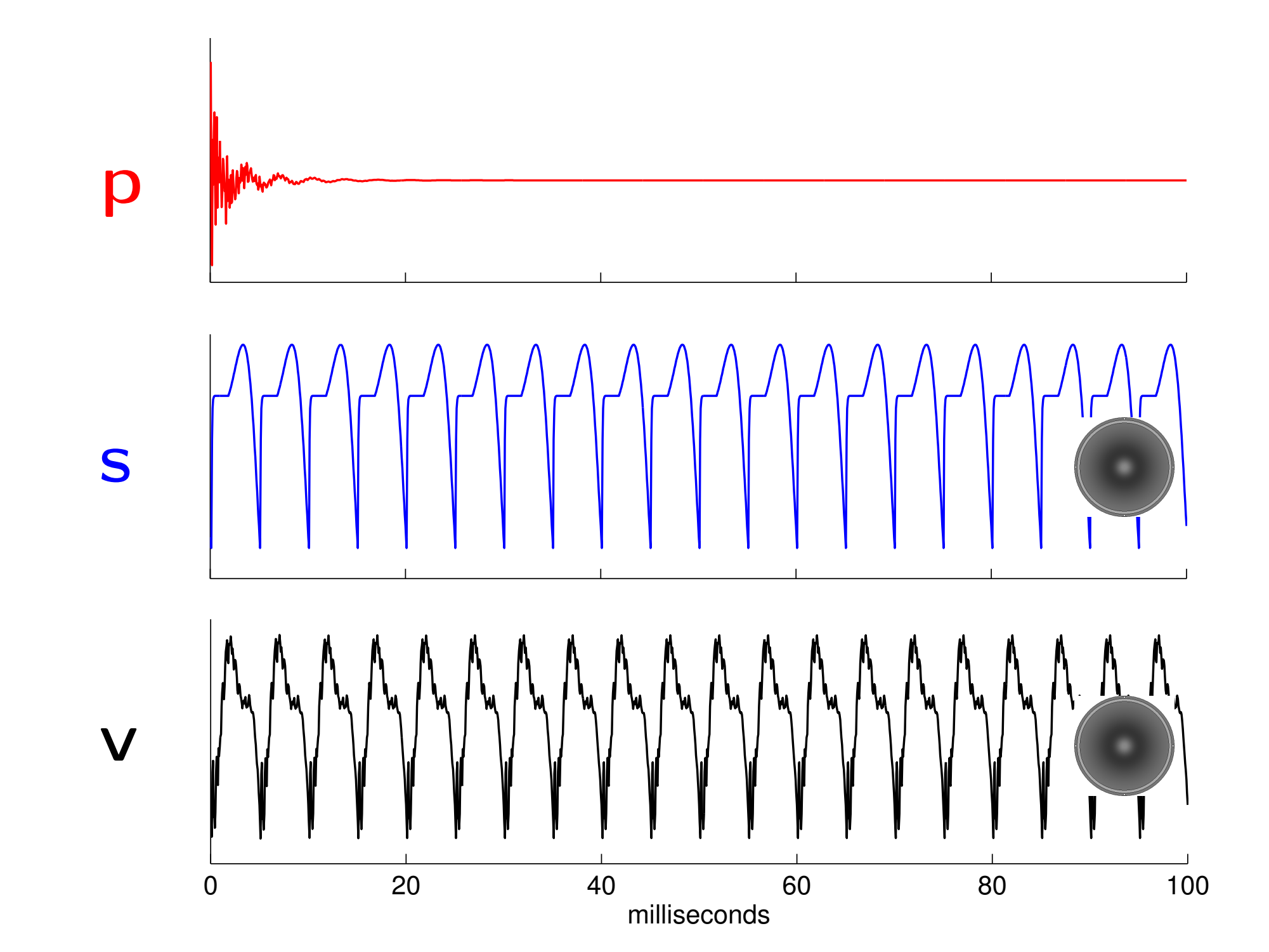
Inverse problem,
from **effect to cause**:

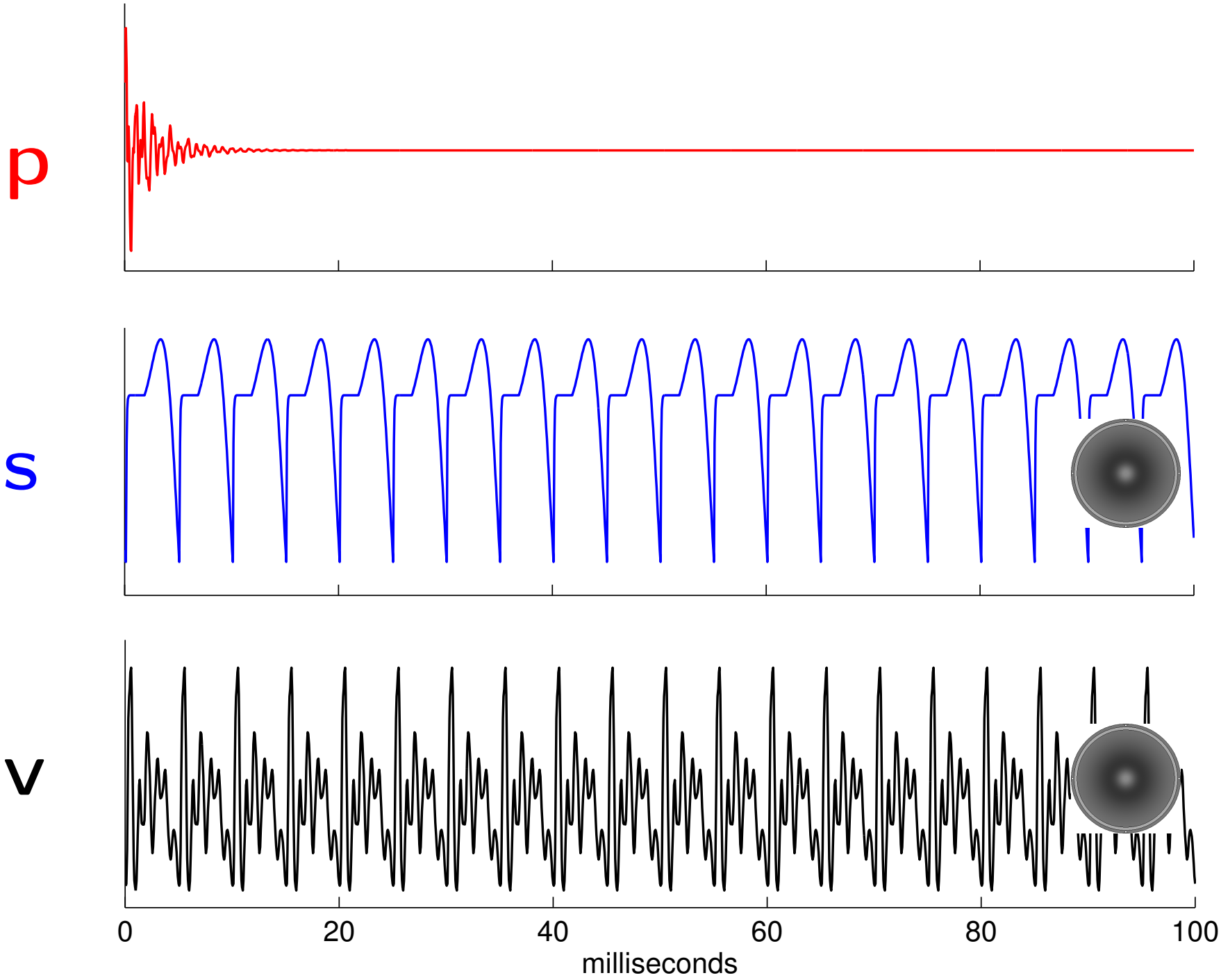
Given a microphone recording $v(t)$ of a vowel sound, use the model

$$v(t) = (p * s)(t) + \varepsilon.$$

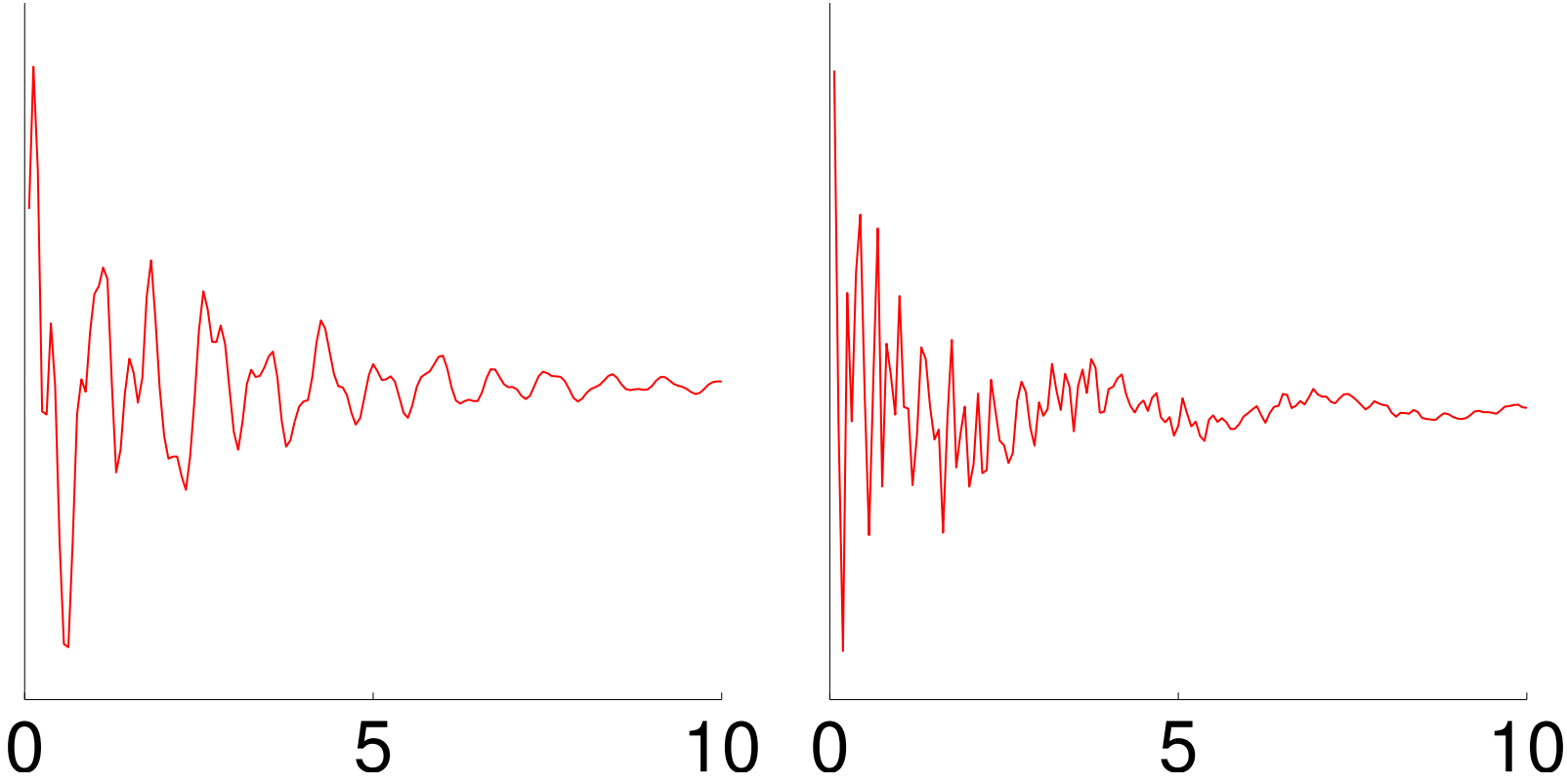
Recover the excitation signal $s(t)$ and the filter $p(t)$.

This blind deconvolution problem is called Glottal Inverse Filtering (GIF).


p
s
v
0
20
40
60
80
100
milliseconds


p
s
v
0
20
40
60
80
100
milliseconds

Close-ups of filters for vowels /a/ and /i/



The main application of GIF is synthetic speech

Here are examples of low-quality and high-quality speech signals generated by a computer (it is not a person speaking).

Sample 1, low quality:  high quality: 

Sample 2, low quality:  high quality: 

The high-quality samples, developed in **Raitio, Suni, Yamagishi, Pulakka, Nurminen, Vainio** & **Alku** (2010), are based on a Hidden Markov Model (HMM).


Thank you for your attention!