Needles in haystacks

text

music

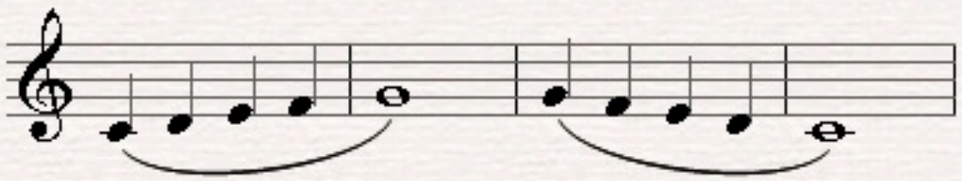
pictures

text

- Google index
  - for every word a list of documents and counts
- Google pagerank
  - deciding which documents are most important
- efficiency – distributed algorithms
  - building tables
  - retrieving documents

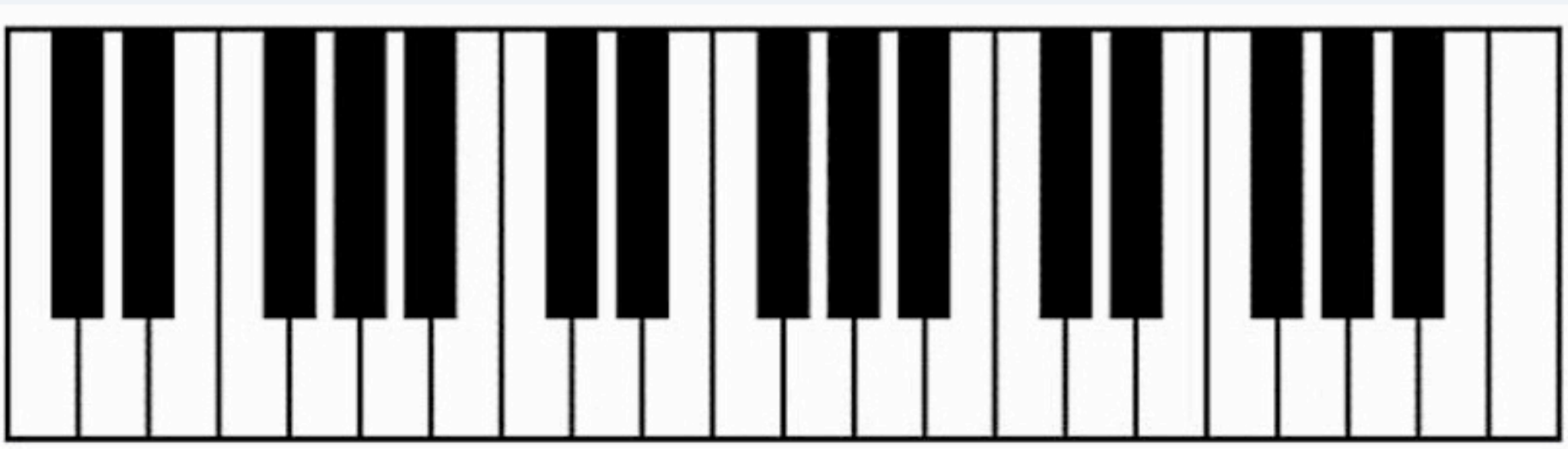
sounds

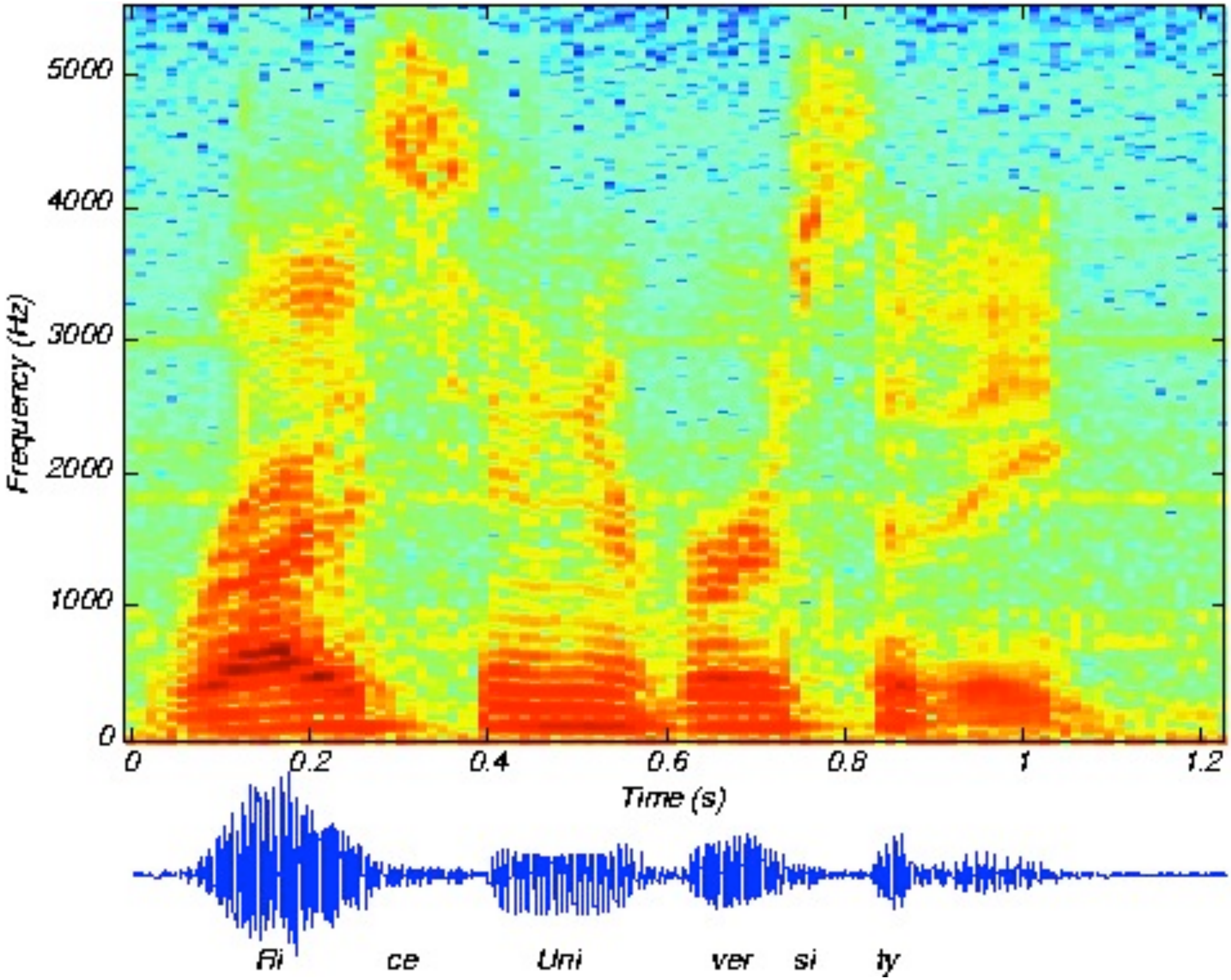
A CURVED LINE over DIFFERENT NOTES
is called A SLUR. It means play LEGATO
(smoothly connected)

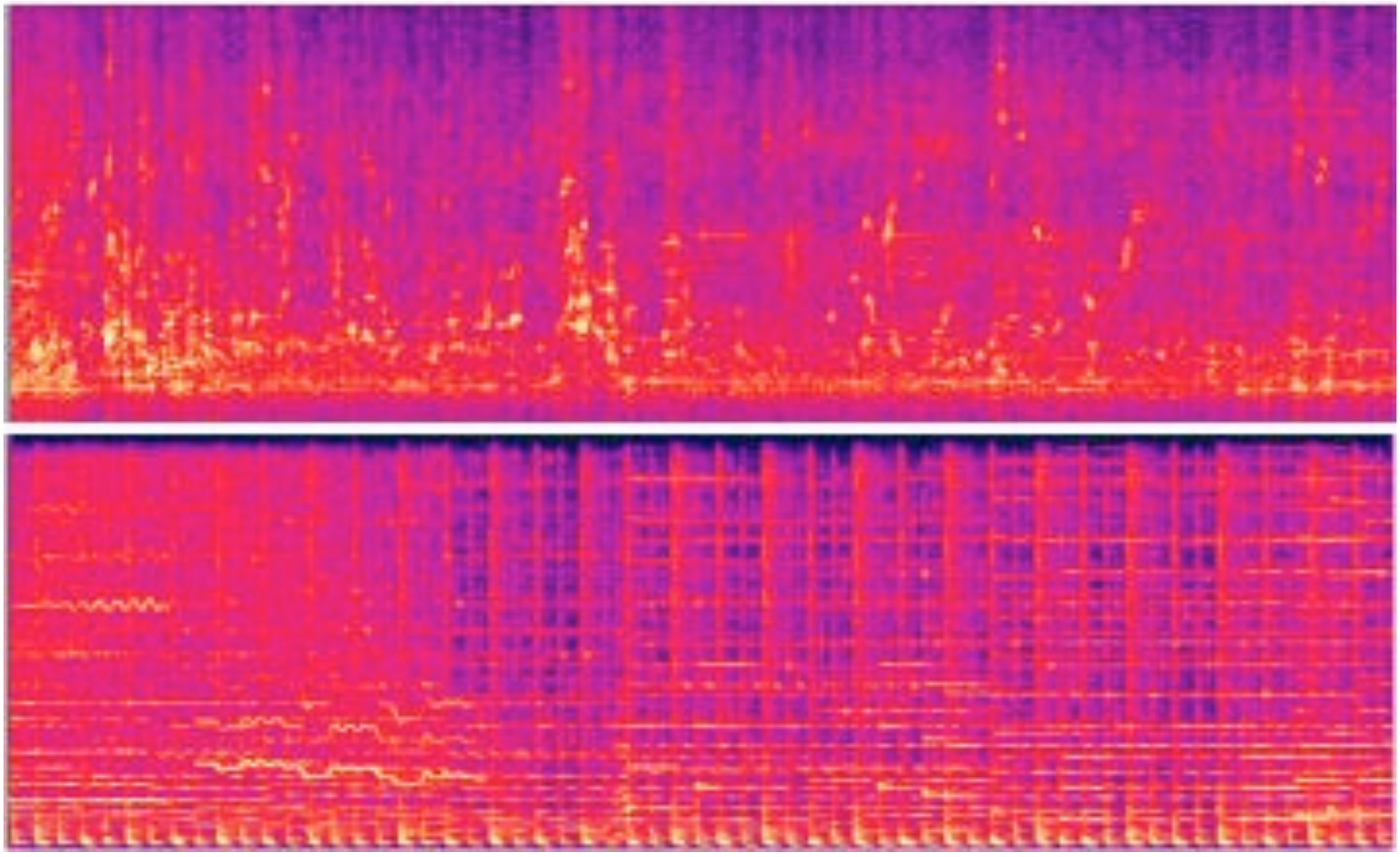


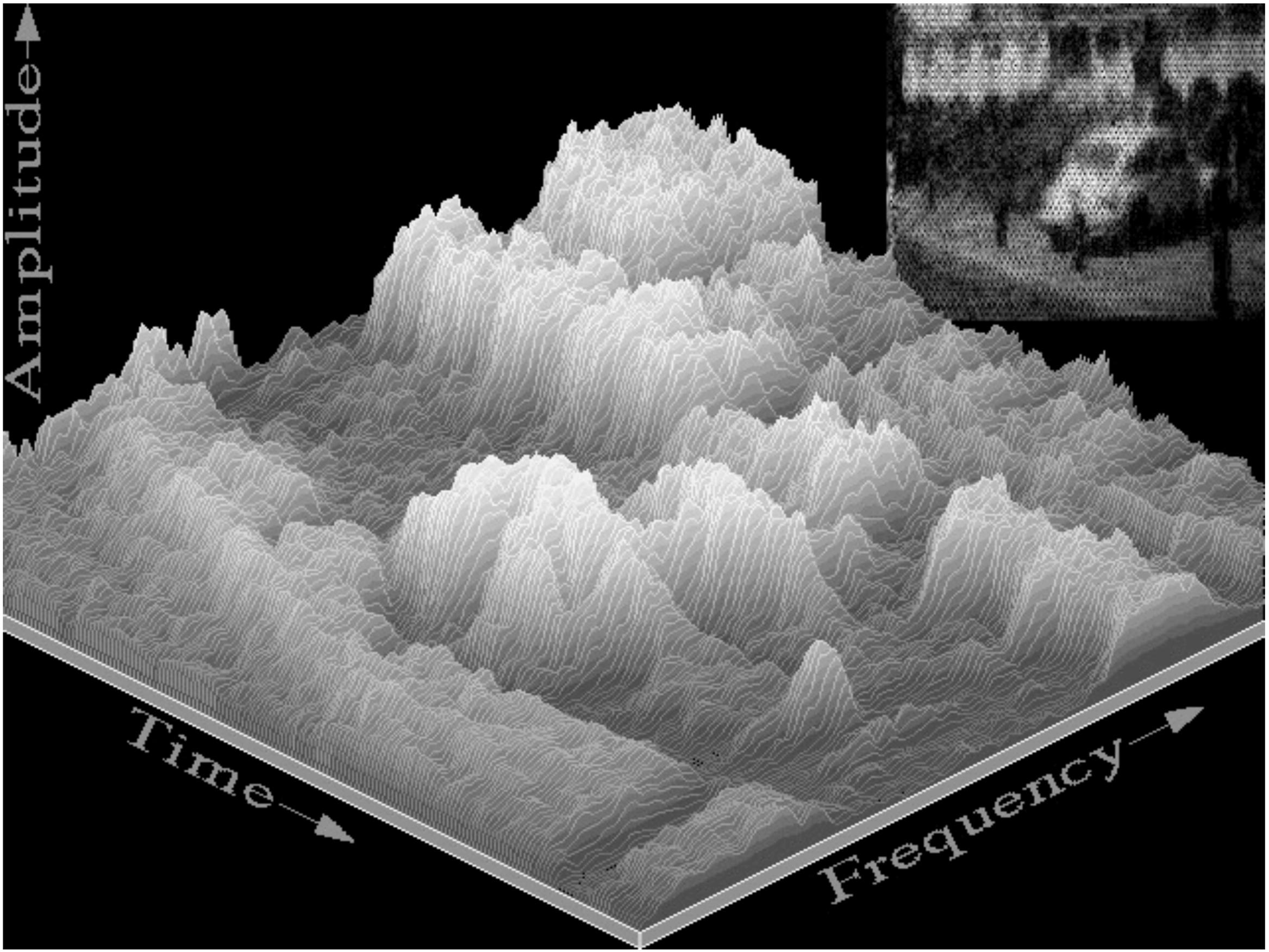
Slurs often divide the music into PHRASES.
A PHRASE is a musical thought or sentence

- no words

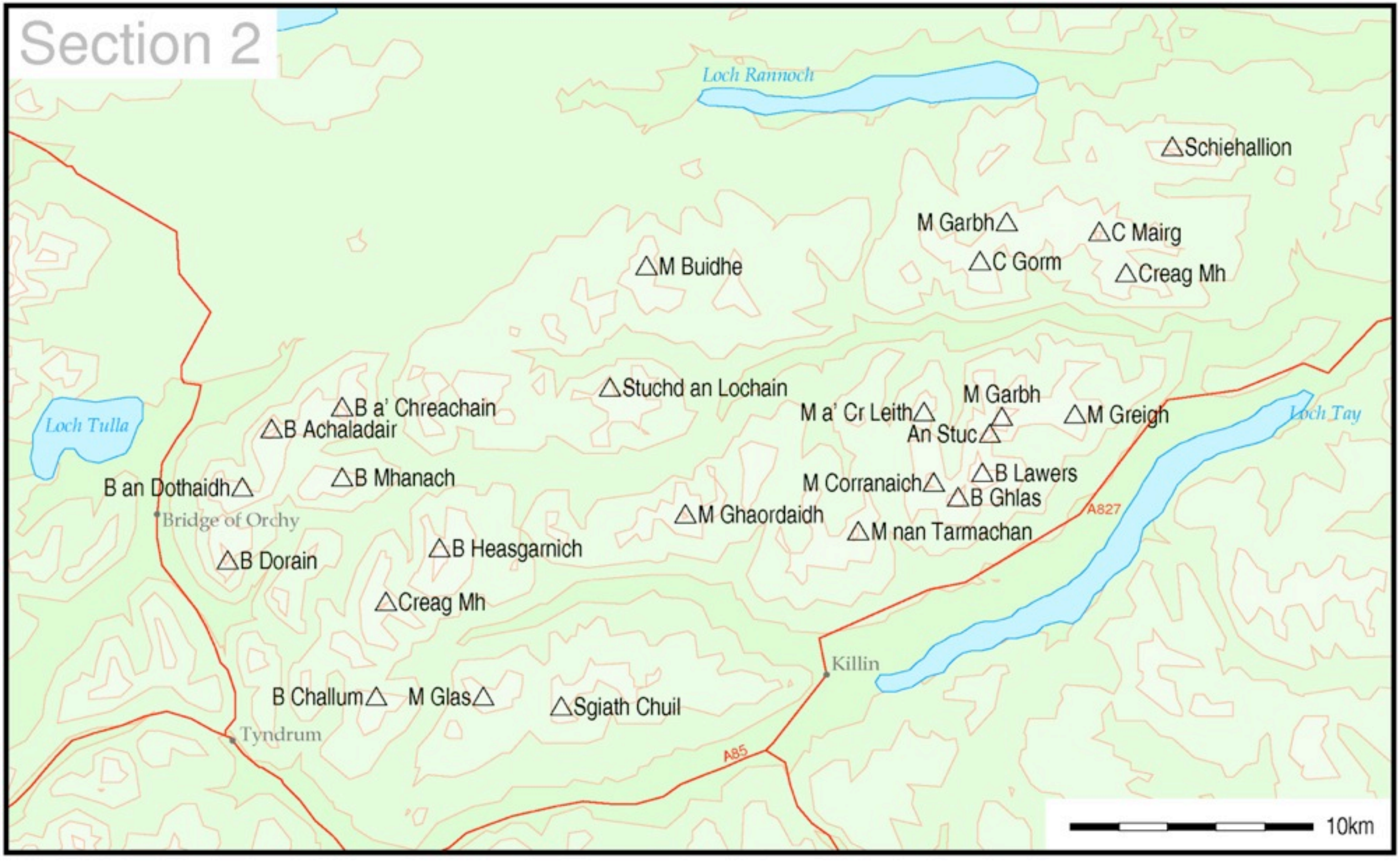



Frequency (Hz)
5000
4000
3000
2000
1000
0
0
0.2
0.4
0.6
0.8
1
1.2
Time (s)
Ri
ce
Uni
ver
si
ty

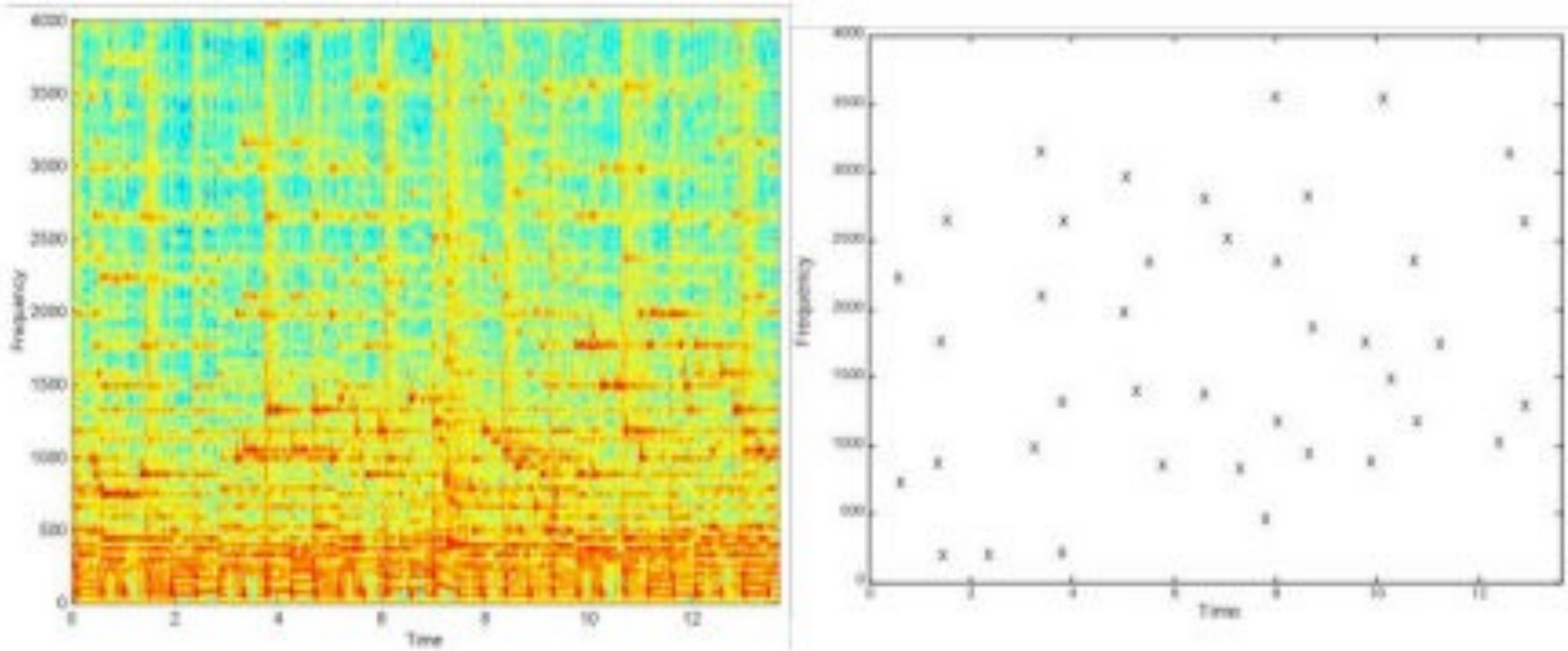


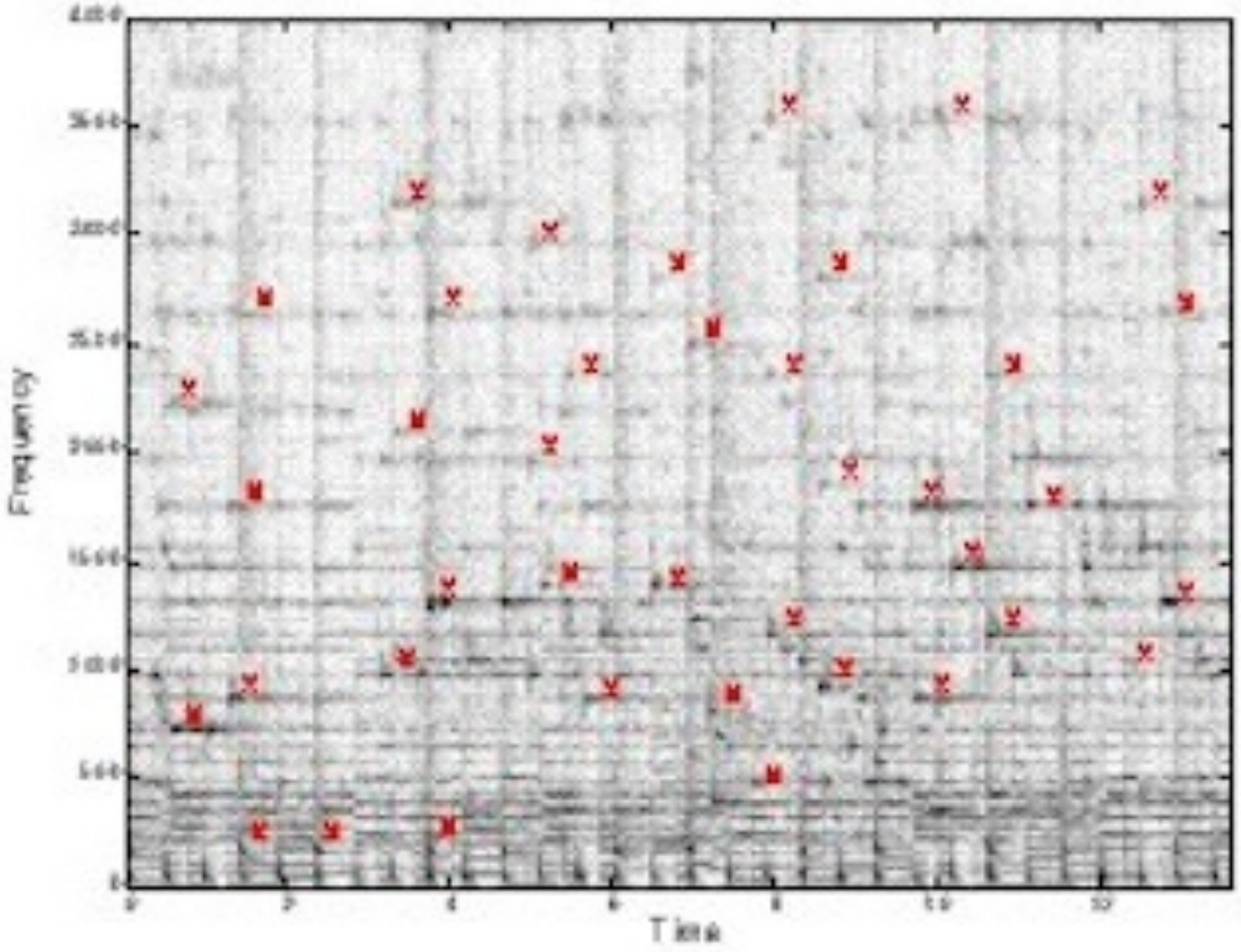

Amplitude
Time
Frequency

Section 2


Loch Rannoch
Schiehallion
M Garbh
C Mairg
C Gorm
Creag Mh
M Buidhe
Stuchd an Lochain
B a' Chreachain
B Achaladair
Loch Tulla
M a' Cr Leith
M Garbh
An Stuc
M Greigh
Loch Tay
B Lawers
M Corranaich
B Ghlas
B Mhanach
B an Dothaidh
Bridge of Orchy
M Ghaordaidh
M nan Tarmachan
A827
B Heasgarnich
B Dorain
Creag Mh
Killin
B Challum
M Glas
Sgiath Chuil
Tyndrum
A85
10km

Munros – isolated peaks




Frequency
Time

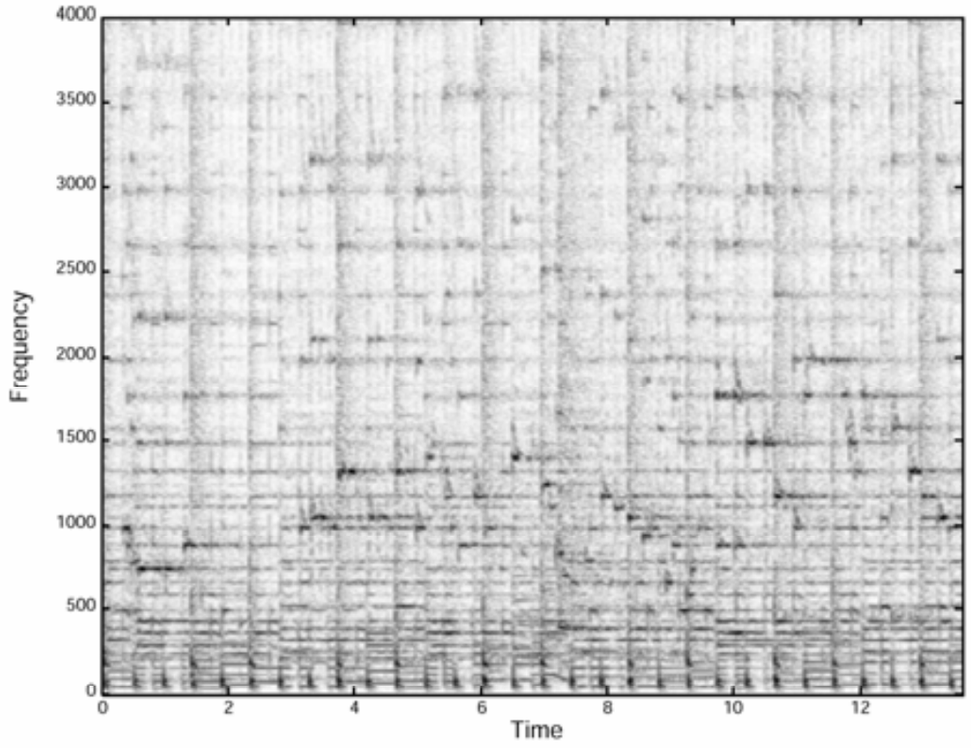


Fig. 1A - Spectrogram

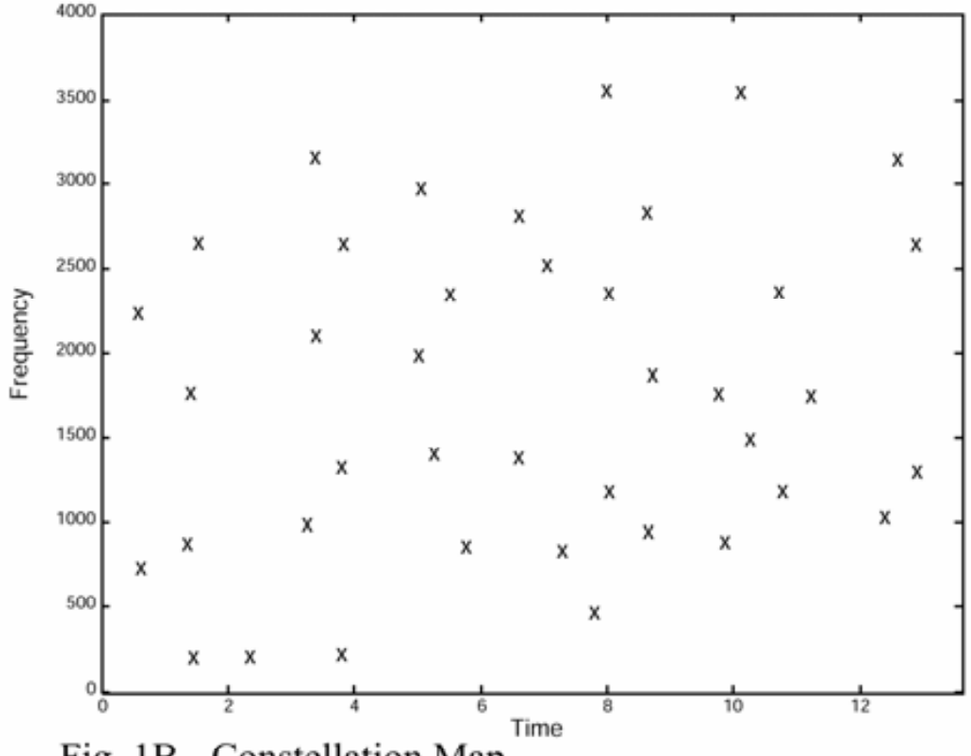


Fig. 1B - Constellation Map

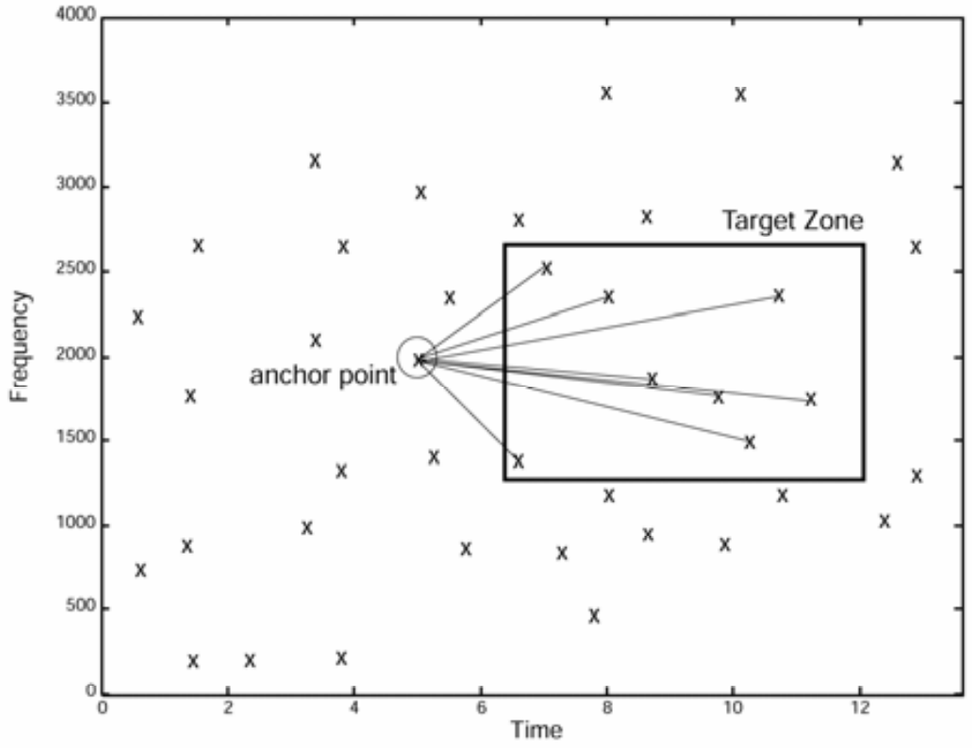


Fig. 1C - Combinatorial Hash Generation

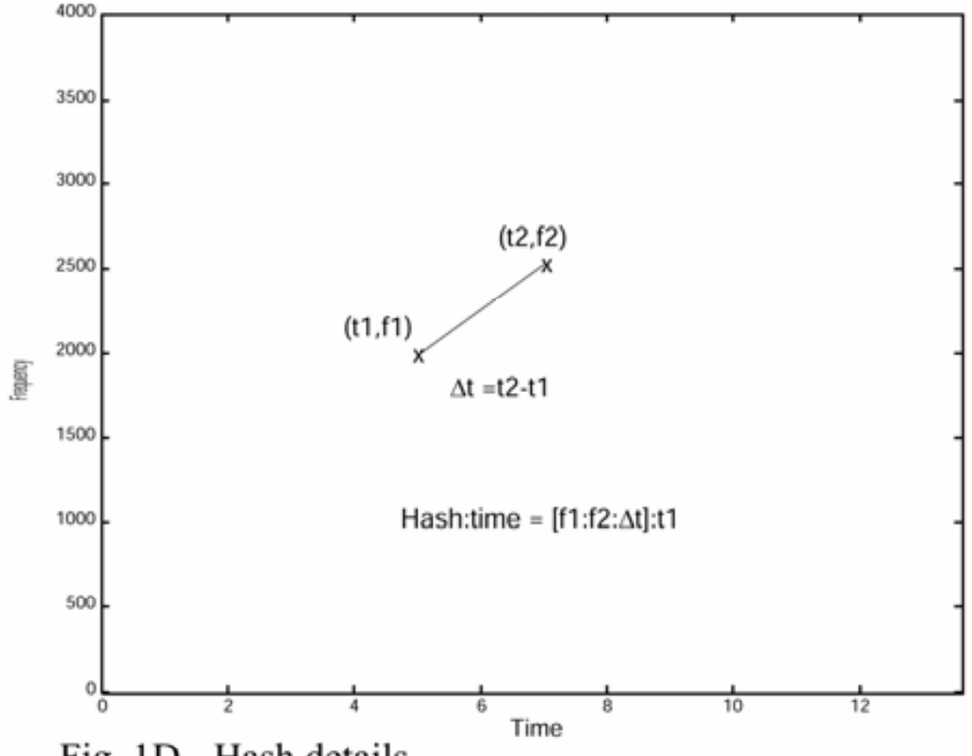


Fig. 1D - Hash details

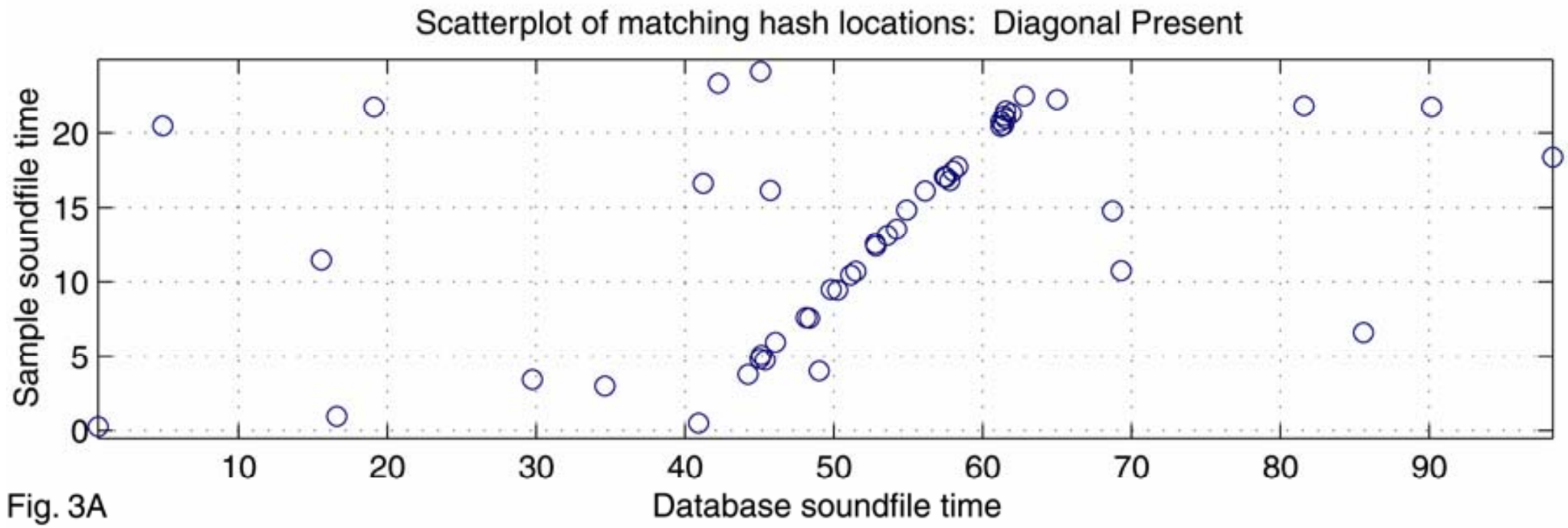

Scatterplot of matching hash locations: Diagonal Present
Sample soundfile time
0
5
10
15
20
10
20
30
40
50
60
70
80
90
Database soundfile time

Fig. 3A

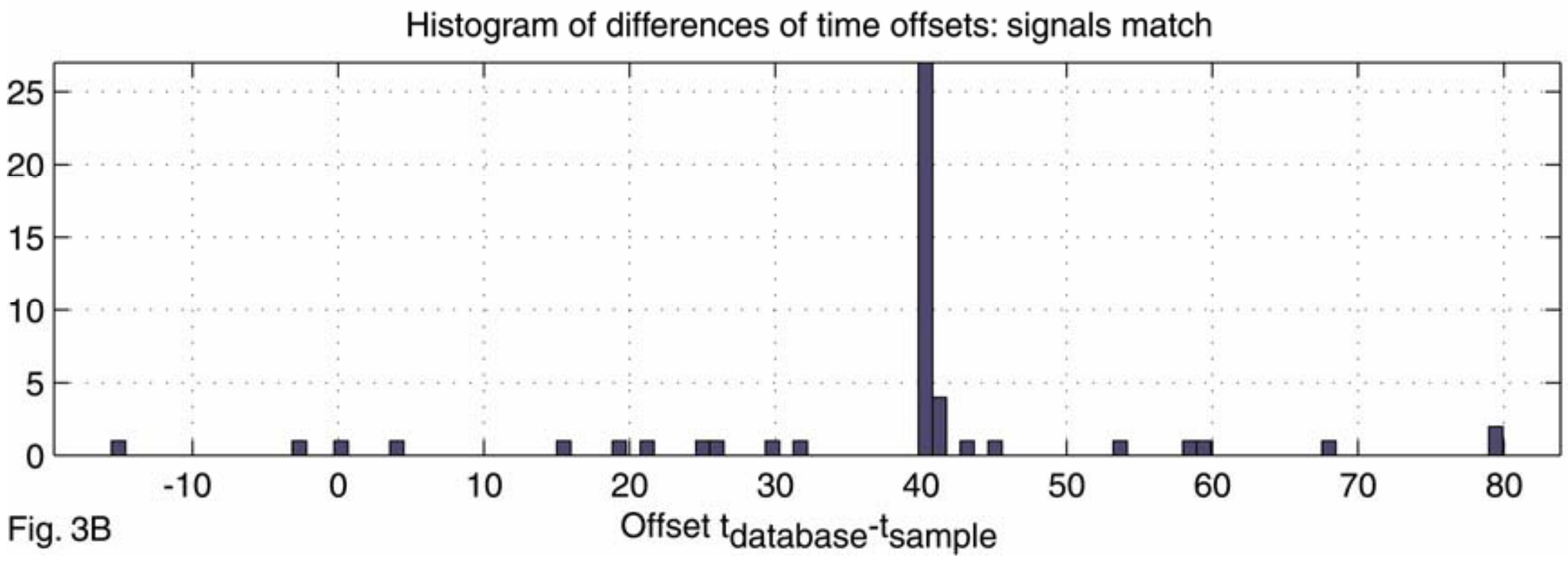

Histogram of differences of time offsets: signals match
25
20
15
10
5
0
-10
0
10
20
30
40
50
60
70
80
Offset $t_{database}$-t_{sample}

Fig. 3B

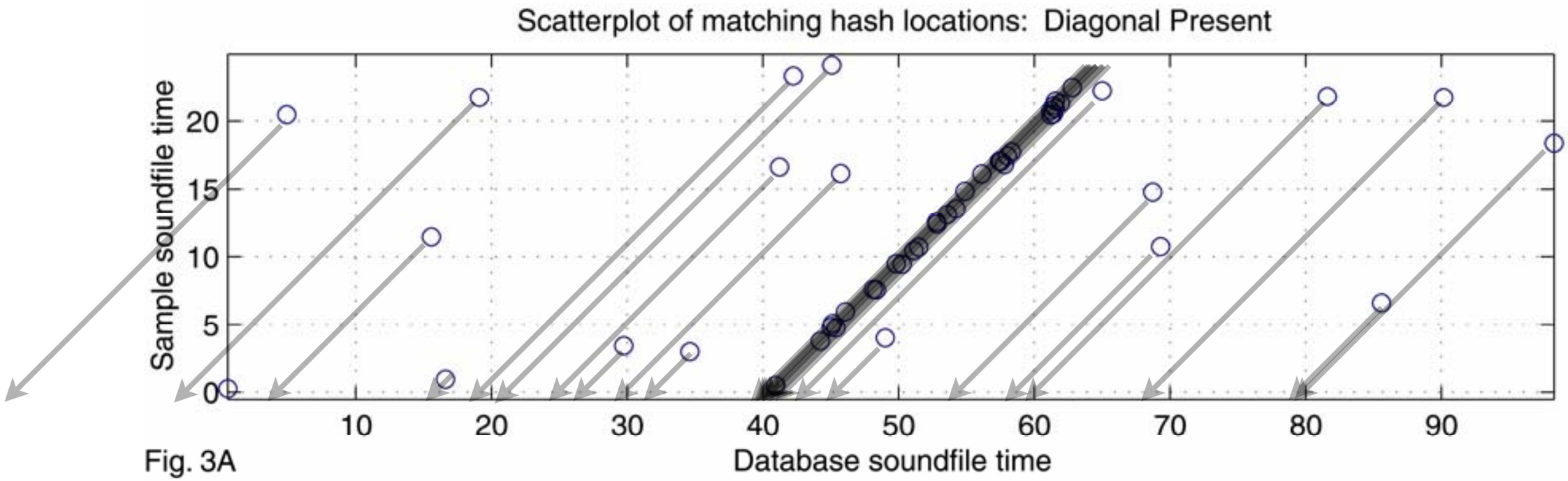

Scatterplot of matching hash locations: Diagonal Present
Sample soundfile time
0
5
10
15
20
10
20
30
40
50
60
70
80
90
Database soundfile time

Fig. 3A

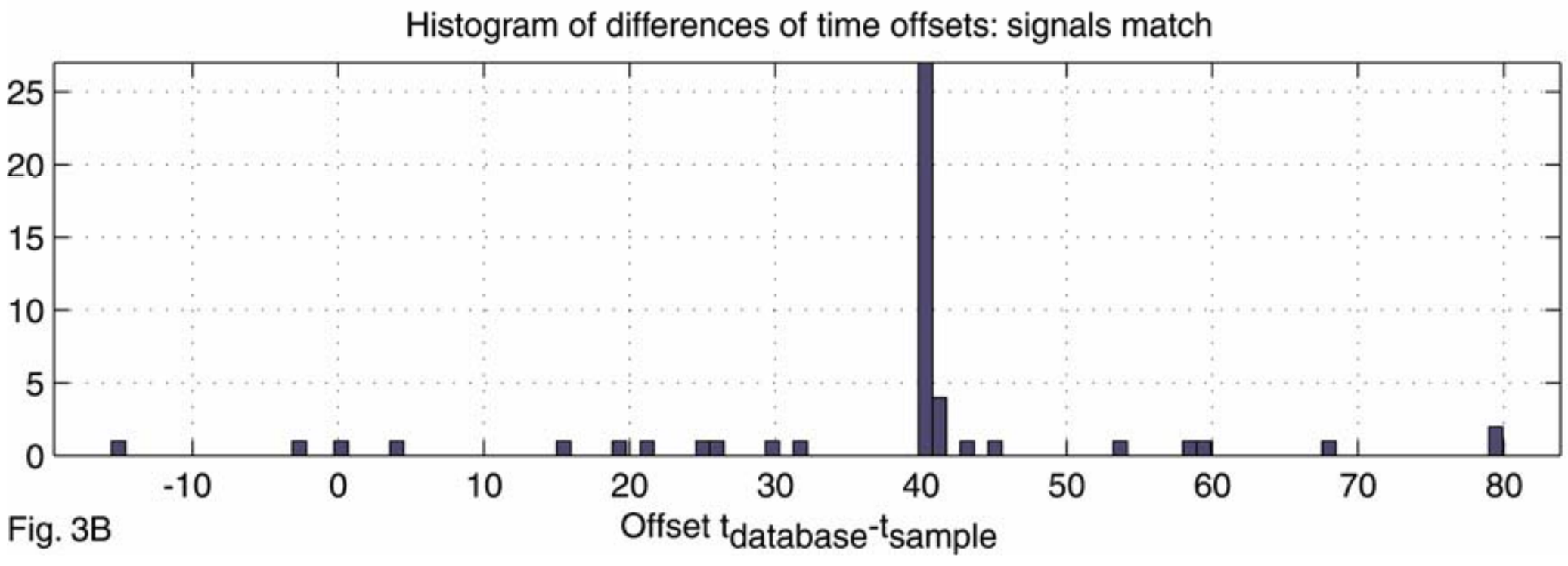

Histogram of differences of time offsets: signals match
25
20
15
10
5
0
-10
0
10
20
30
40
50
60
70
80
Offset t_database-t_sample

Fig. 3B

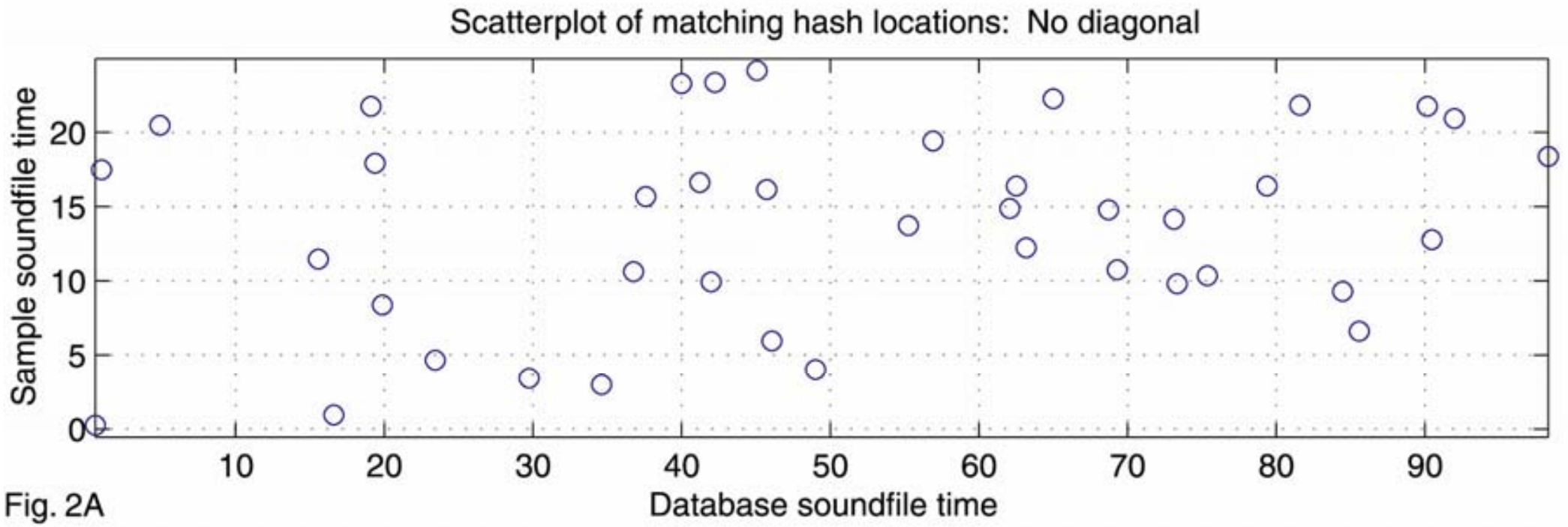

Scatterplot of matching hash locations: No diagonal
Sample soundfile time
0
5
10
15
20
10
20
30
40
50
60
70
80
90
Database soundfile time

Fig. 2A

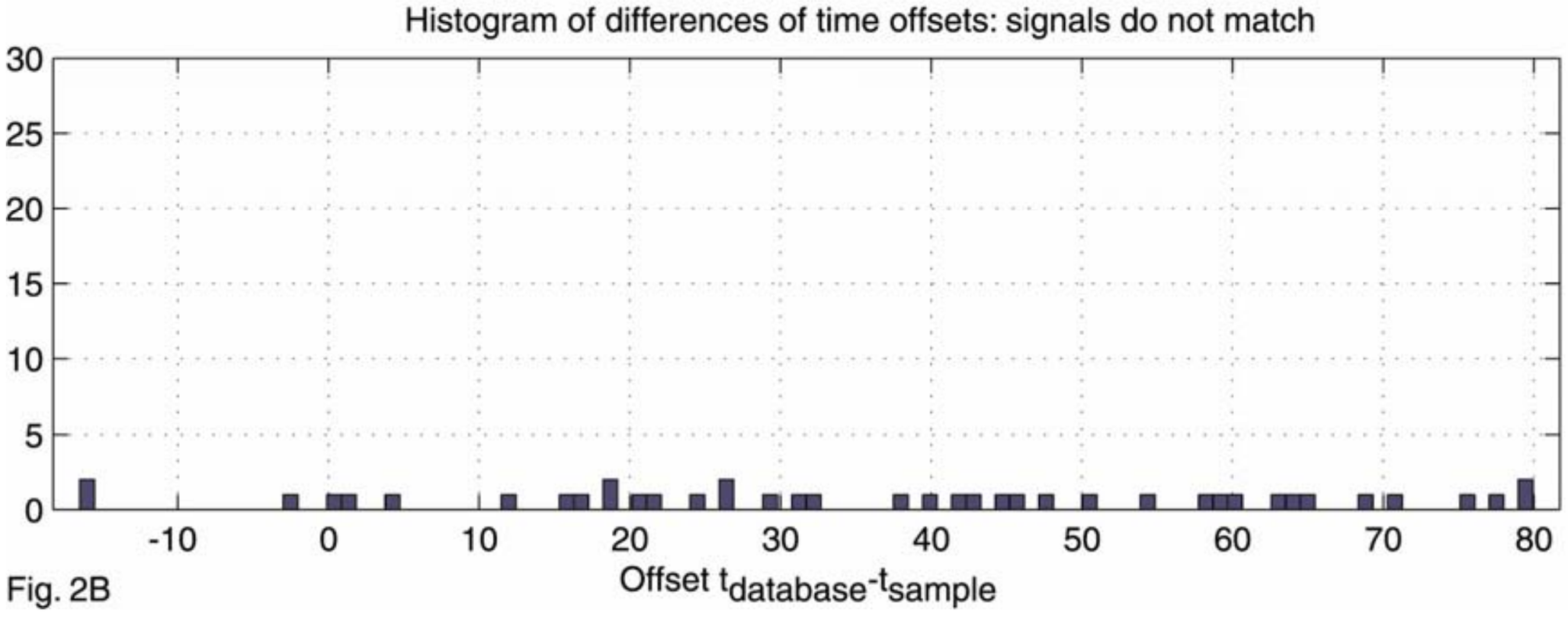
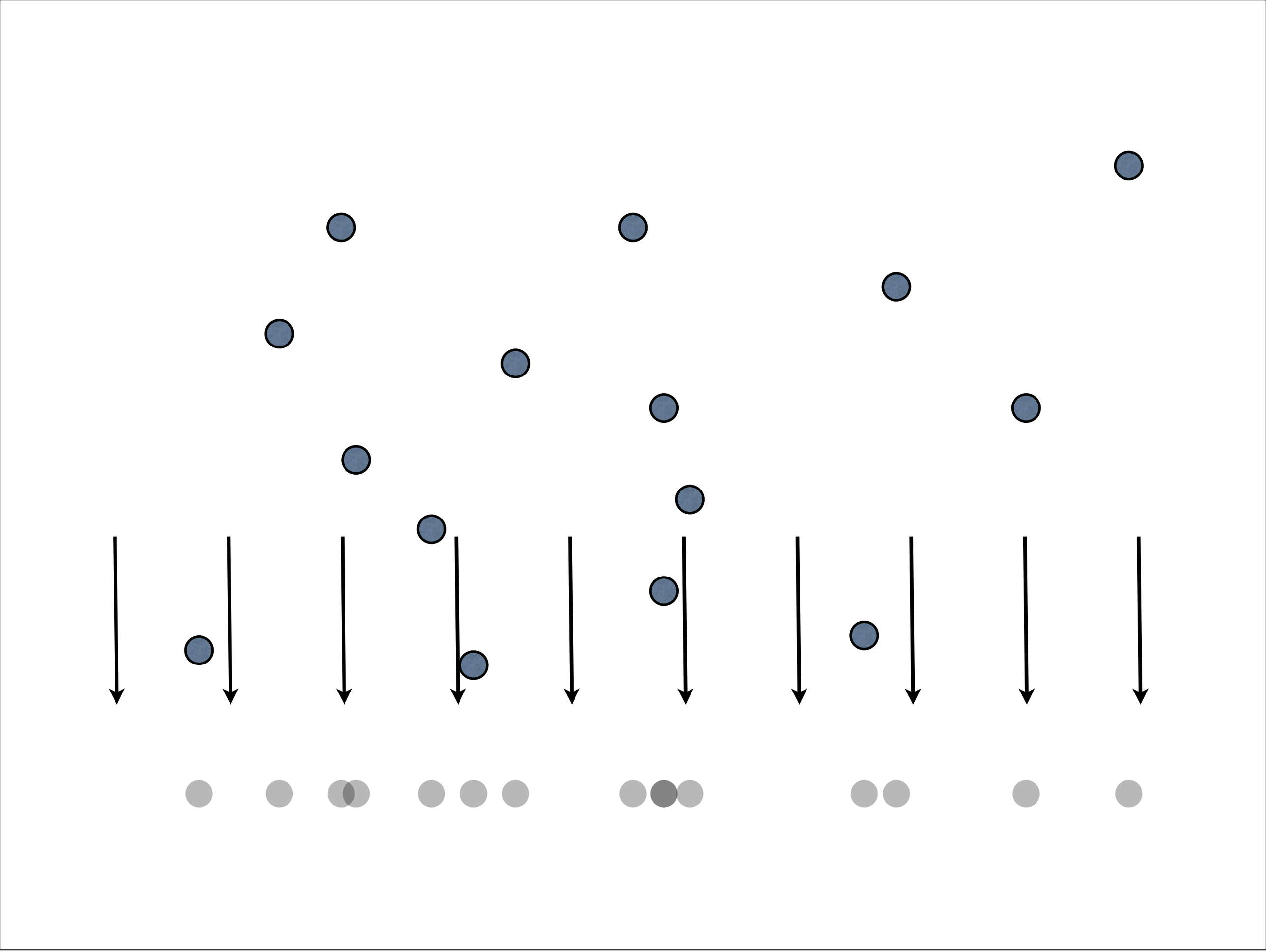
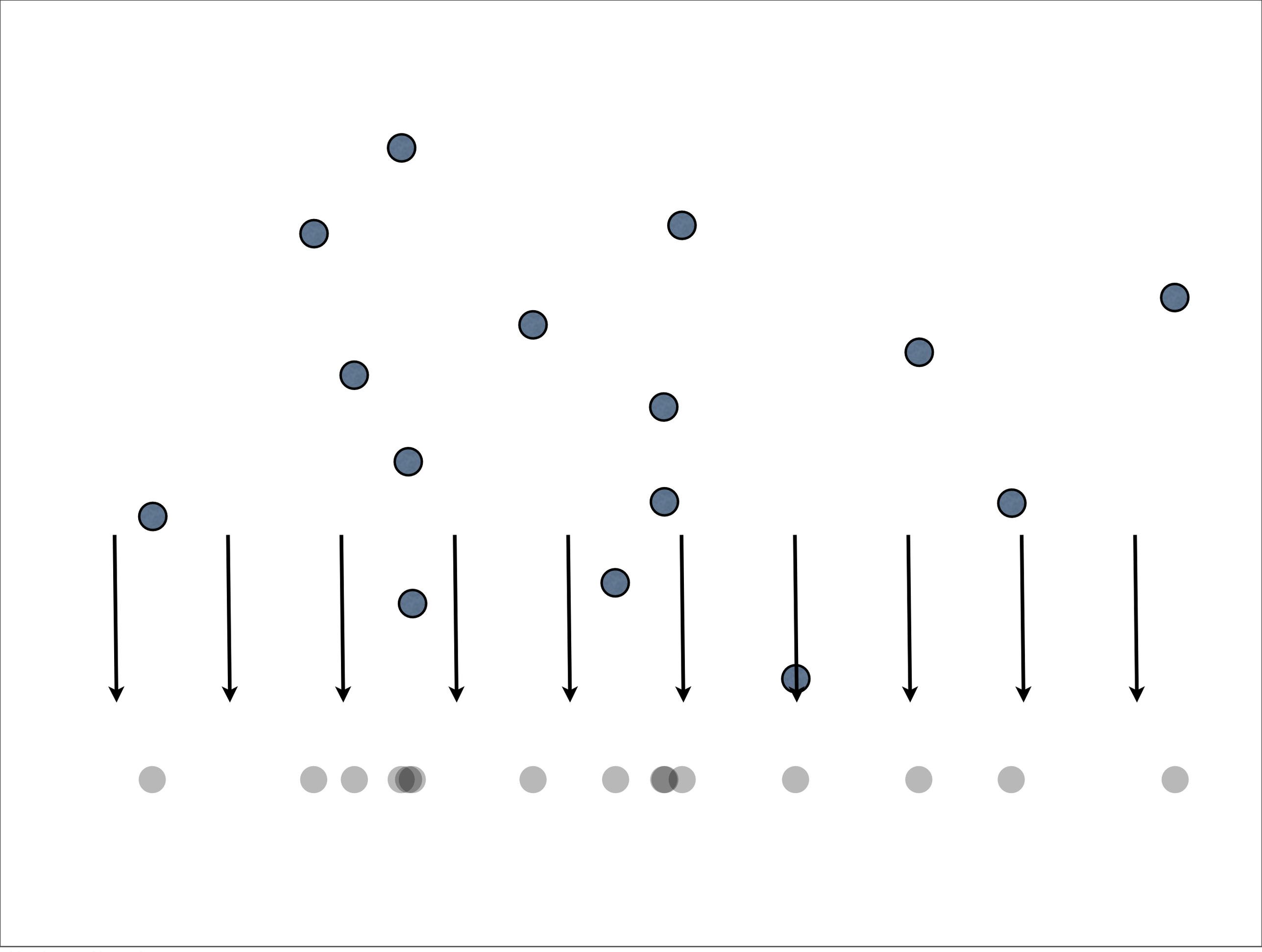

Histogram of differences of time offsets: signals do not match
30
25
20
15
10
5
0
-10
0
10
20
30
40
50
60
70
80
Offset t_database-t_sample

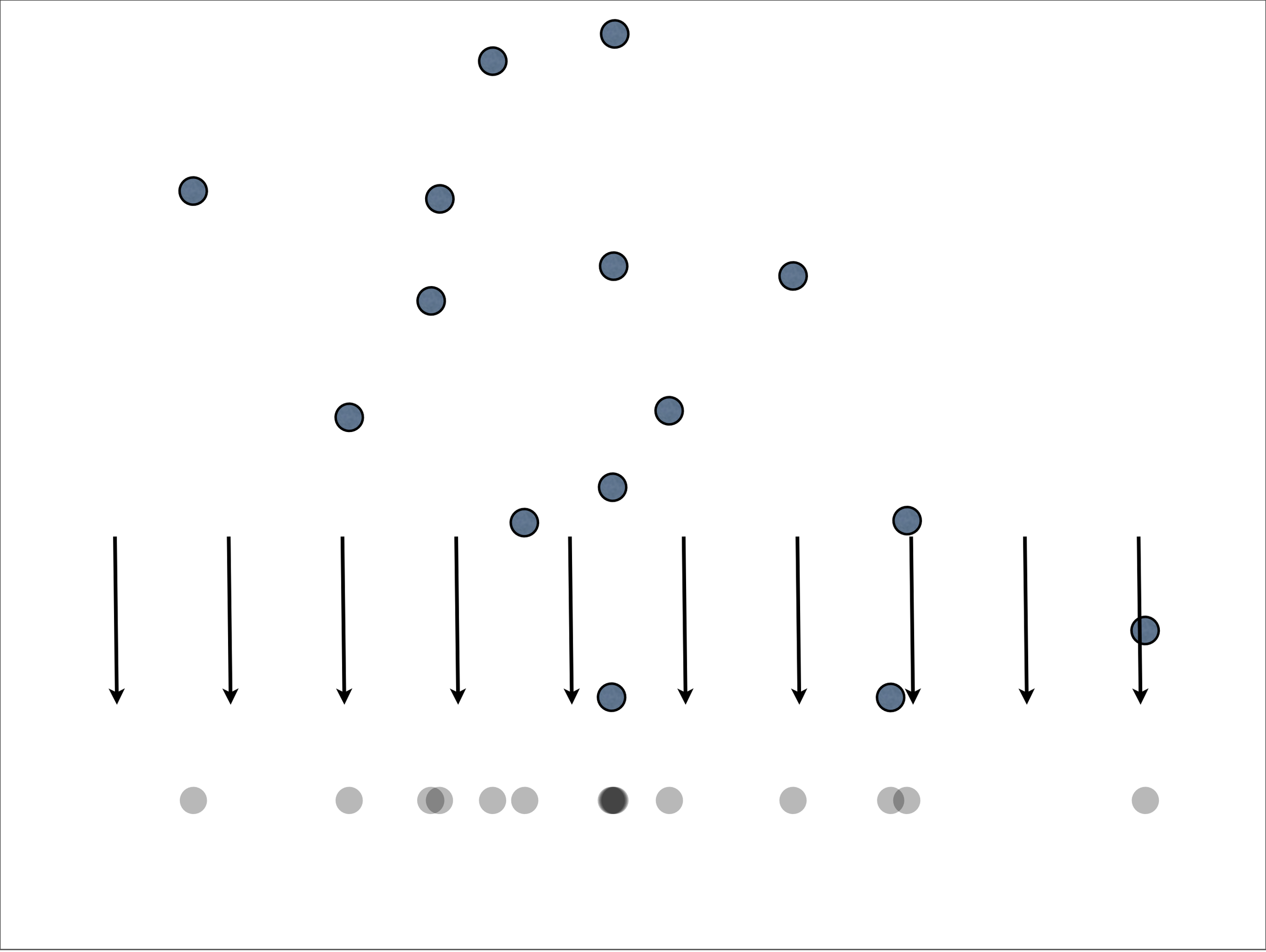
Fig. 2B

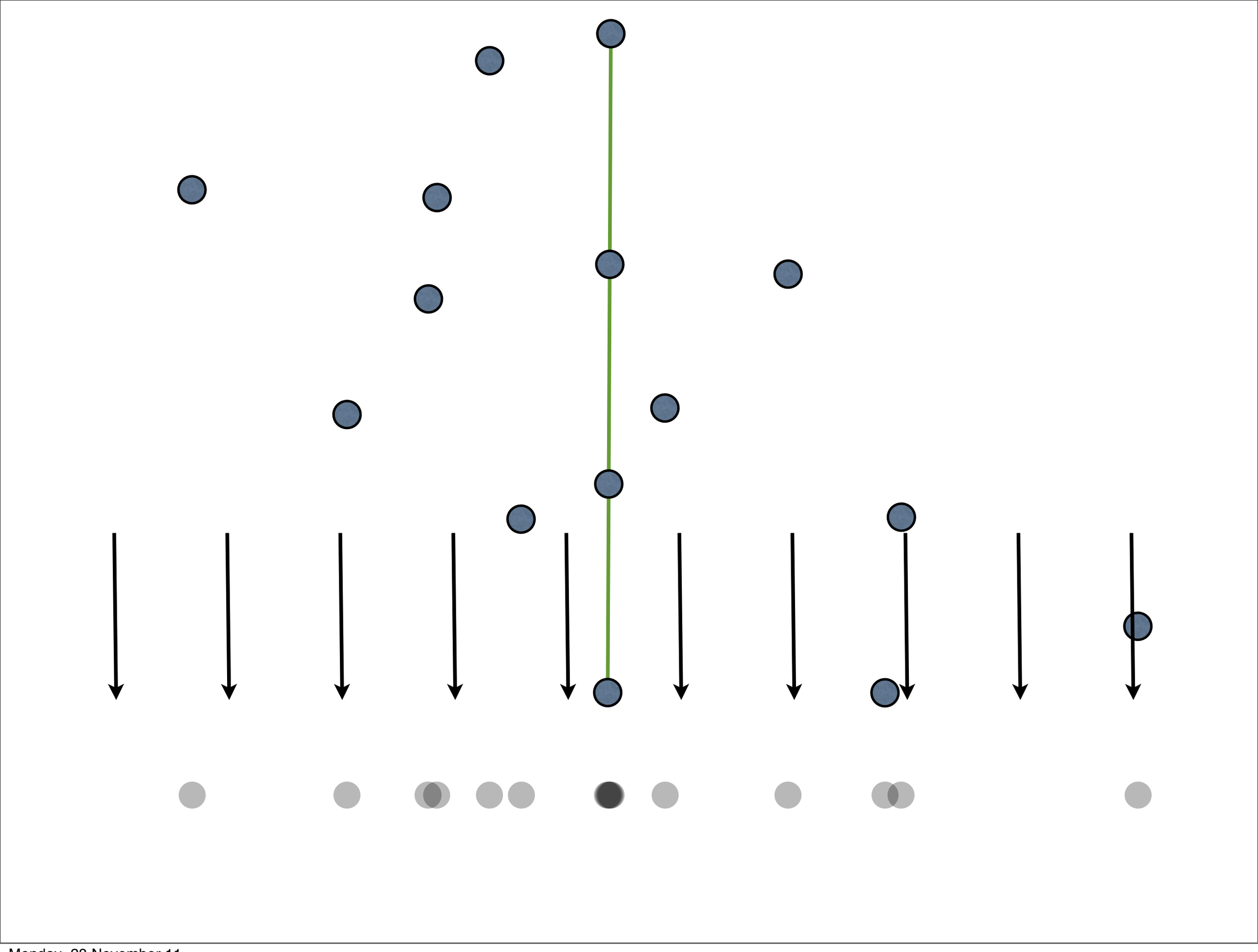
pictures

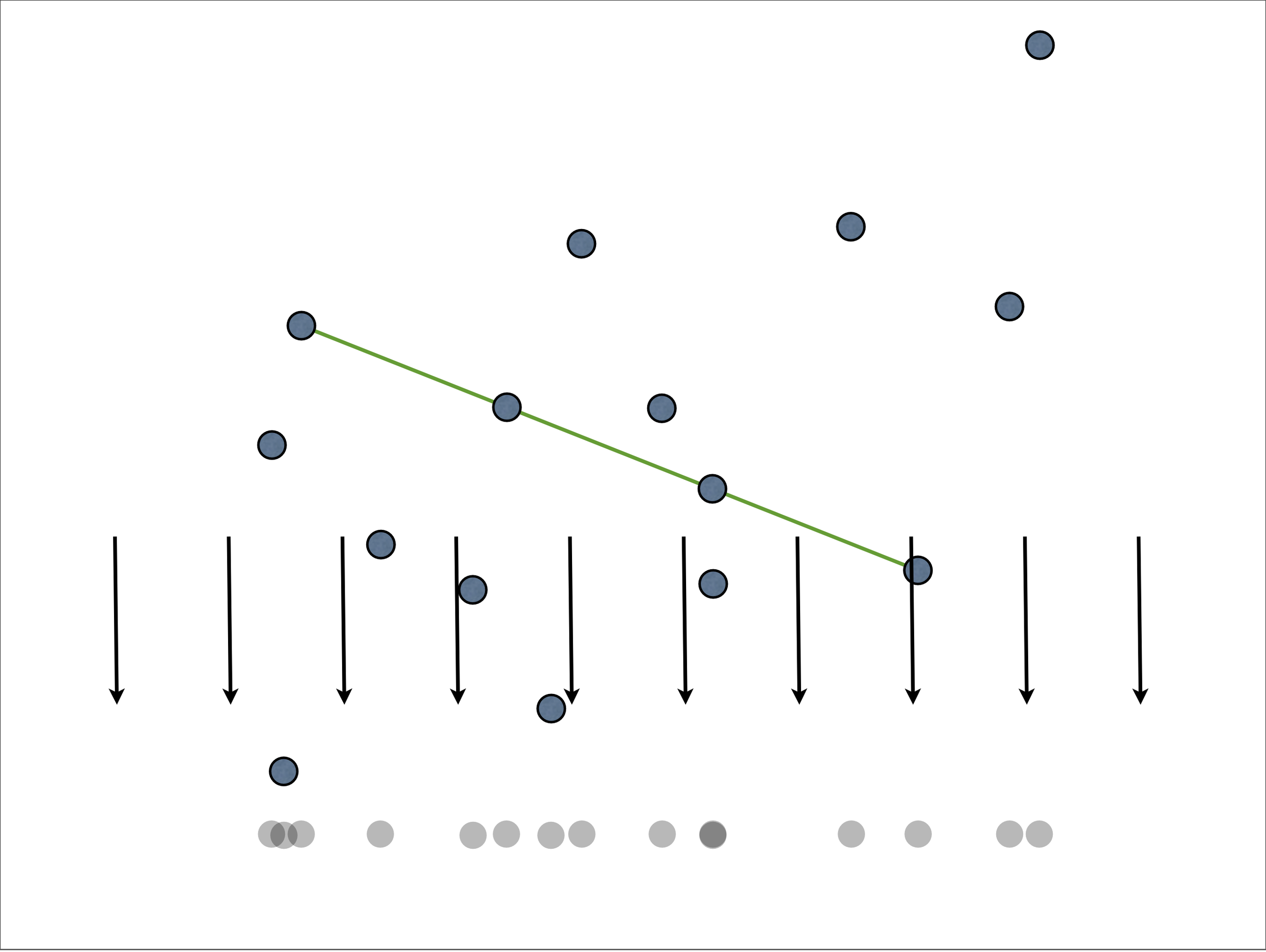
finding lines in images

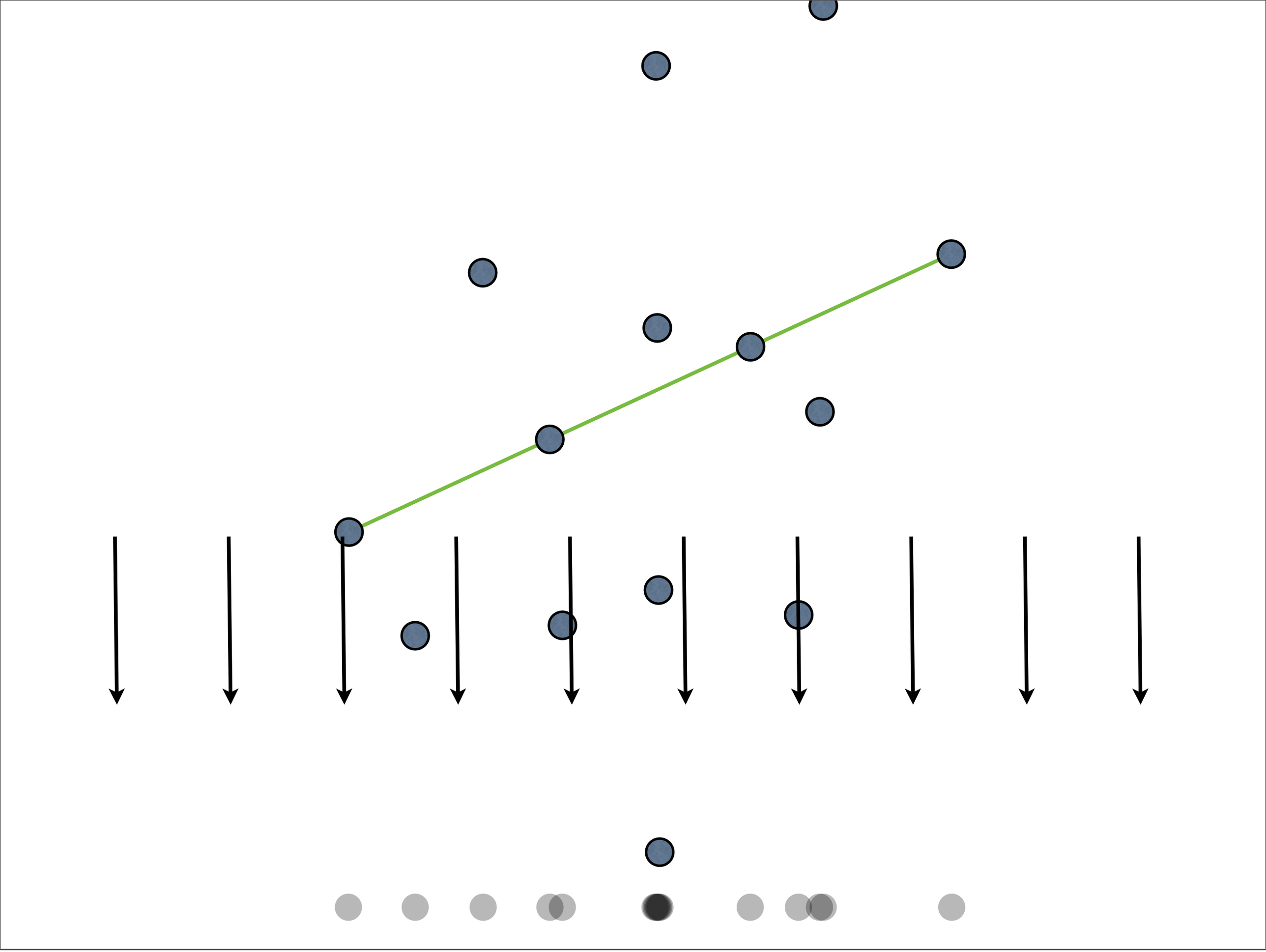


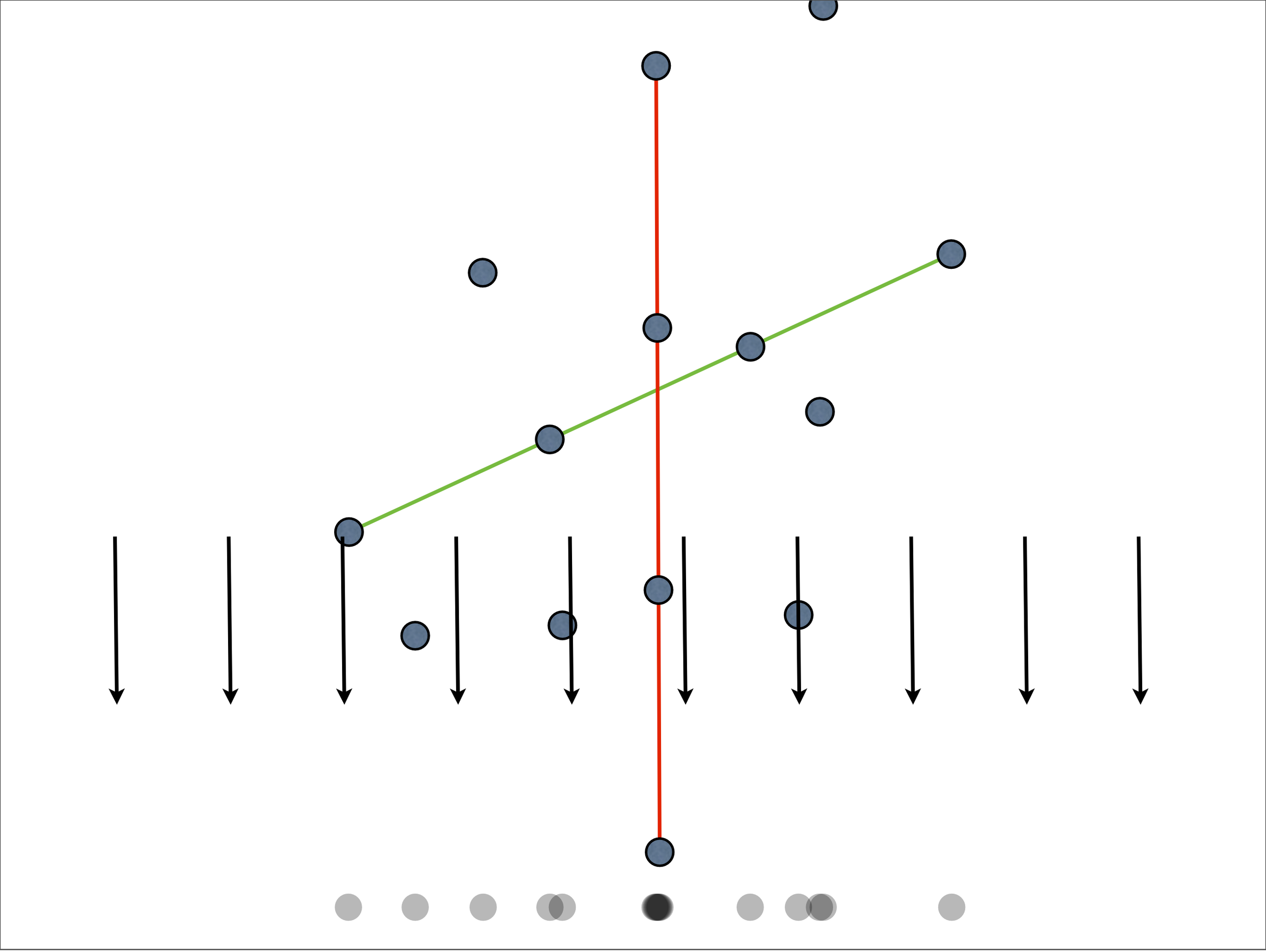


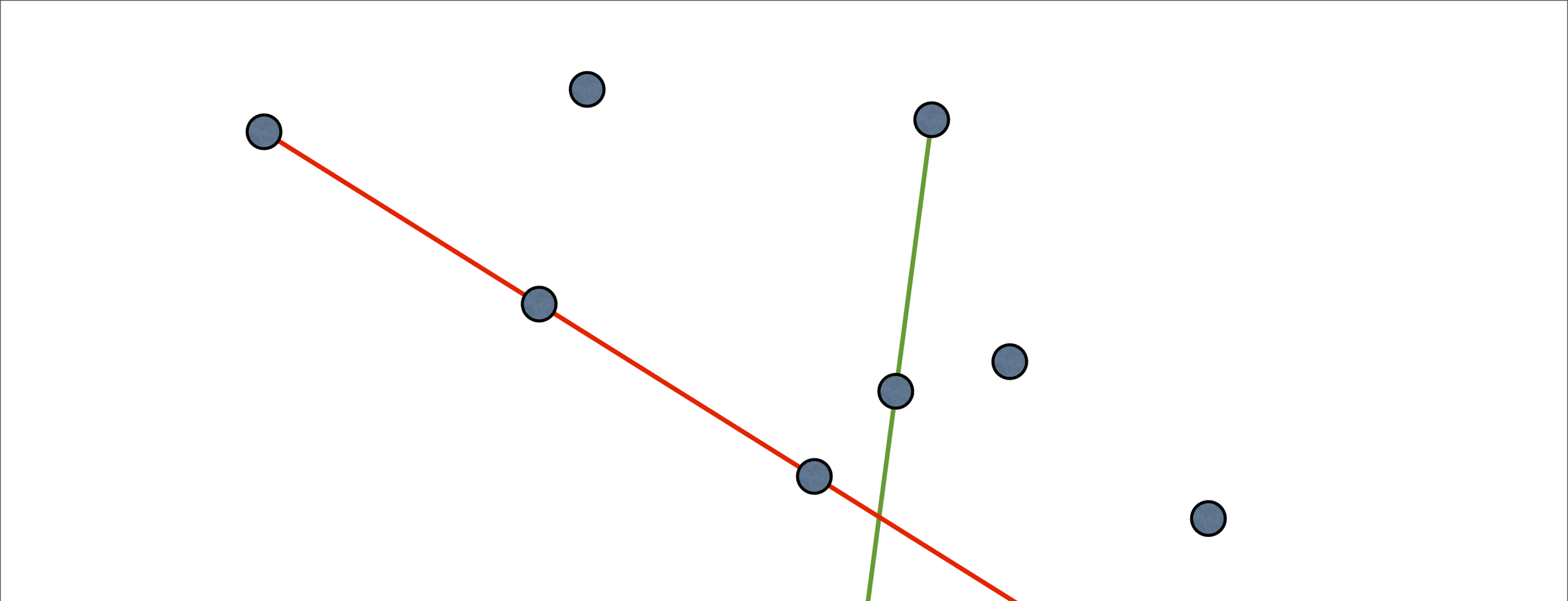


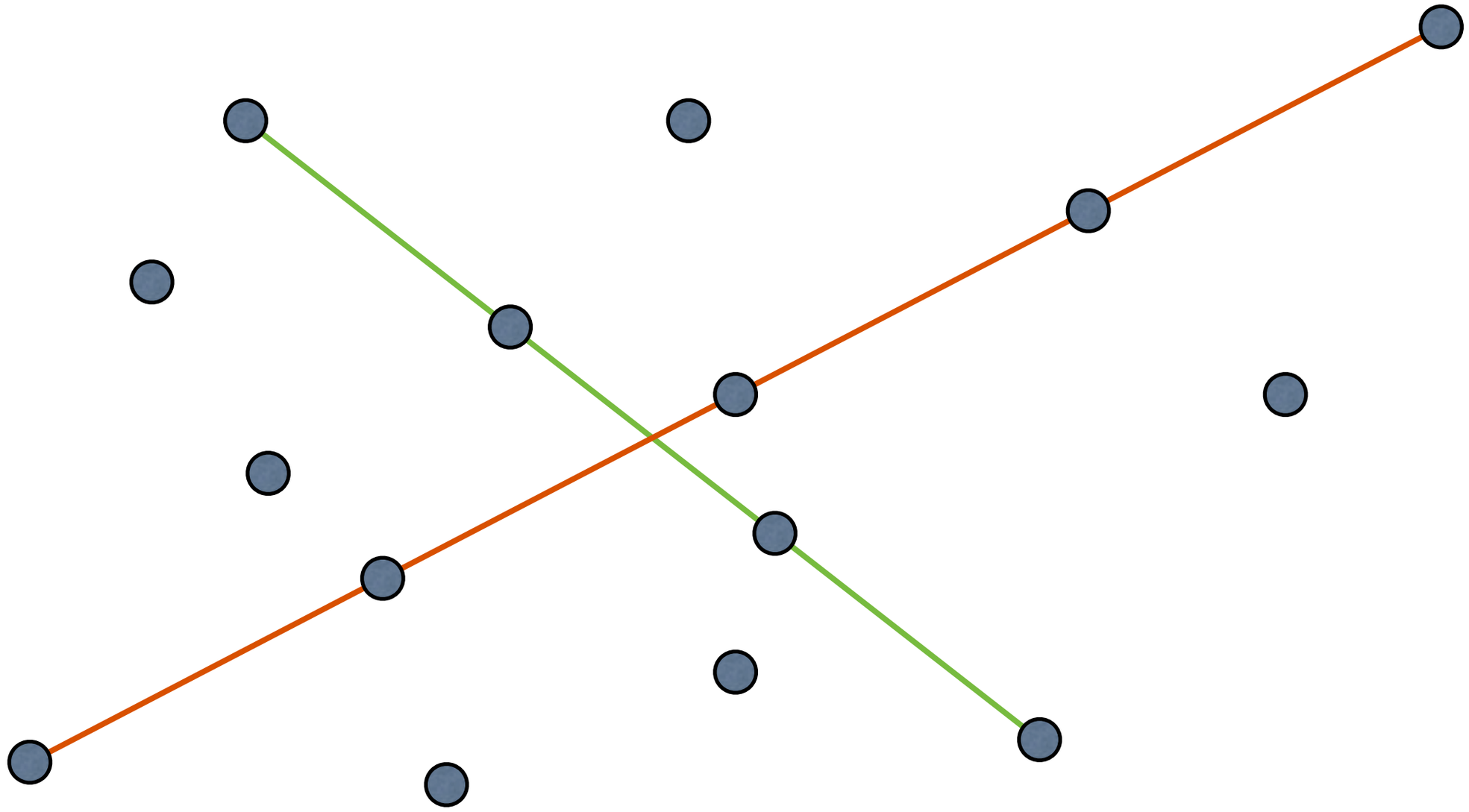




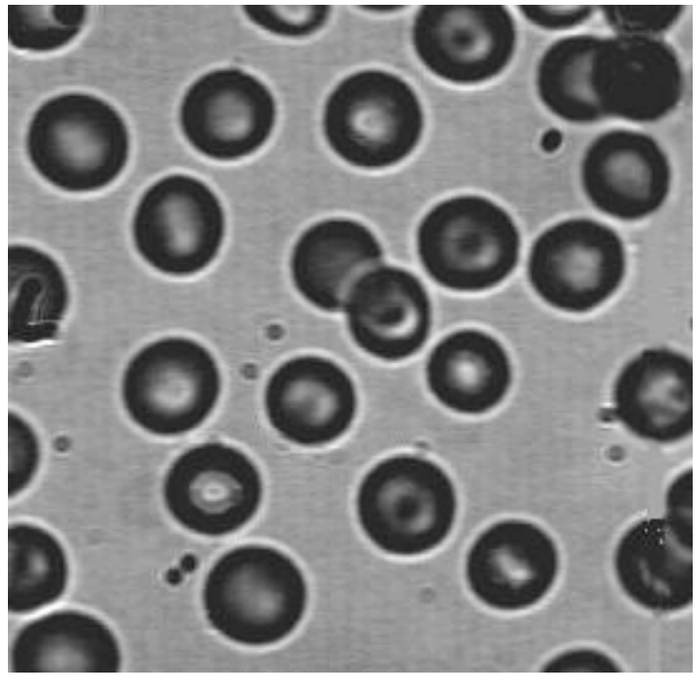




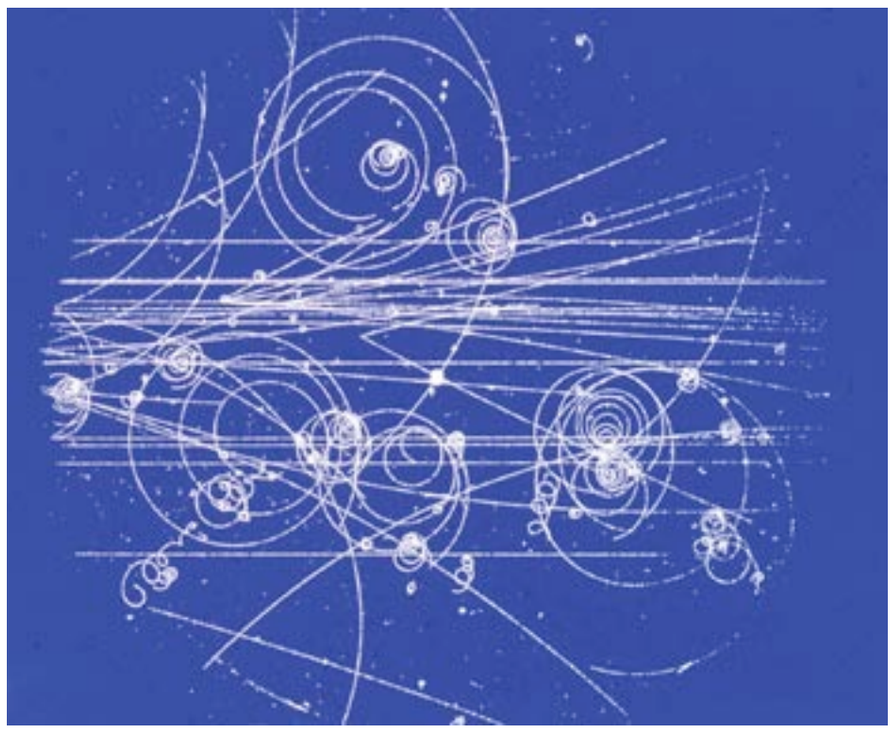




finding blood cells in images



finding sub-atomic particles in images



needles in haystacks

- Find features and create an index pointing to the *documents* that have the relevant features.