ECE 598HH: Advanced Wireless Networks and Sensing Systems

Lecture 18: Internet of Acoustic Things
Haitham Hassanieh

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CSL: COORDINATED SCIENCE LAB

*These slides are courtesy of Nirupam Roy (UIUC)


50 million voice assistants are sold in US
amazon

Google
amazon
amazon
nura
BOSE
QUALCOMM
The Future of Personal Audio
Google
Symphony 29...
Cello Suite no...
Fantasia and Fu...
Flute Sonata in...
Partita no. 5,...
Johnson & Johnson
BioStampRC
AUDIO SPOT LIGHT
QUALCOMM
PHILIPS
NSF
HEARABLES
CHALLENGE

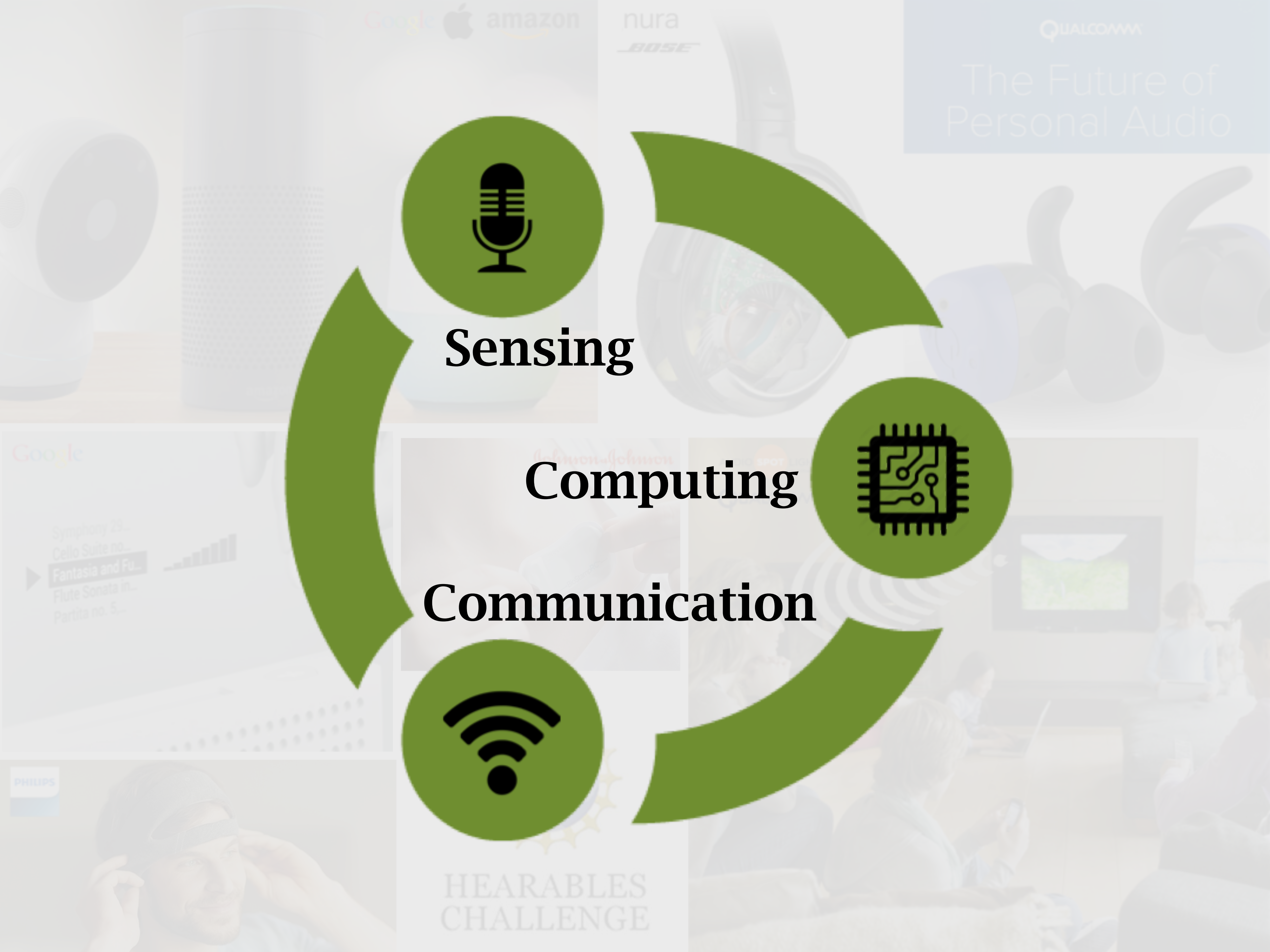

Google
amazon
nura
BOSE
QUALCOMM
The Future of Personal Audio
Google
Symphony 29...
Cello Suite no...
Fantasia and Fu...
Flute Sonata in...
Partita no. 5,...
Johnson & Johnson
AUDIO SPOTLIGHT
QUALCOMM
PHILIPS
NSF
HEARABLES CHALLENGE


Google
amazon
nura
BOSE
QUALCOMM
The Future of Personal Audio
Google
Symphony 29...
Cello Suite no...
Fantasia and Fu...
Flute Sonata in...
Partita no. 5,...
Gesture
Johnson&Johnson
AUDIO SPOT LIGHT
QUALCOMM
PHILIPS
NSF
HEARABLES CHALLENGE

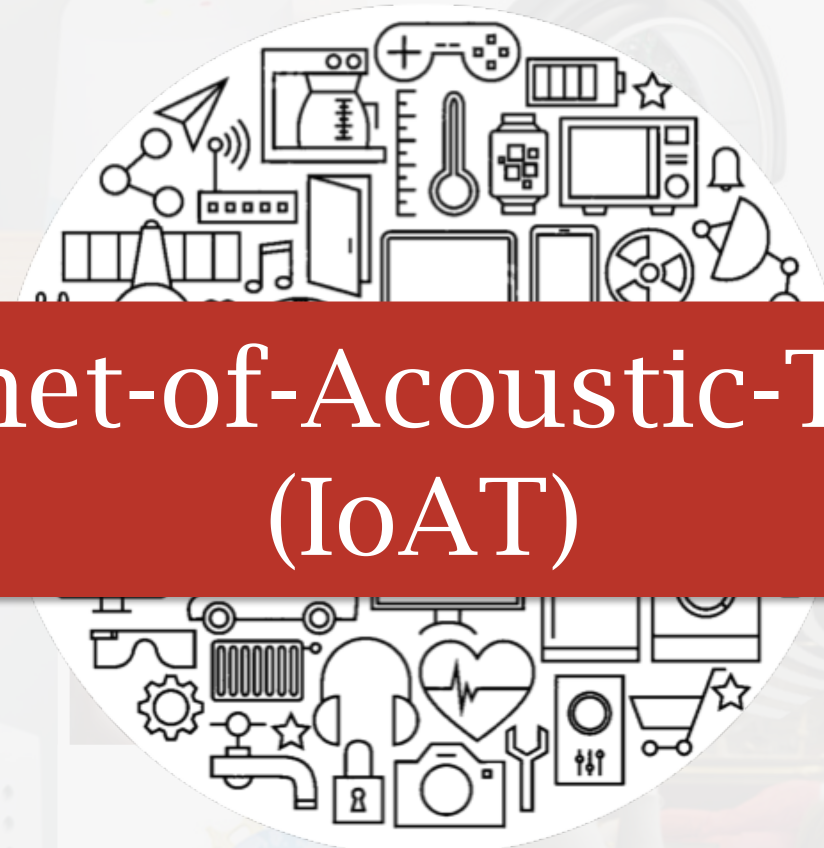

Google
amazon
nura
BOSE
QUALCOMM
The Future of Personal Audio
Google
Symphony 29...
Cello Suite no...
Fantasia and Fu...
Flute Sonata in...
Partita no. 5,...
Johnson & Johnson
Body vibration
AUDIO SPOT LIGHT
QUALCOMM
PHILIPS
NSF
HEARABLES CHALLENGE


Google
nura
BOSE
QUALCOMM
The Future of Personal Audio
Google
Symphony 29...
Cello Suite no...
Fantasia and Fu...
Flute Sonata in...
Partita no. 5,...
PHILIPS
NSF
HEARABLES CHALLENGE
AUDIO SPOT LIGHT
QUALCOMM
Holosonics: Pencil beam sound

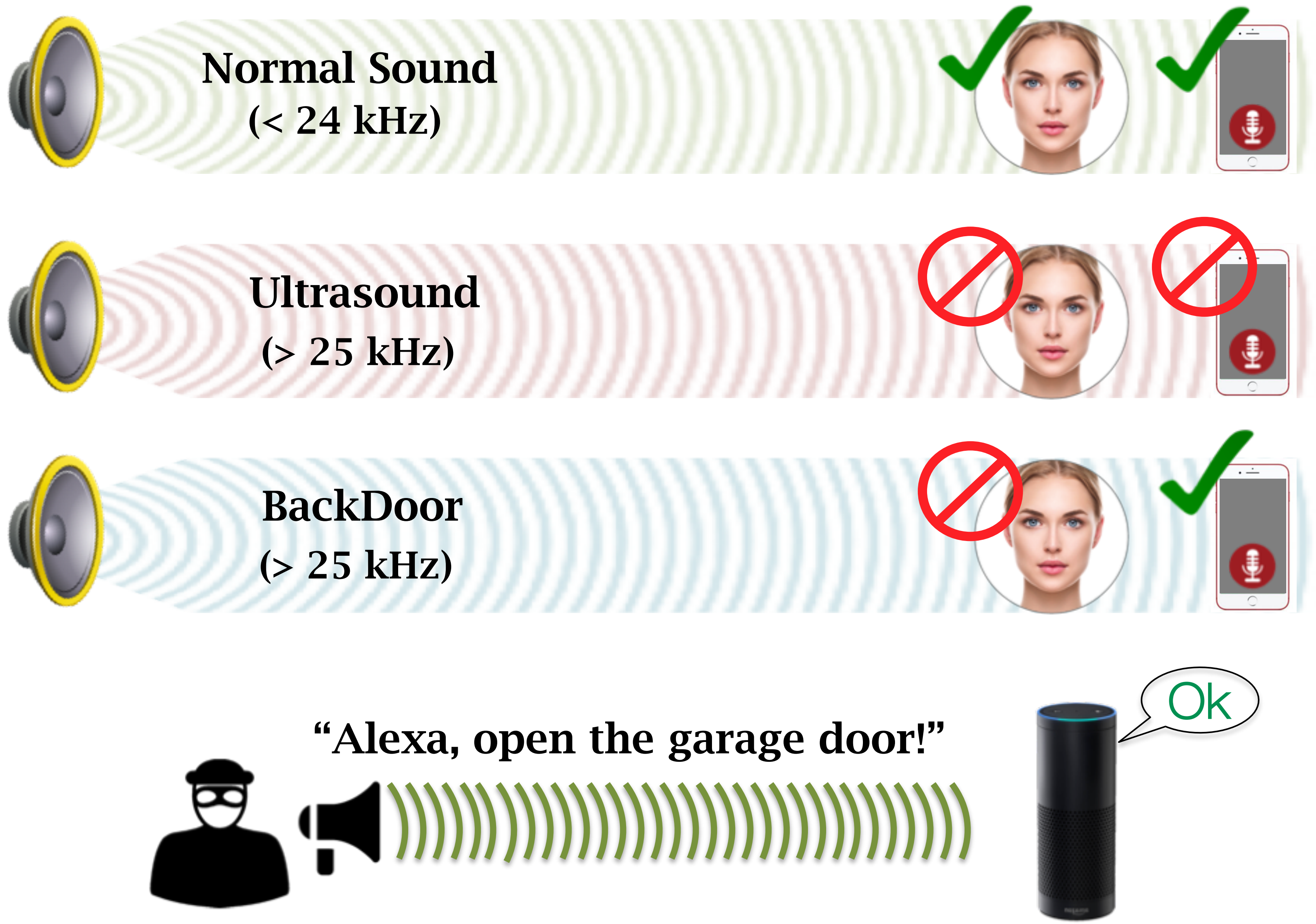

Google
amazon
nura
BOSE
QUALCOMM
The Future of Personal Audio
Google
Symphony 29...
Cello Suite no...
Fantasia and Fu...
Flute Sonata in...
Partita no. 5,...
Johnson & Johnson
AUDIO SPOT LIGHT
QUALCOMM
PHILIPS
NSF
HEARABLES CHALLENGE


The Future of Personal Audio
Sensing
Computing
Communication
HEARABLES CHALLENGE

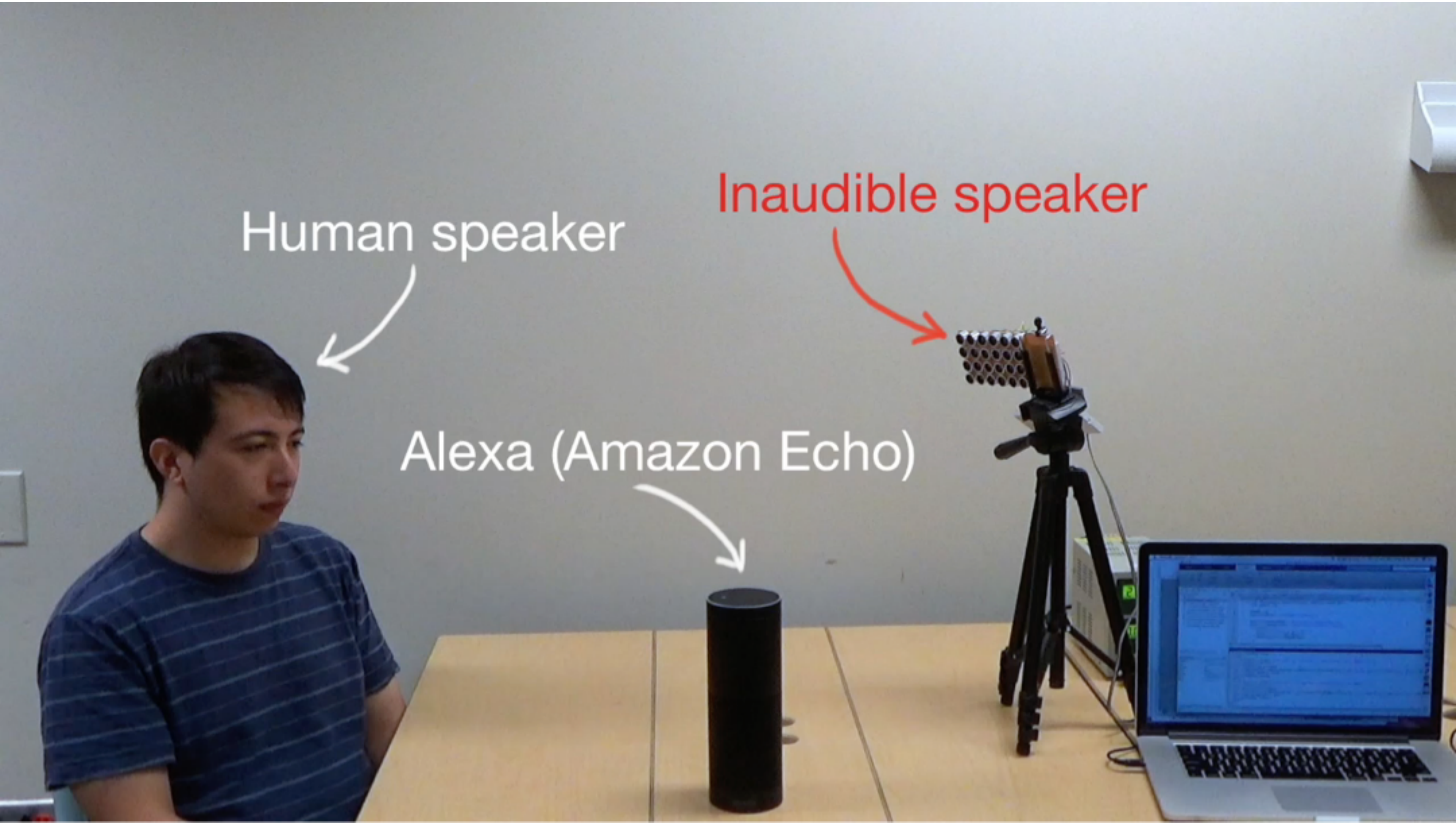
Internet-of-Acoustic-Things (IoAT)

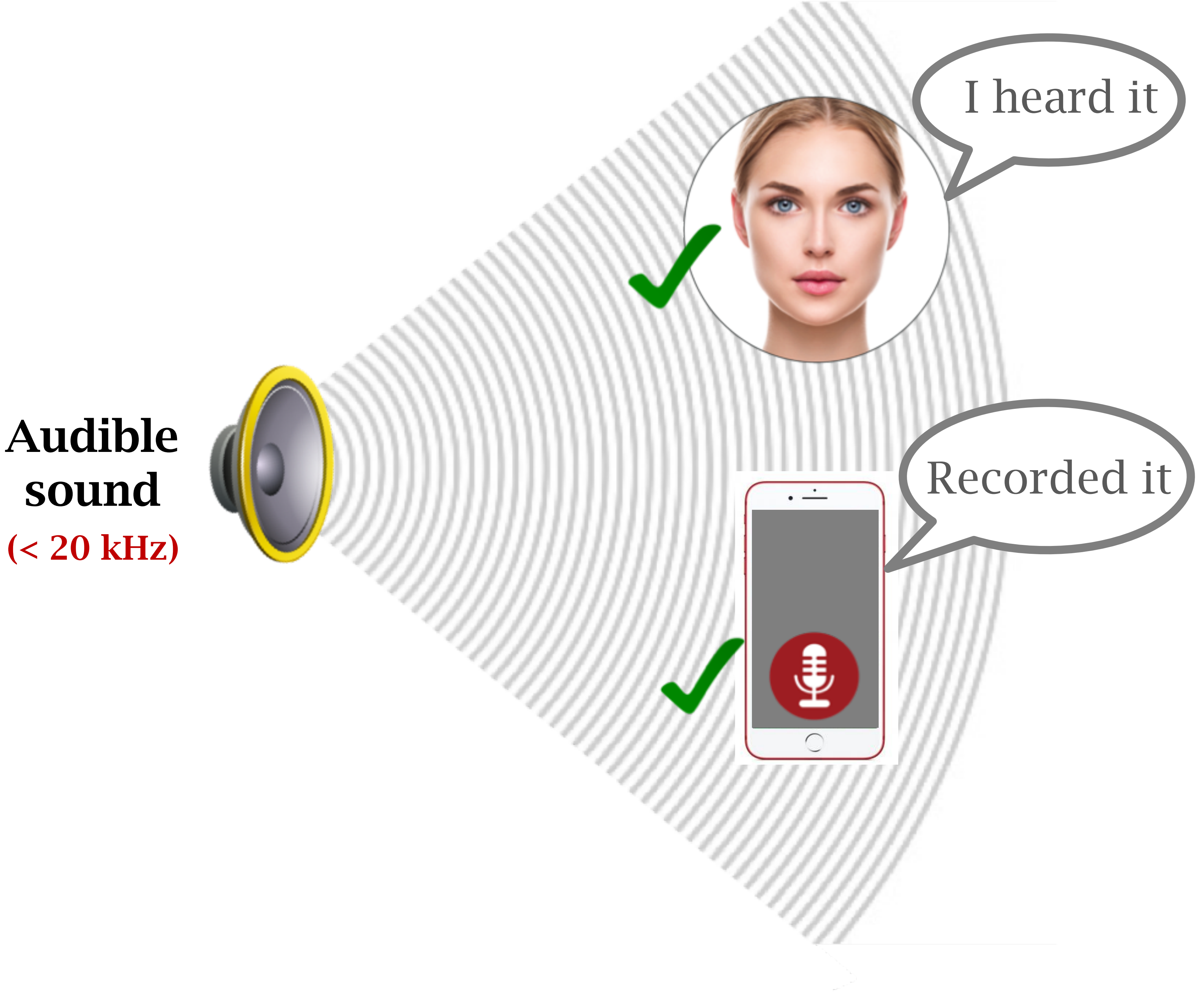


Inaudible Acoustics

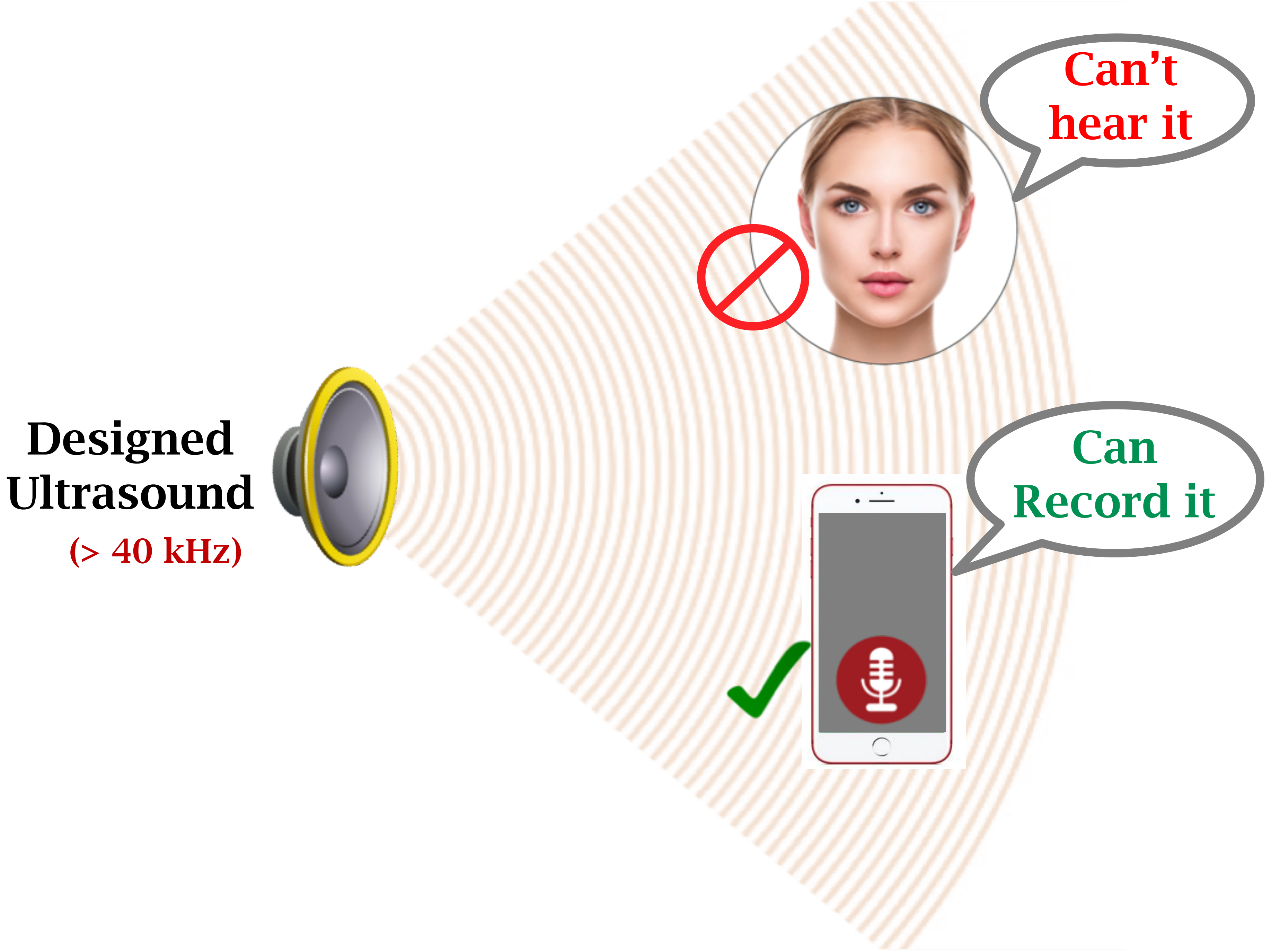


Inaudible Acoustics

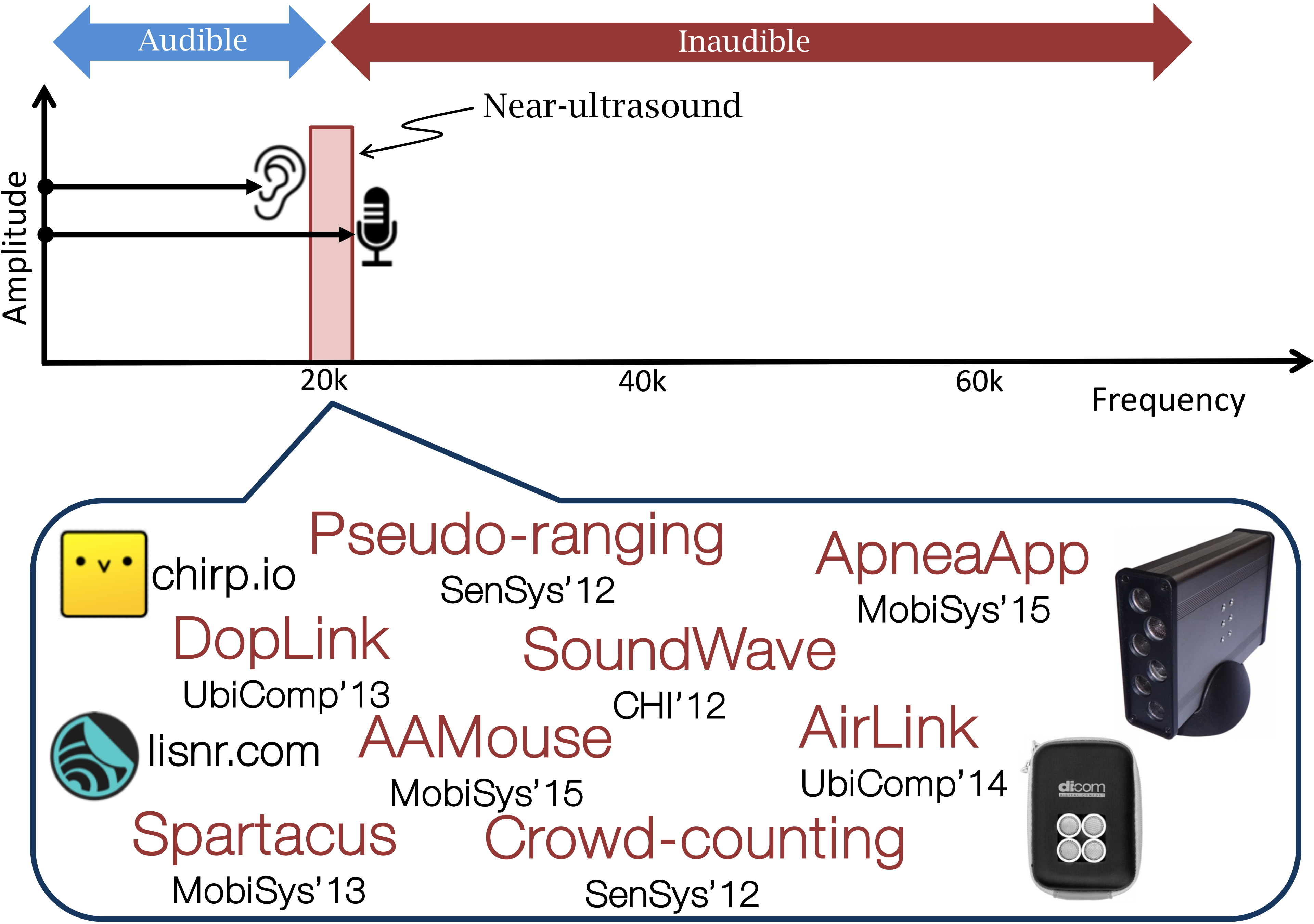


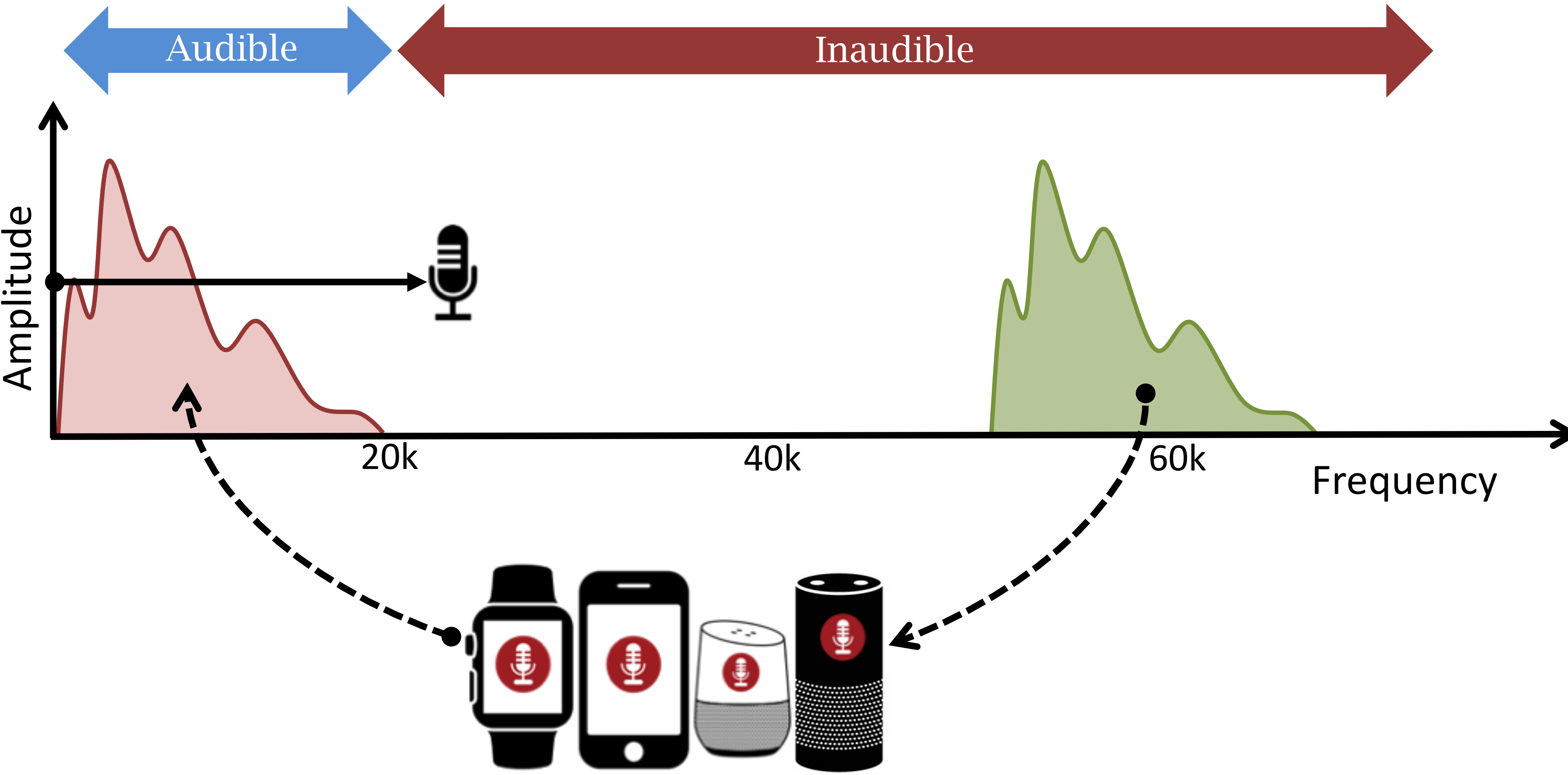


Audible sound
(< 20 kHz)
I heard it
Recorded it



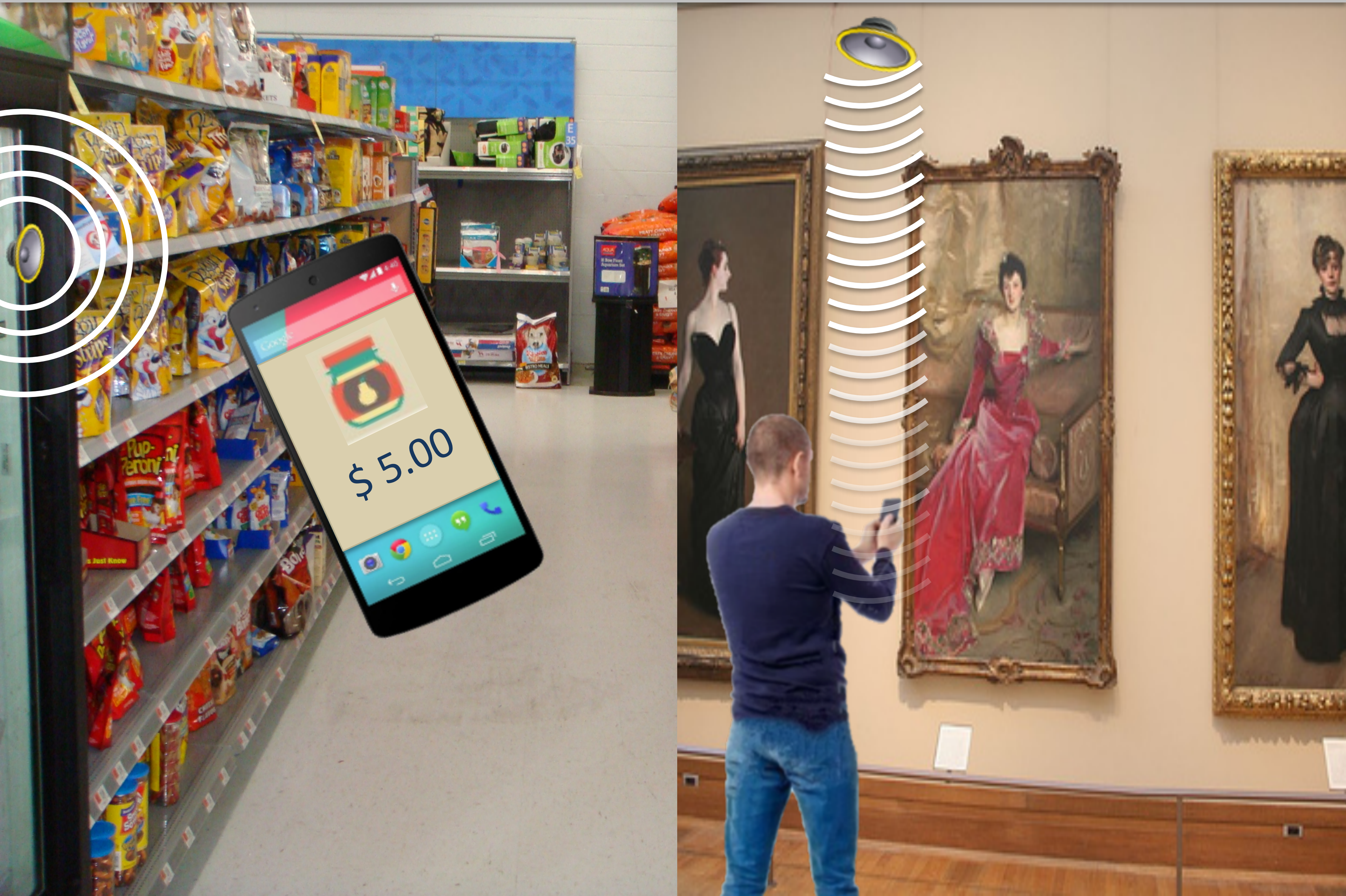
Designed
Ultrasound
(> 40 kHz)
Can't
hear it
Can
Record it


Audible
Inaudible
Near-ultrasound
Amplitude
20k
40k
60k
Frequency
chirp.io
Pseudo-ranging
SenSys'12
ApneaApp
MobiSys'15
DopLink
UbiComp'13
SoundWave
CHI'12
lisnr.com
AAMouse
MobiSys'15
AirLink
UbiComp'14
Spartacus
MobiSys'13
Crowd-counting
SenSys'12
dicom



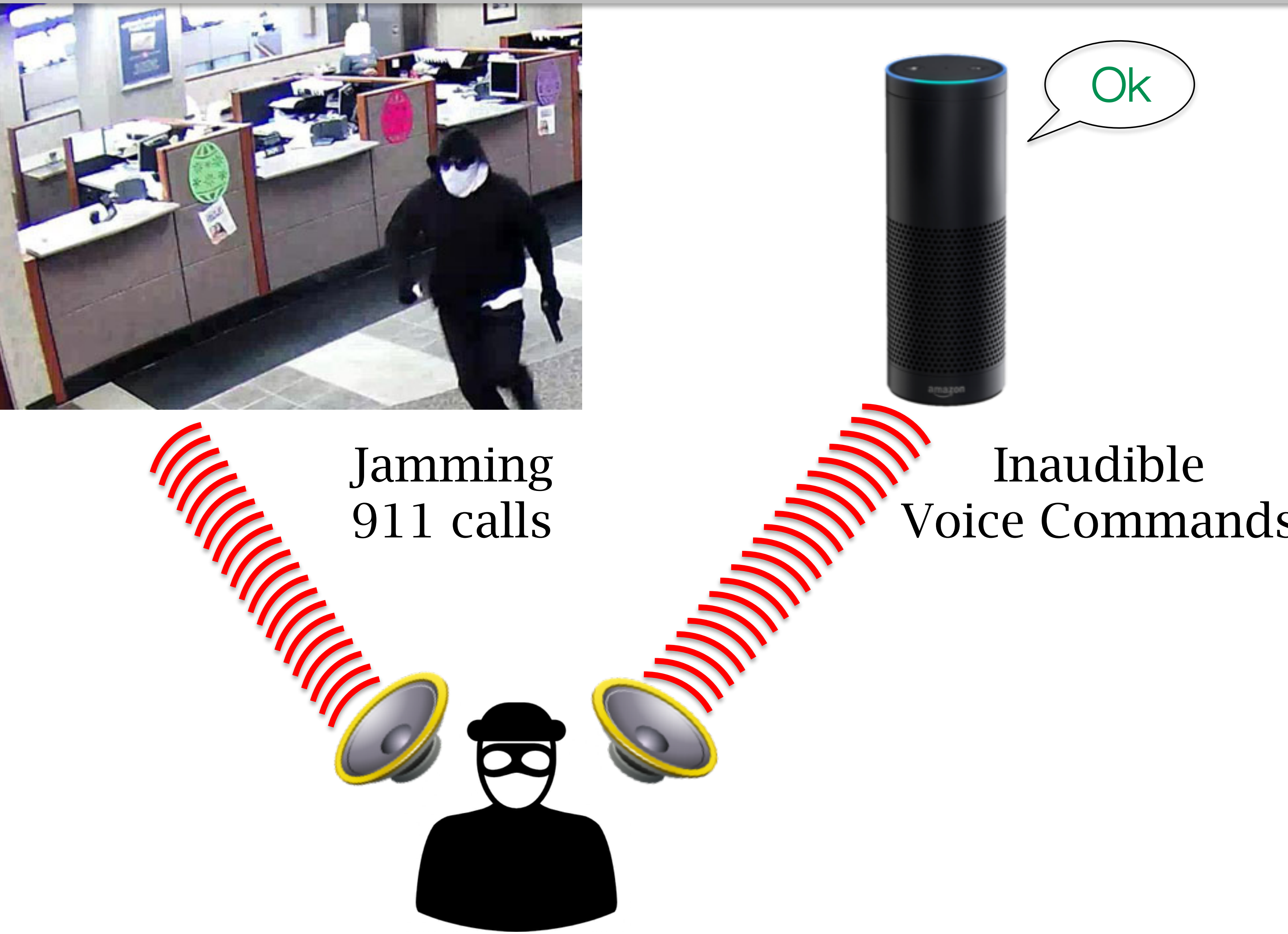
Audible
Inaudible
Amplitude
20k
40k
60k
Frequency

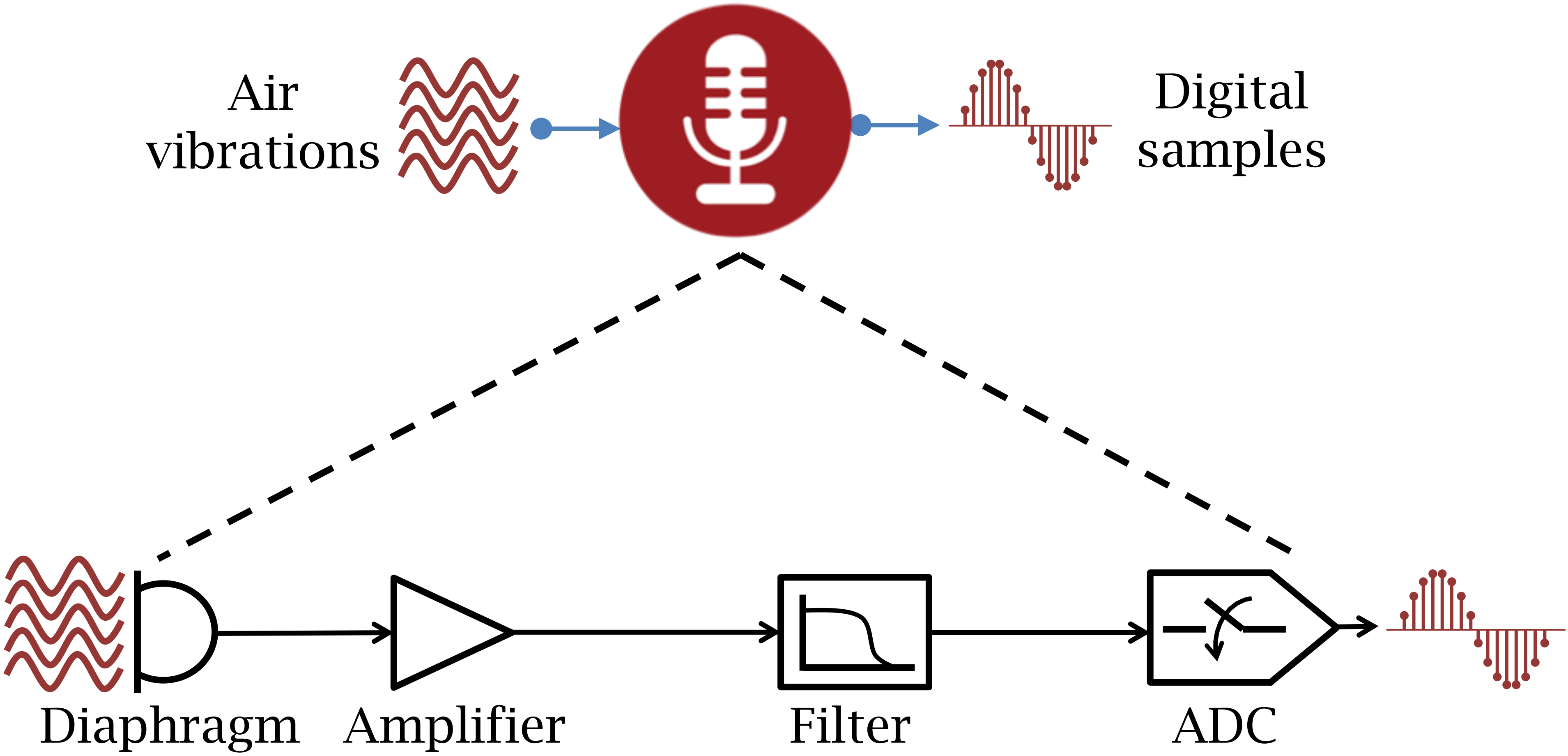
Inaudible Acoustic Communication

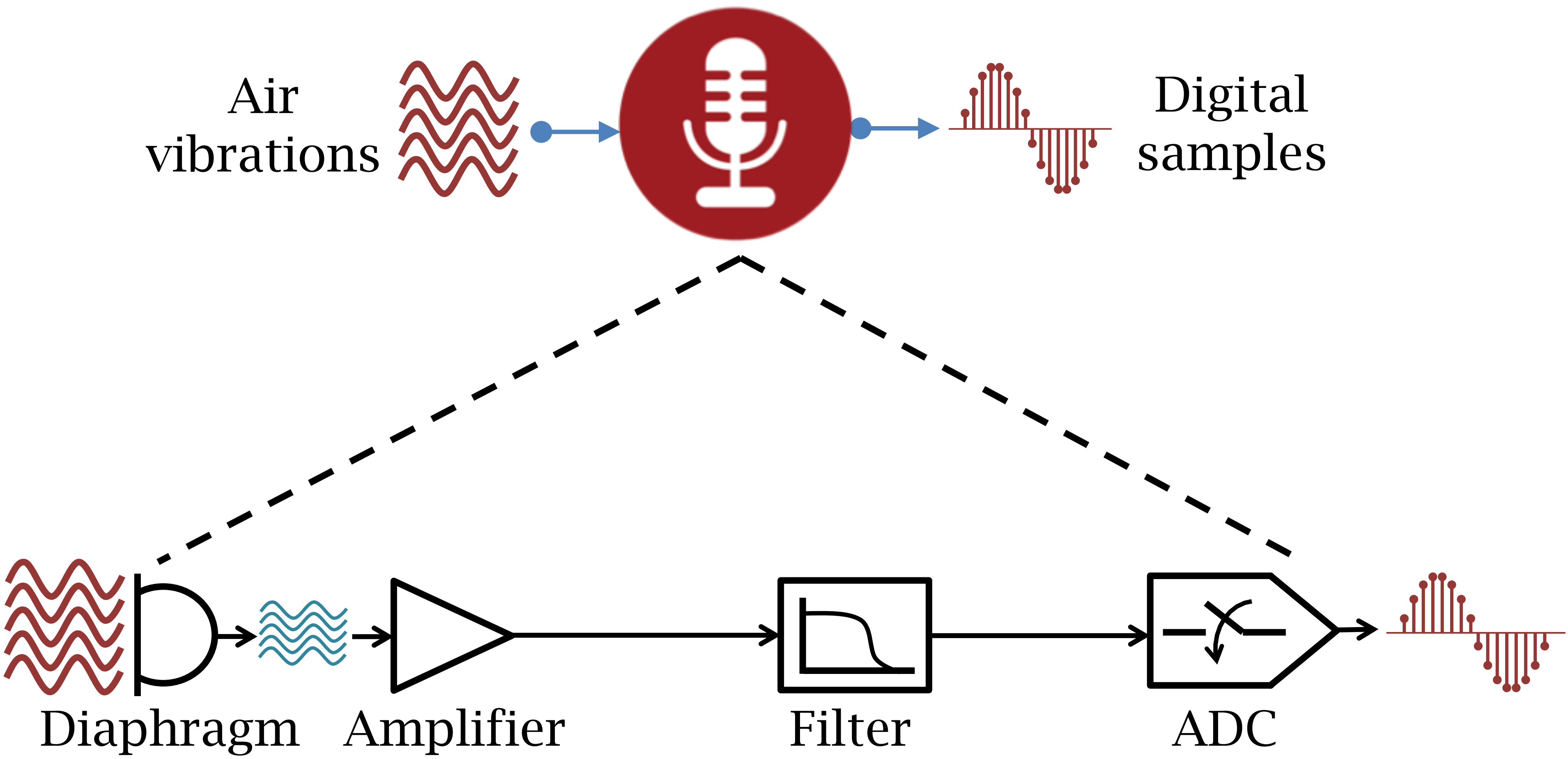


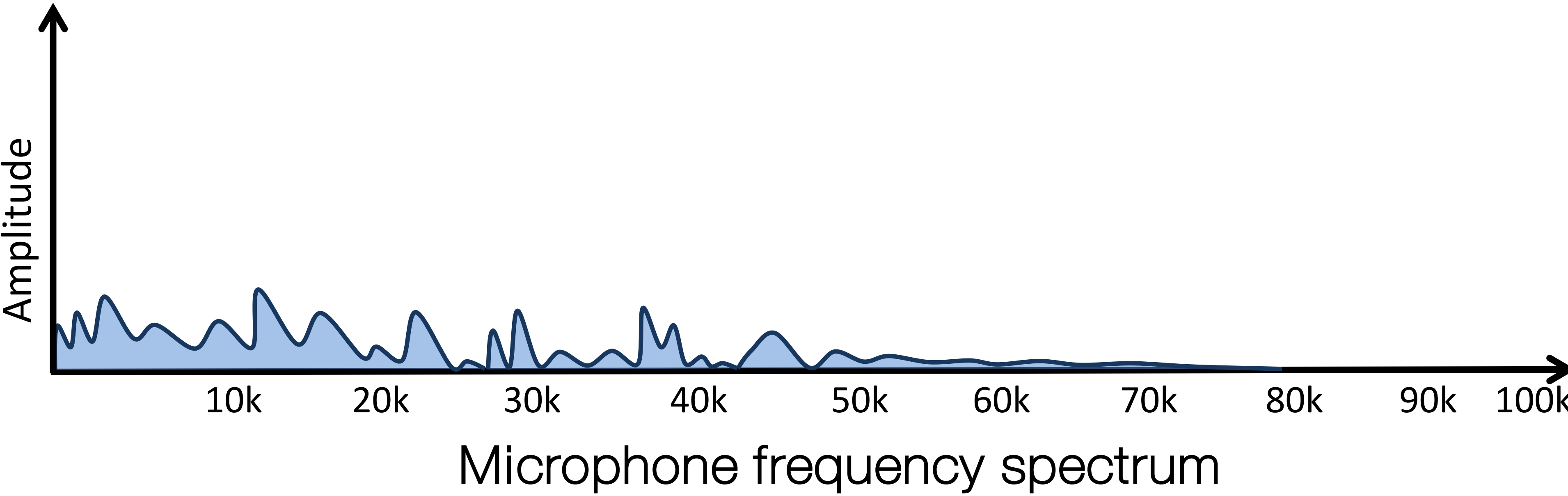
Protecting Privacy (Inaudible Jammer)

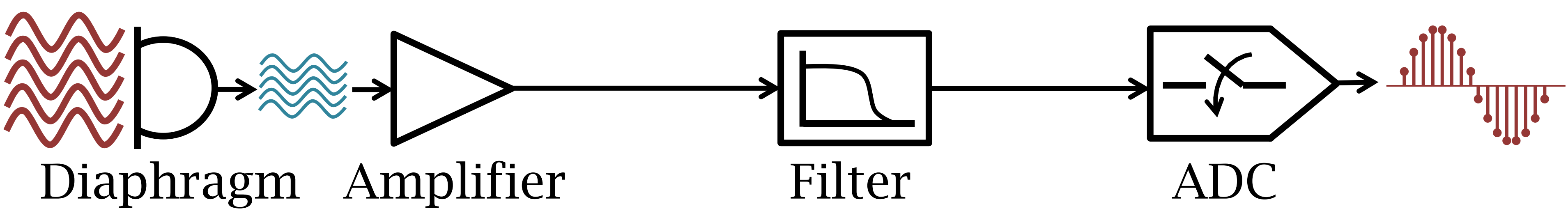
Threats: Inaudible Acoustic Attack

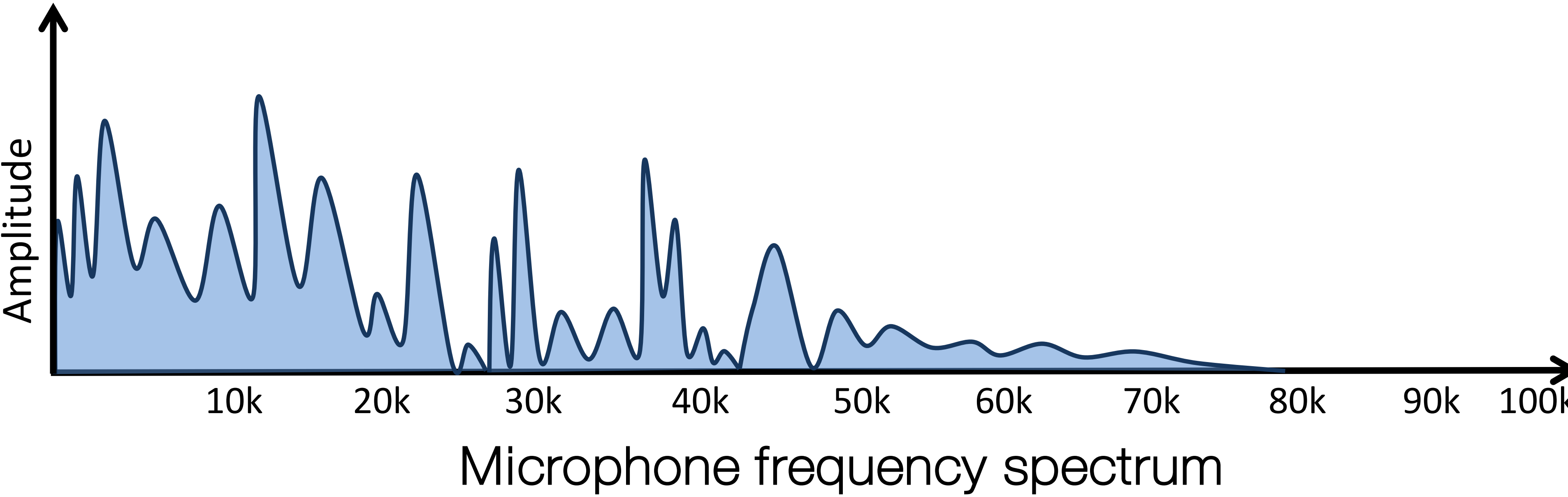


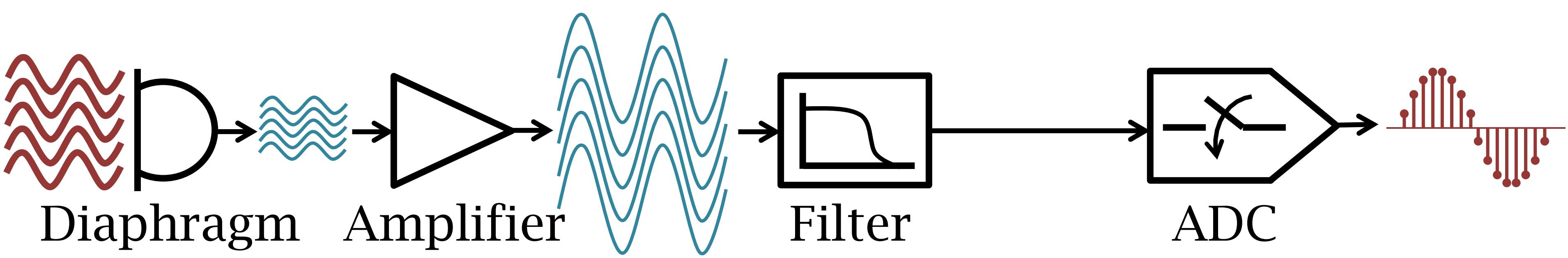

Air vibrations
Digital samples
Diaphragm
Amplifier
Filter
ADC

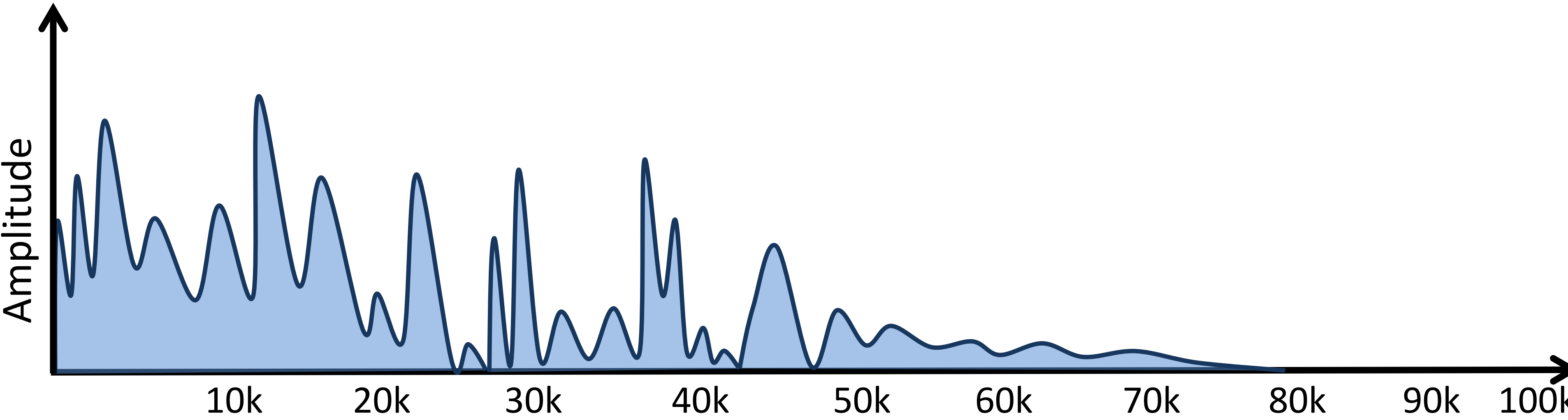

Air vibrations
Digital samples
Diaphragm
Amplifier
Filter
ADC


Amplitude
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Microphone frequency spectrum

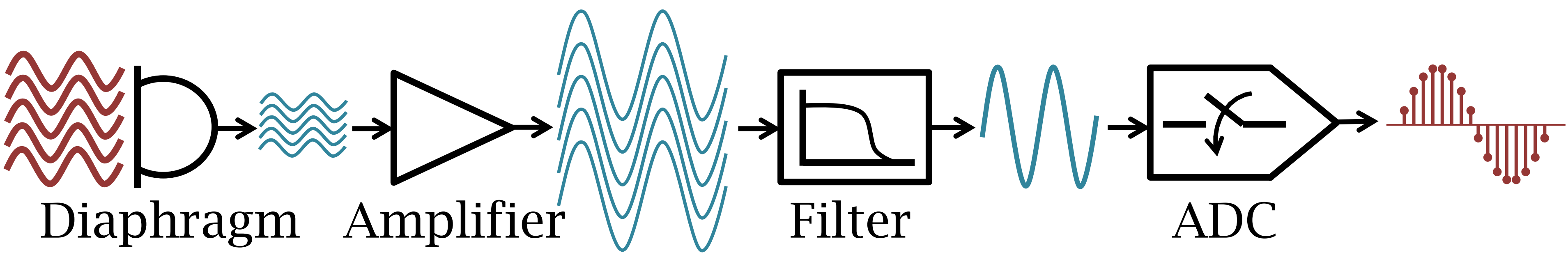

Diaphragm
Amplifier
Filter
ADC

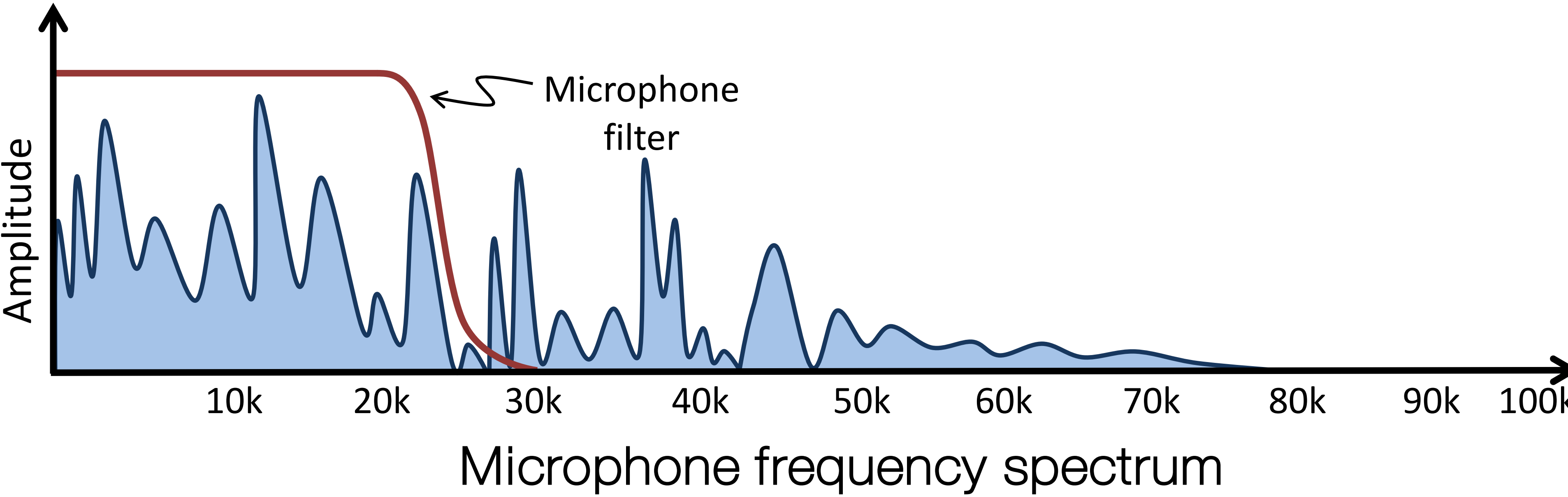

Amplitude
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Microphone frequency spectrum

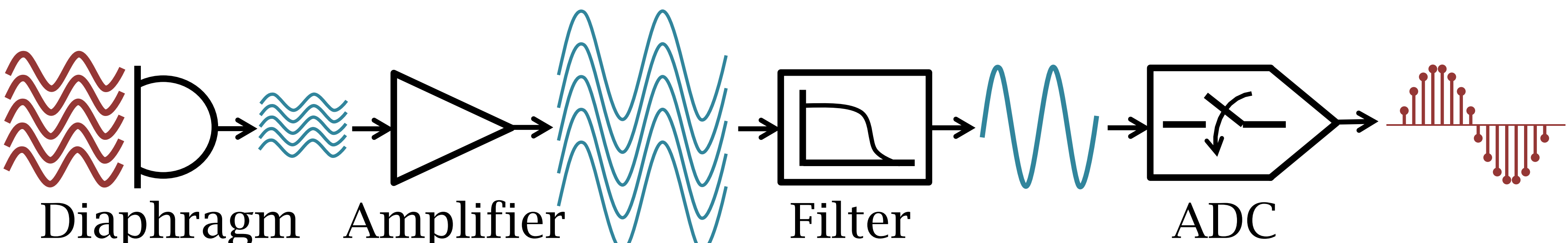

Diaphragm
Amplifier
Filter
ADC

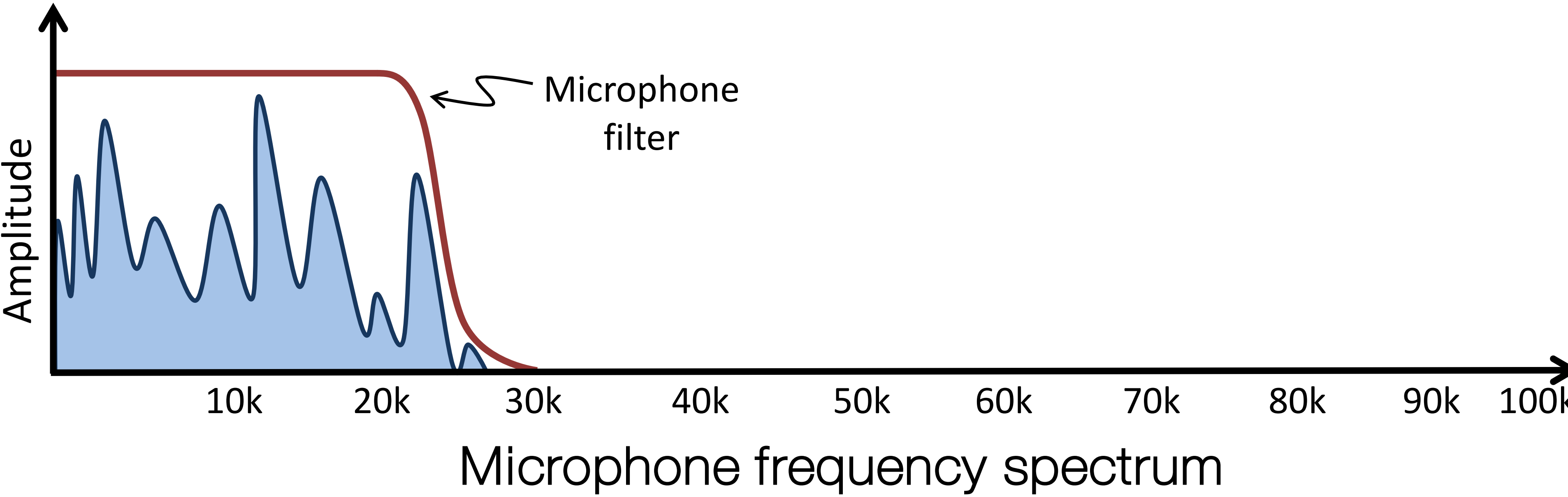


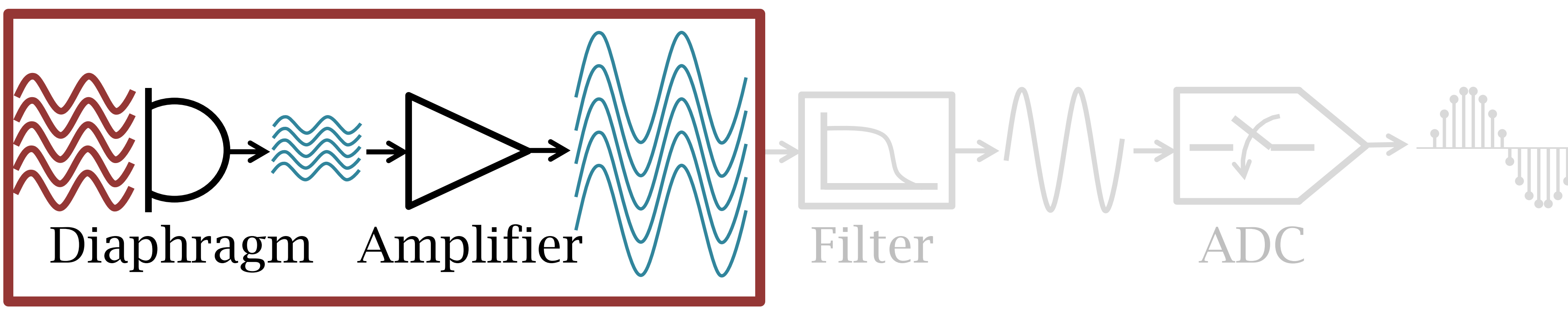
Microphone frequency spectrum

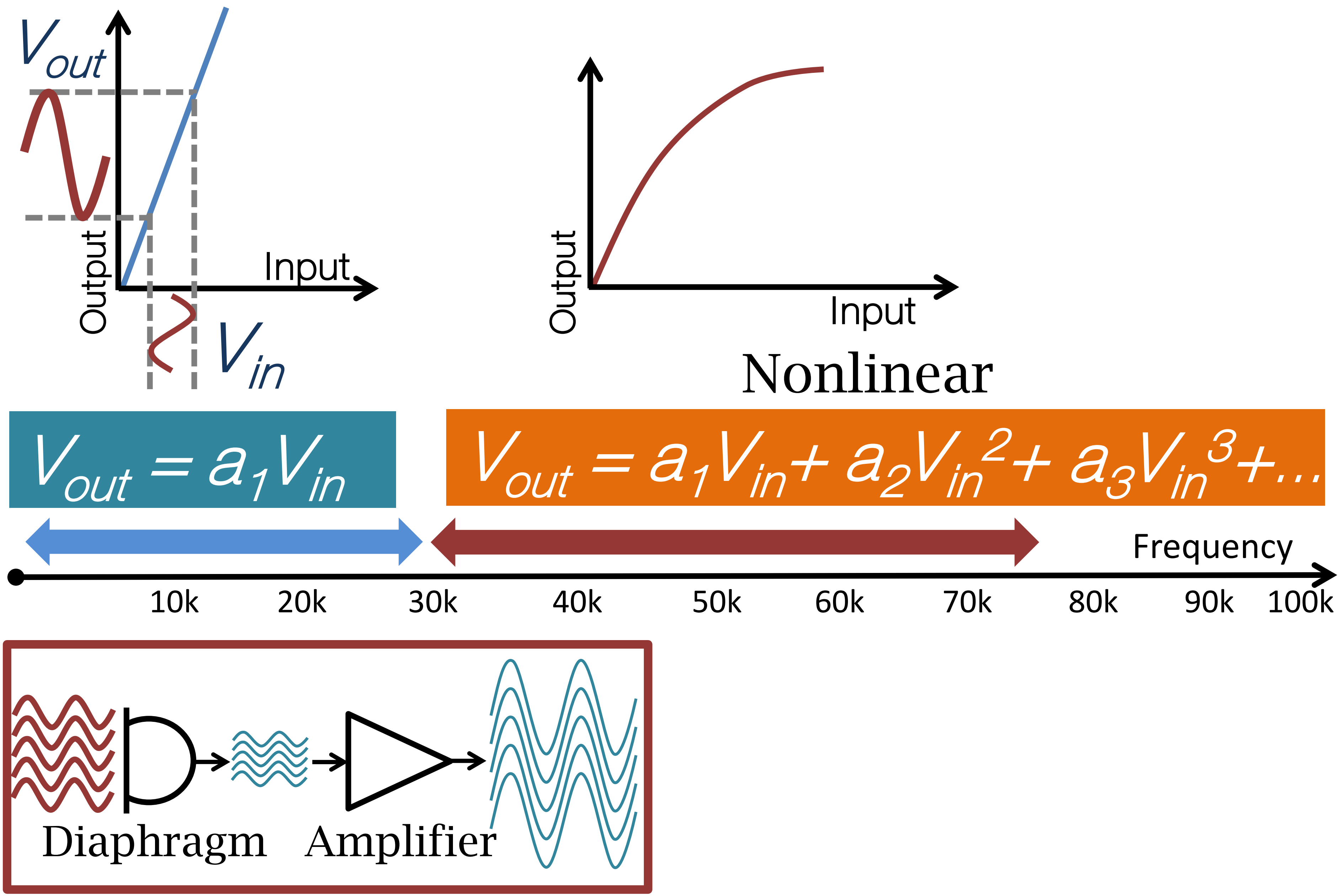


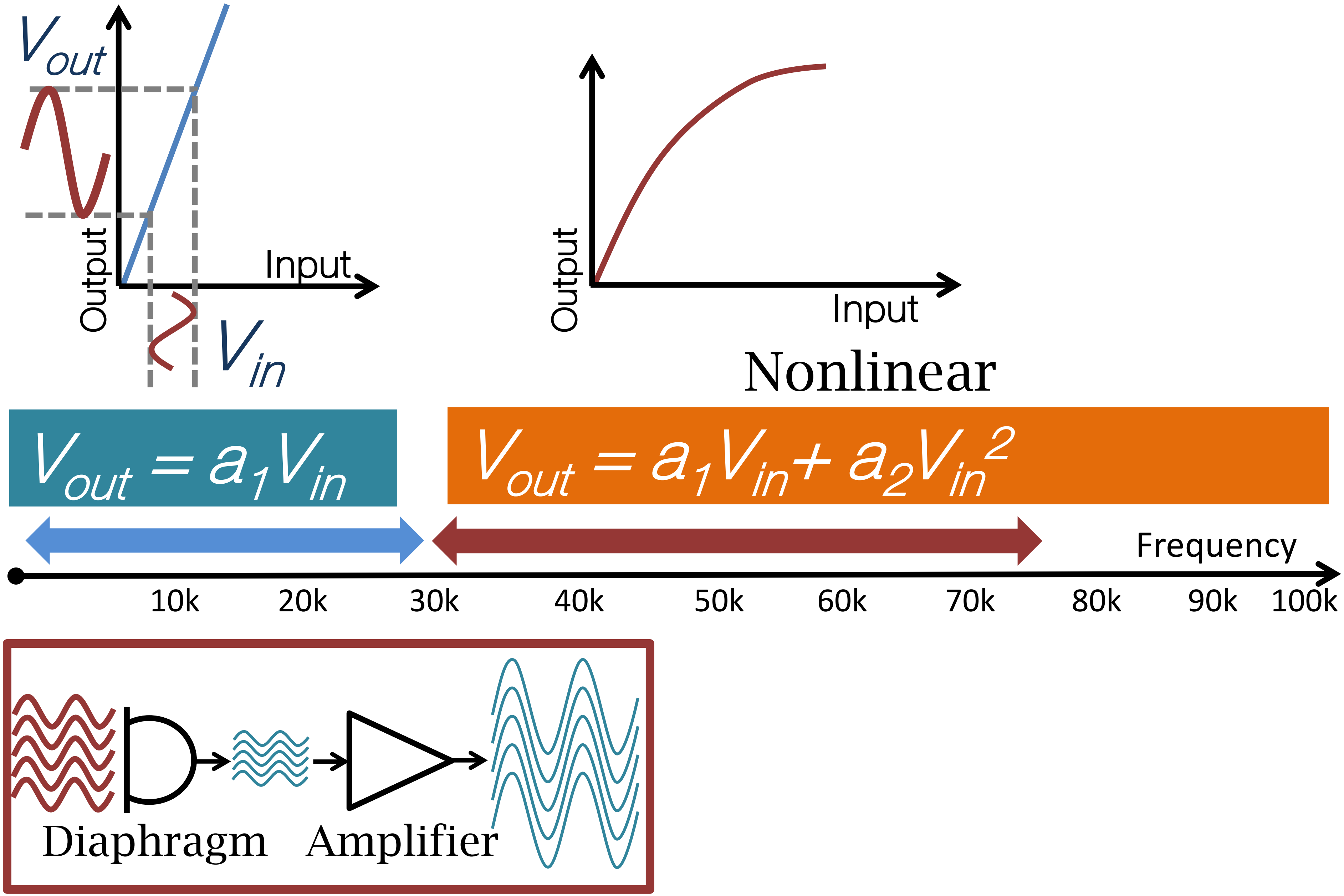

Amplitude
Microphone
filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Microphone frequency spectrum


Diaphragm
Amplifier
Filter
ADC

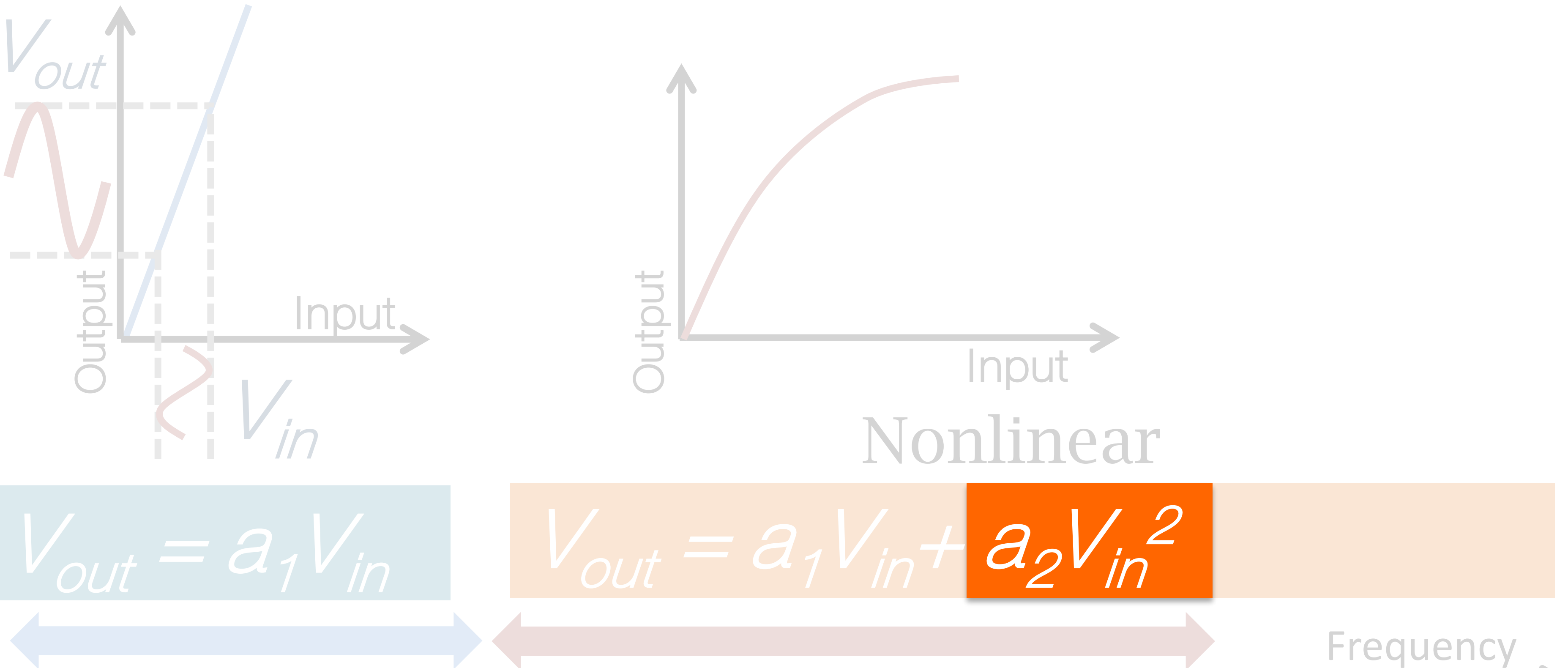

Amplitude
Microphone
filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Microphone frequency spectrum


Diaphragm
Amplifier
Filter
ADC


V_{out}
Output
Input
V_{in}
Output
Input
Nonlinear
$V_{out} = a_1 V_{in}$
$V_{out} = a_1 V_{in} + a_2 V_{in}^2 + a_3 V_{in}^3 + ...$
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Diaphragm
Amplifier

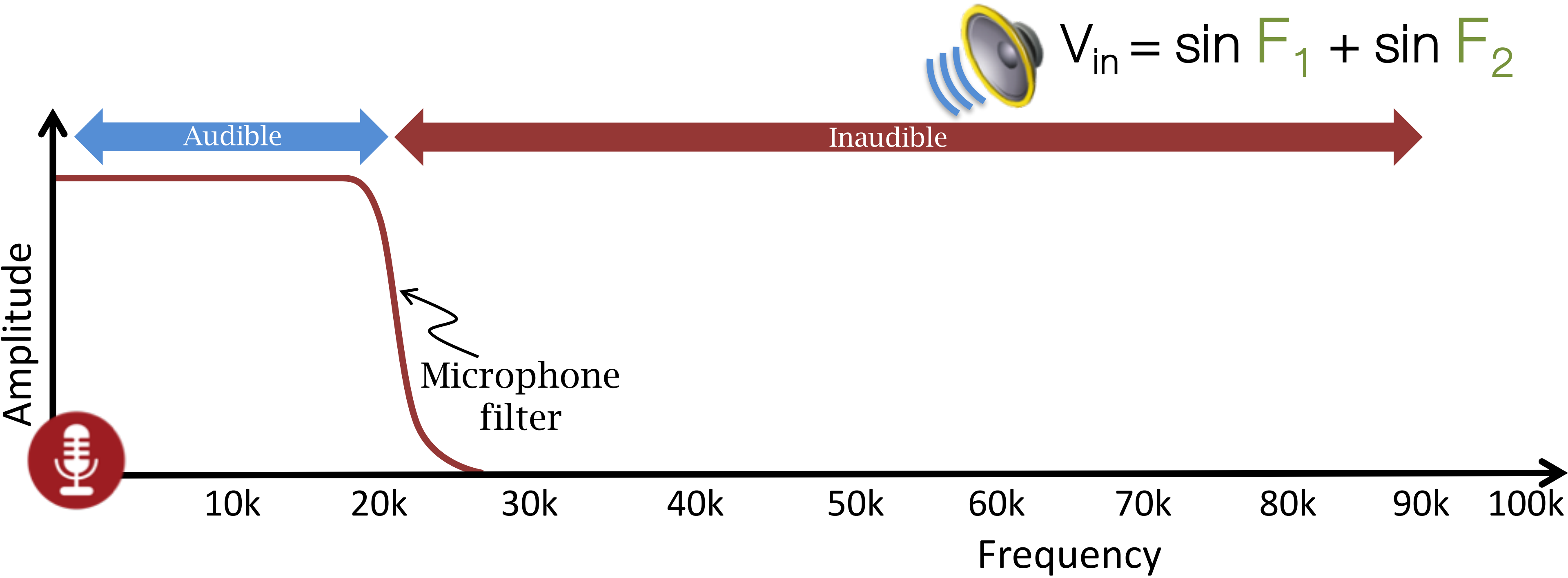


V_{out}
Output
Input
V_{in}
Output
Input
Nonlinear
$V_{out} = a_1 V_{in}$
$V_{out} = a_1 V_{in} + a_2 V_{in}^2$
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Diaphragm
Amplifier

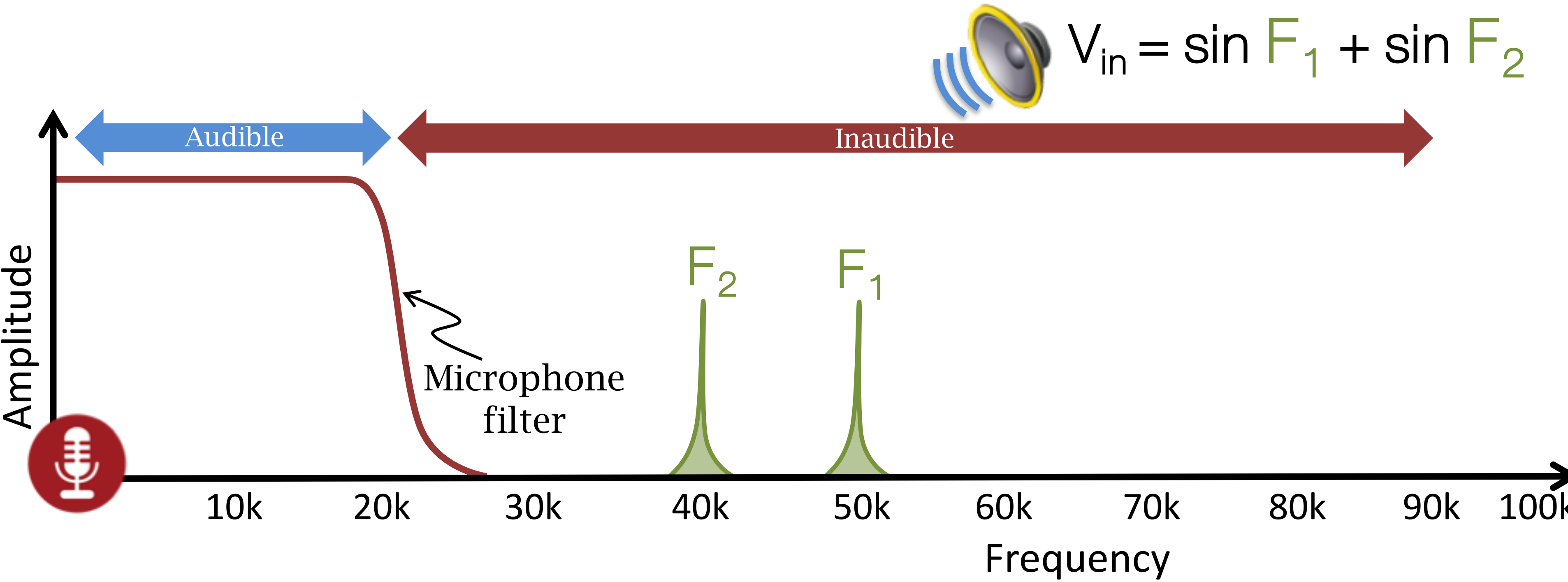


Nonlinearity is fundamental to all microphones

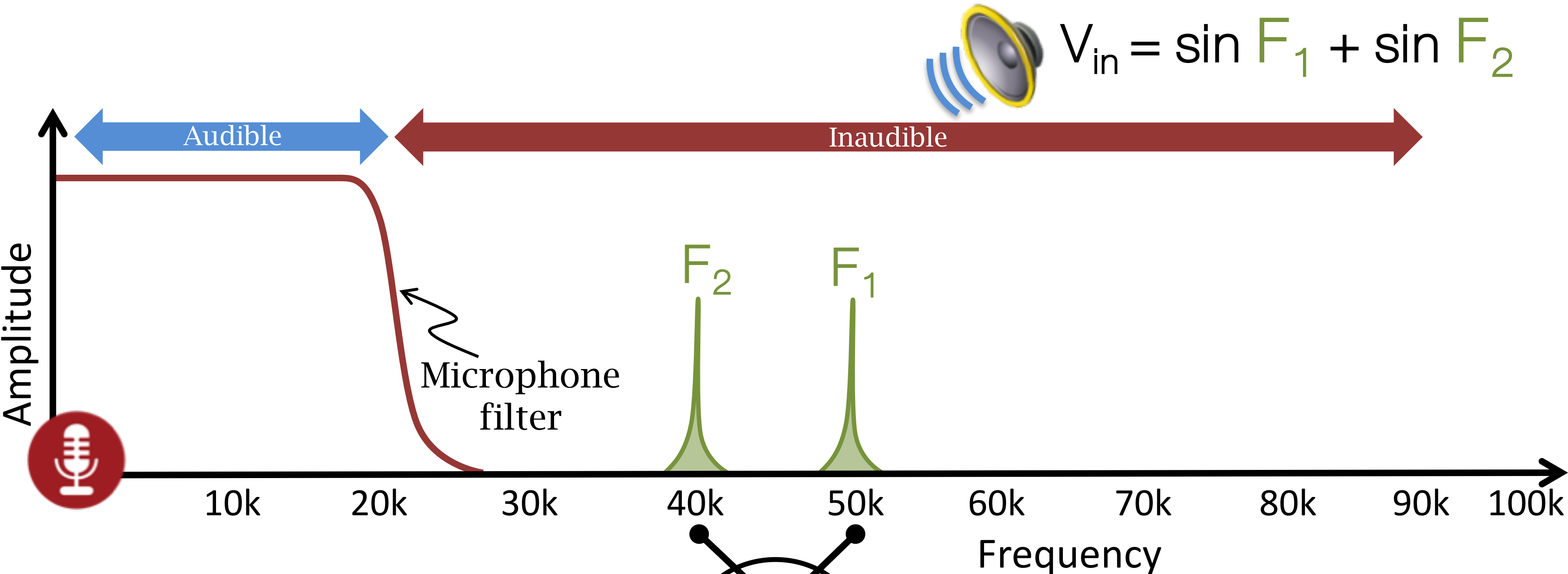
$V_{in} = \sin F_1 + \sin F_2$



$V_{in} = \sin F_1 + \sin F_2$
Audible
Inaudible
Amplitude
Microphone filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency

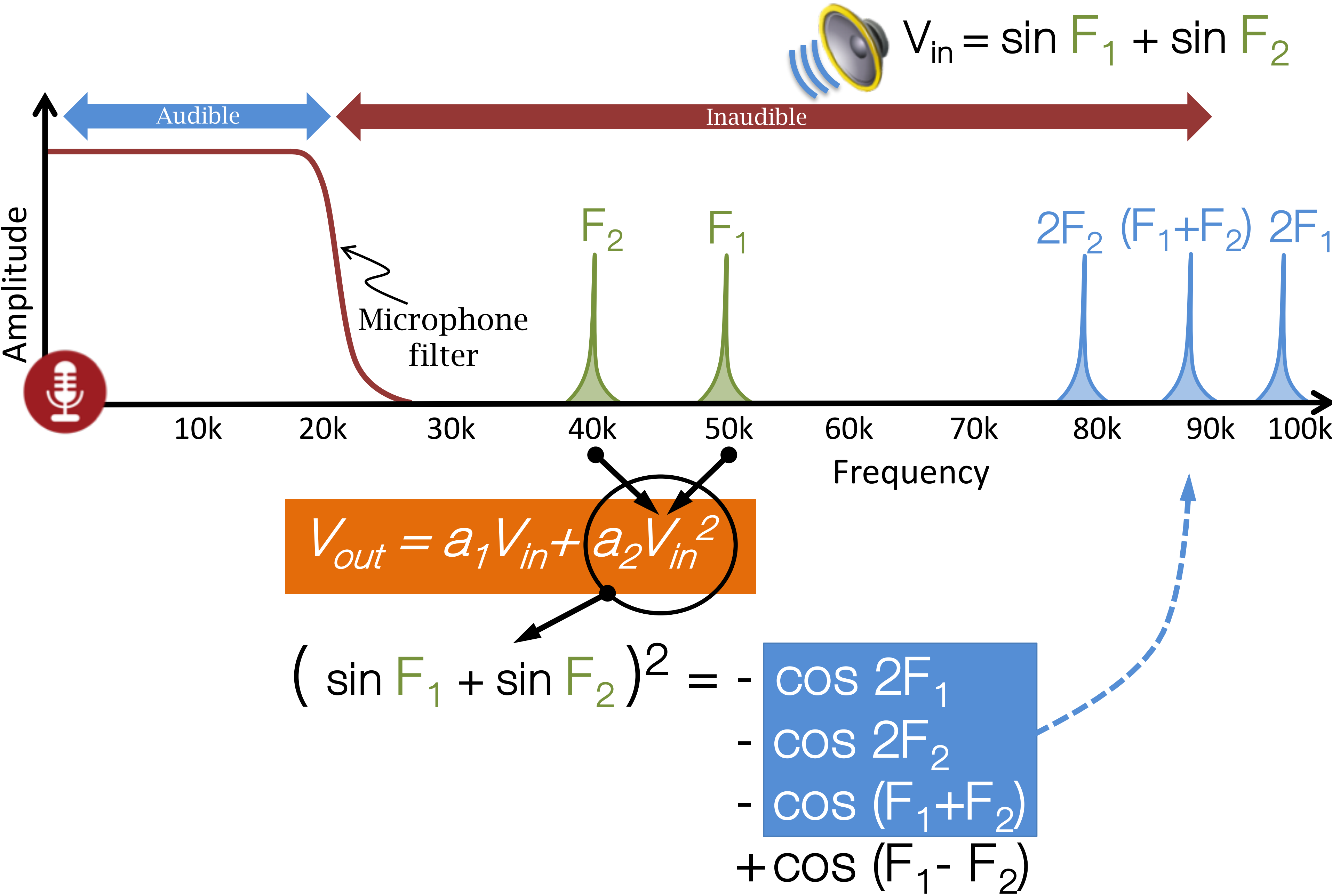


$V_{in} = \sin F_1 + \sin F_2$
Audible
Inaudible
Amplitude
F_2
F_1
Microphone filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency

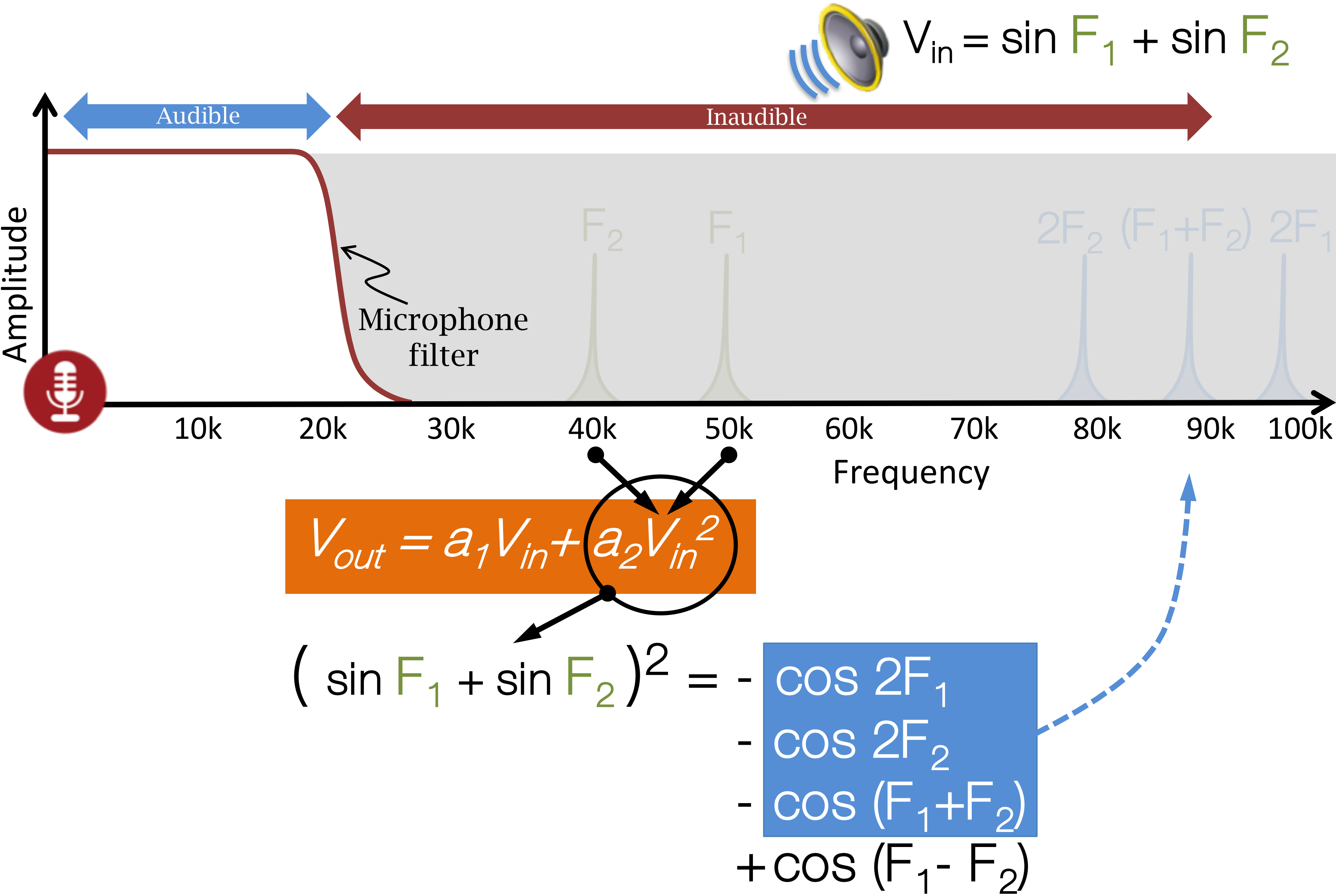


$$V_{out} = a_1 V_{in} + a_2 V_{in}^2$$

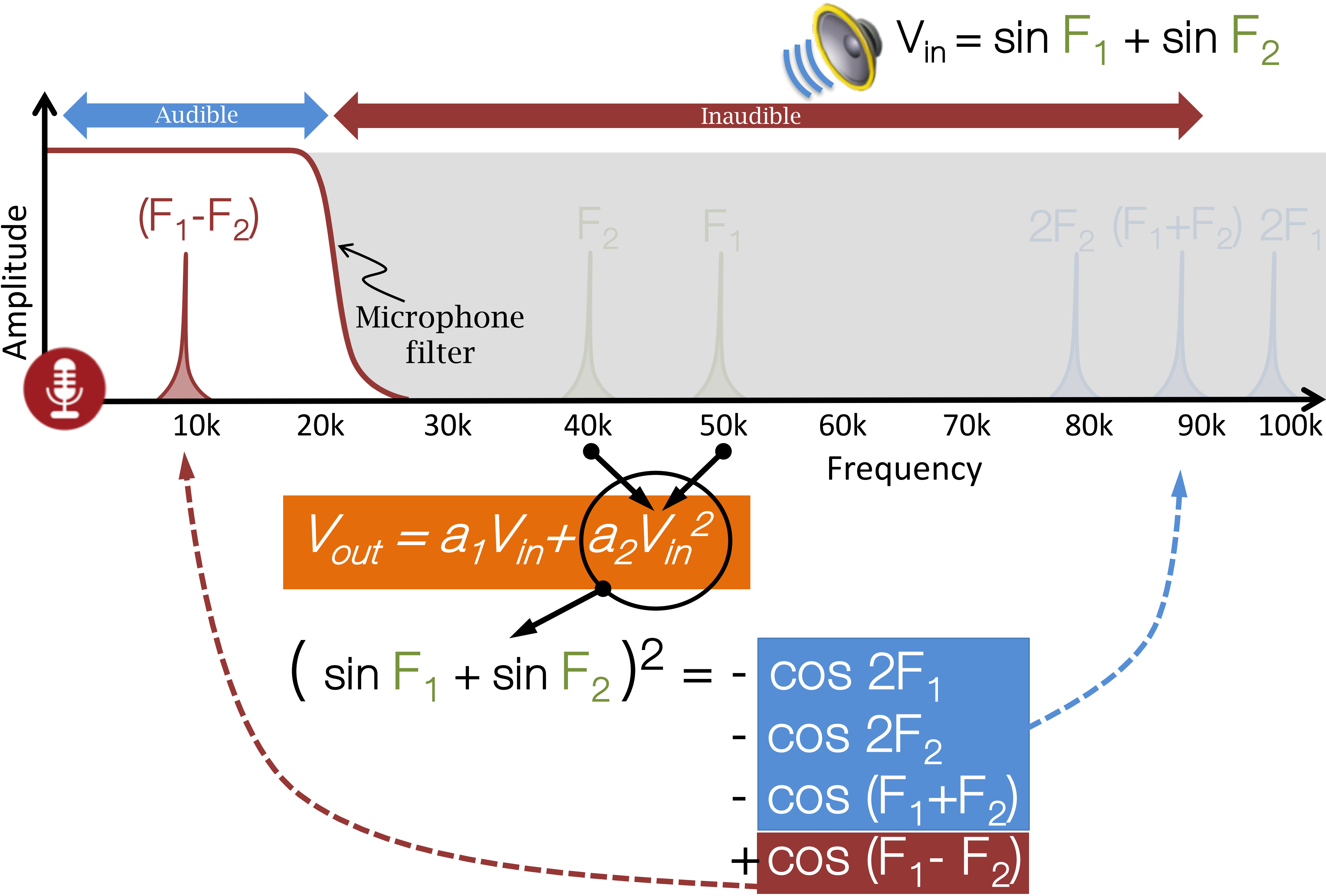
$$(\sin F_1 + \sin F_2)^2 = -\cos 2F_1 - \cos 2F_2 - \cos(F_1 + F_2) + \cos(F_1 - F_2)$$



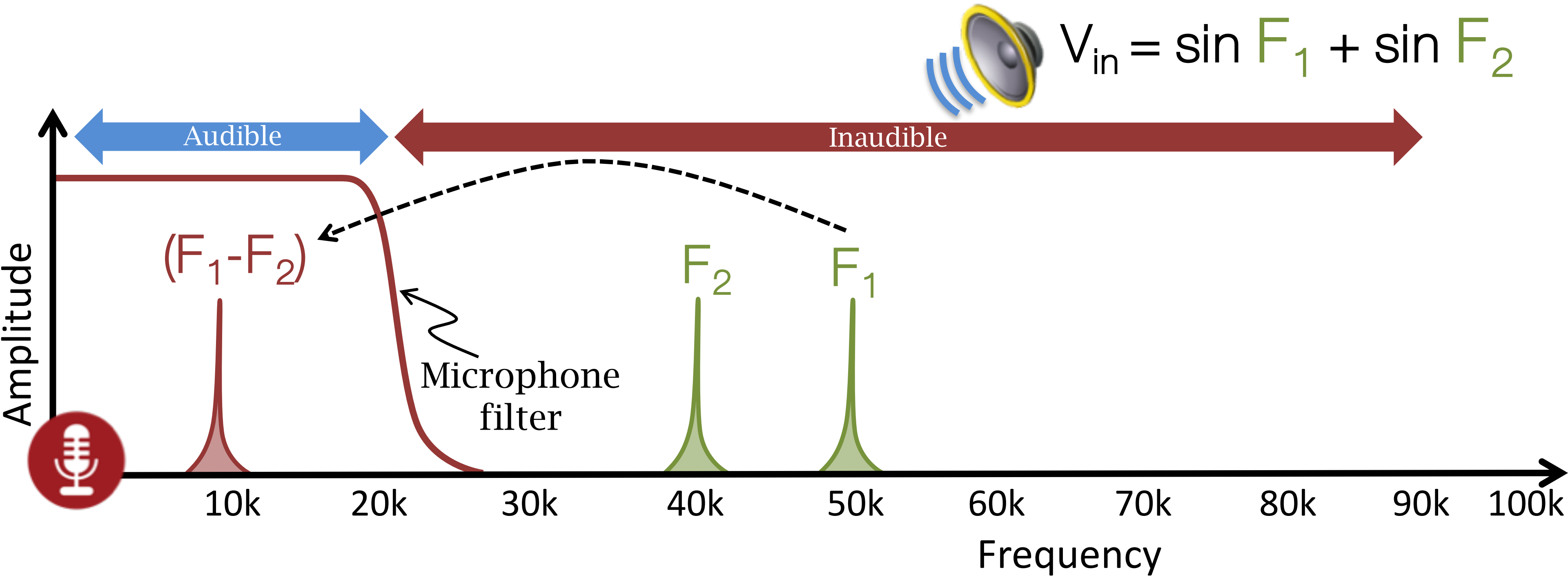
$V_{in} = \sin F_1 + \sin F_2$
Audible
Inaudible
Amplitude
Microphone filter
F_2
F_1
$2F_2$
(F_1+F_2)
$2F_1$
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
$V_{out} = a_1 V_{in} + a_2 V_{in}^2$
$(\sin F_1 + \sin F_2)^2 = -\cos 2F_1 - \cos 2F_2 - \cos (F_1+F_2) + \cos (F_1 - F_2)$

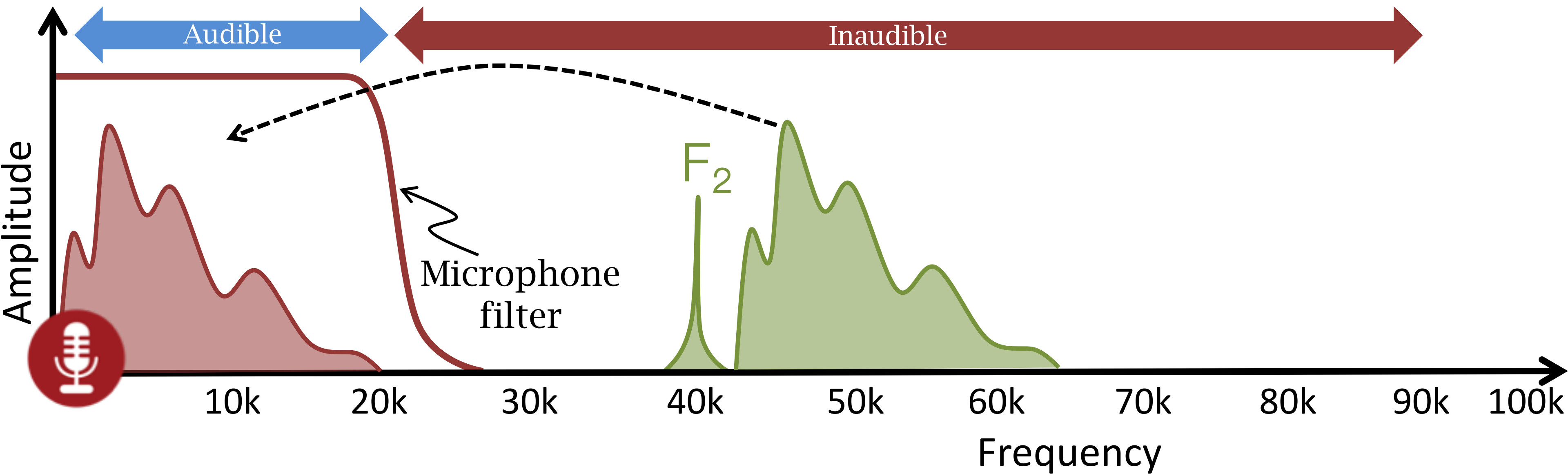


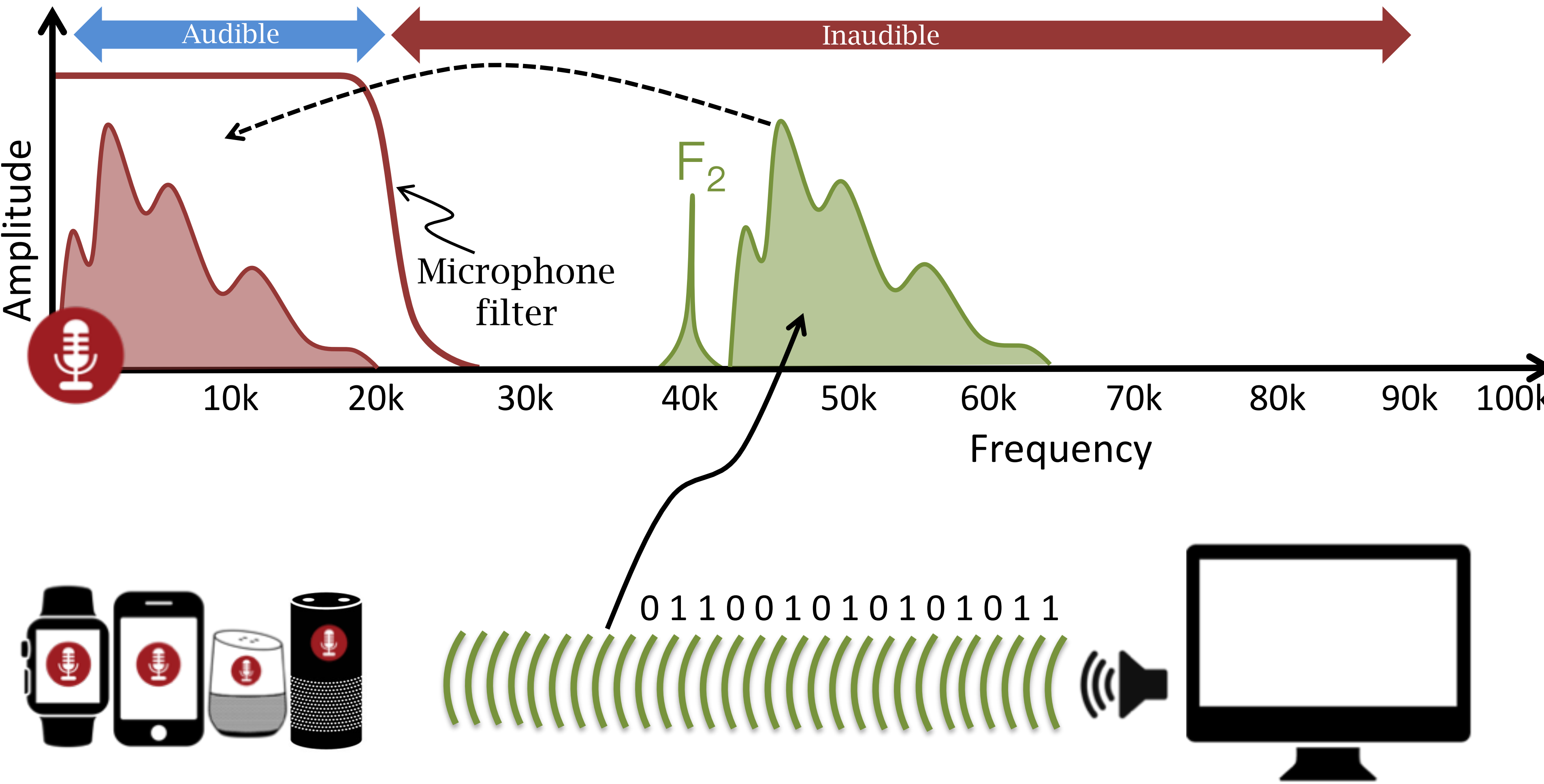
$V_{in} = \sin F_1 + \sin F_2$
Audible
Inaudible
Amplitude
Microphone filter
F_2
F_1
$2F_2$
(F_1+F_2)
$2F_1$
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
$V_{out} = a_1 V_{in} + a_2 V_{in}^2$
$(\sin F_1 + \sin F_2)^2 = - \cos 2F_1 - \cos 2F_2 - \cos (F_1+F_2) + \cos (F_1 - F_2)$

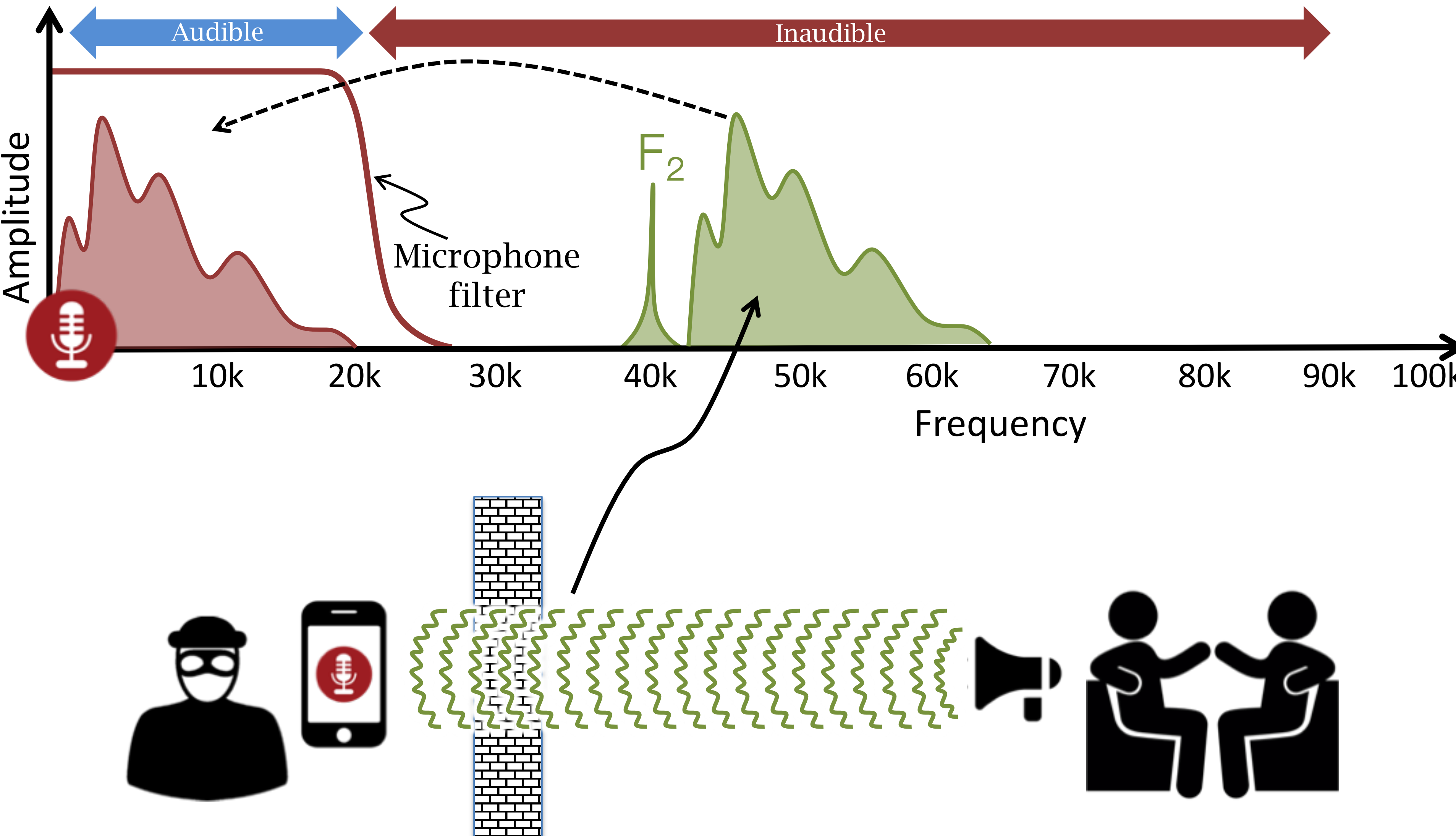


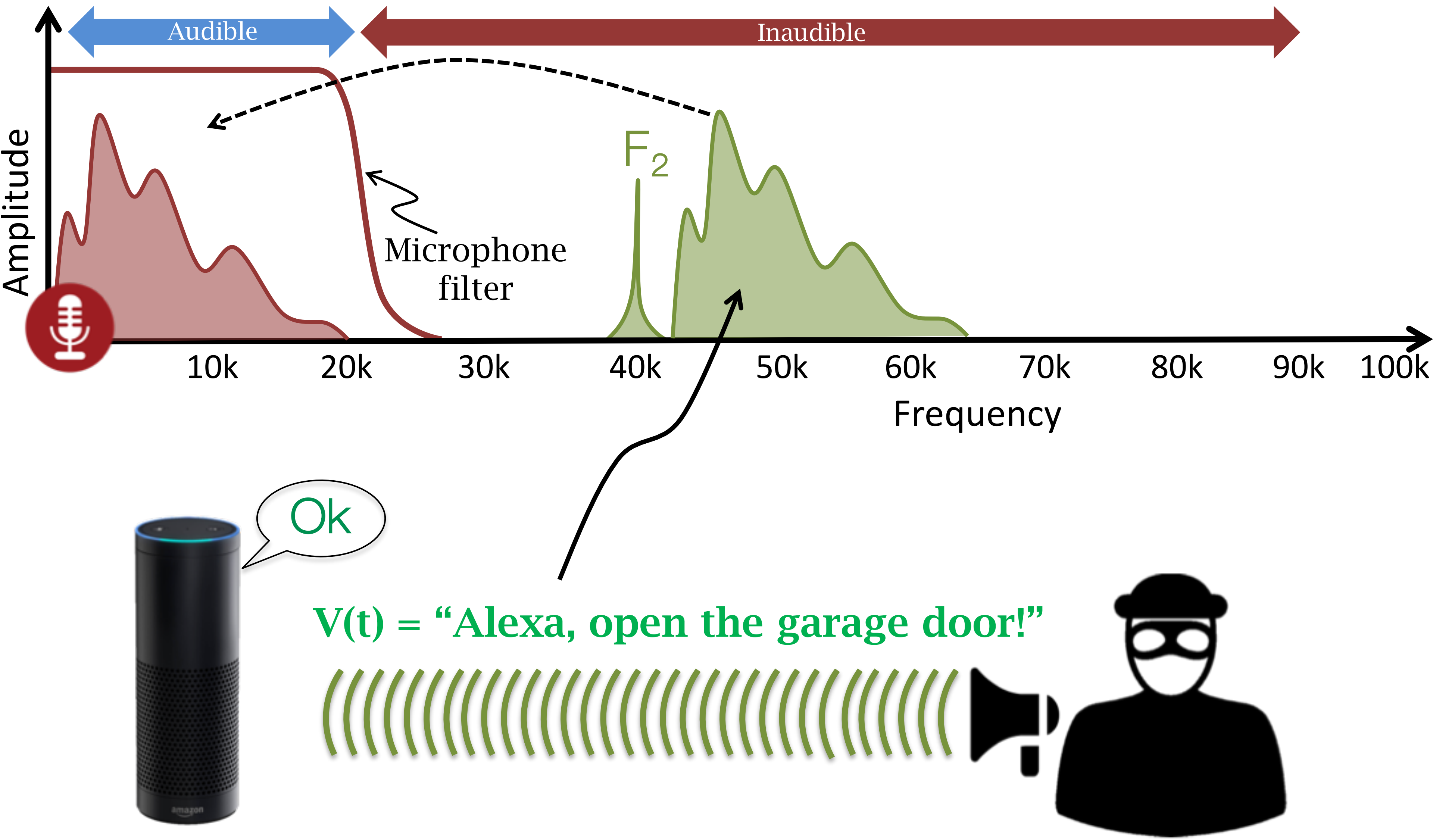
V_in = sin F_1 + sin F_2
Audible
Inaudible
Amplitude
(F_1-F_2)
F_2
F_1
2F_2
(F_1+F_2)
2F_1
Microphone filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
V_out = a_1 V_in + a_2 V_in^2
(sin F_1 + sin F_2)^2 = - cos 2F_1
- cos 2F_2
- cos (F_1+F_2)
+ cos (F_1 - F_2)

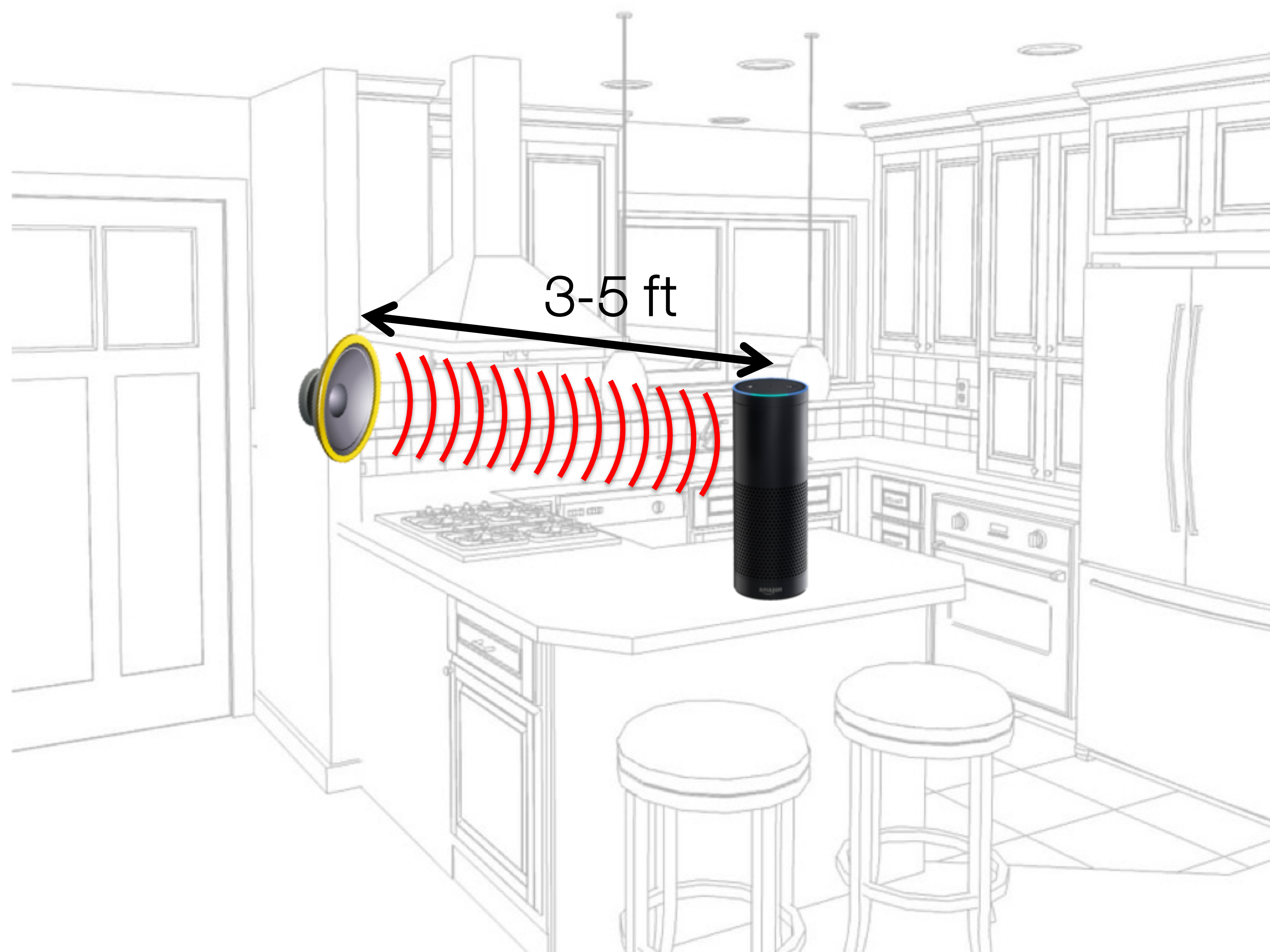

$V_{in} = \sin F_1 + \sin F_2$
Audible
Inaudible
Amplitude
(F_1-F_2)
F_2
F_1
Microphone filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency


Audible
Inaudible
Amplitude
F2
Microphone filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency

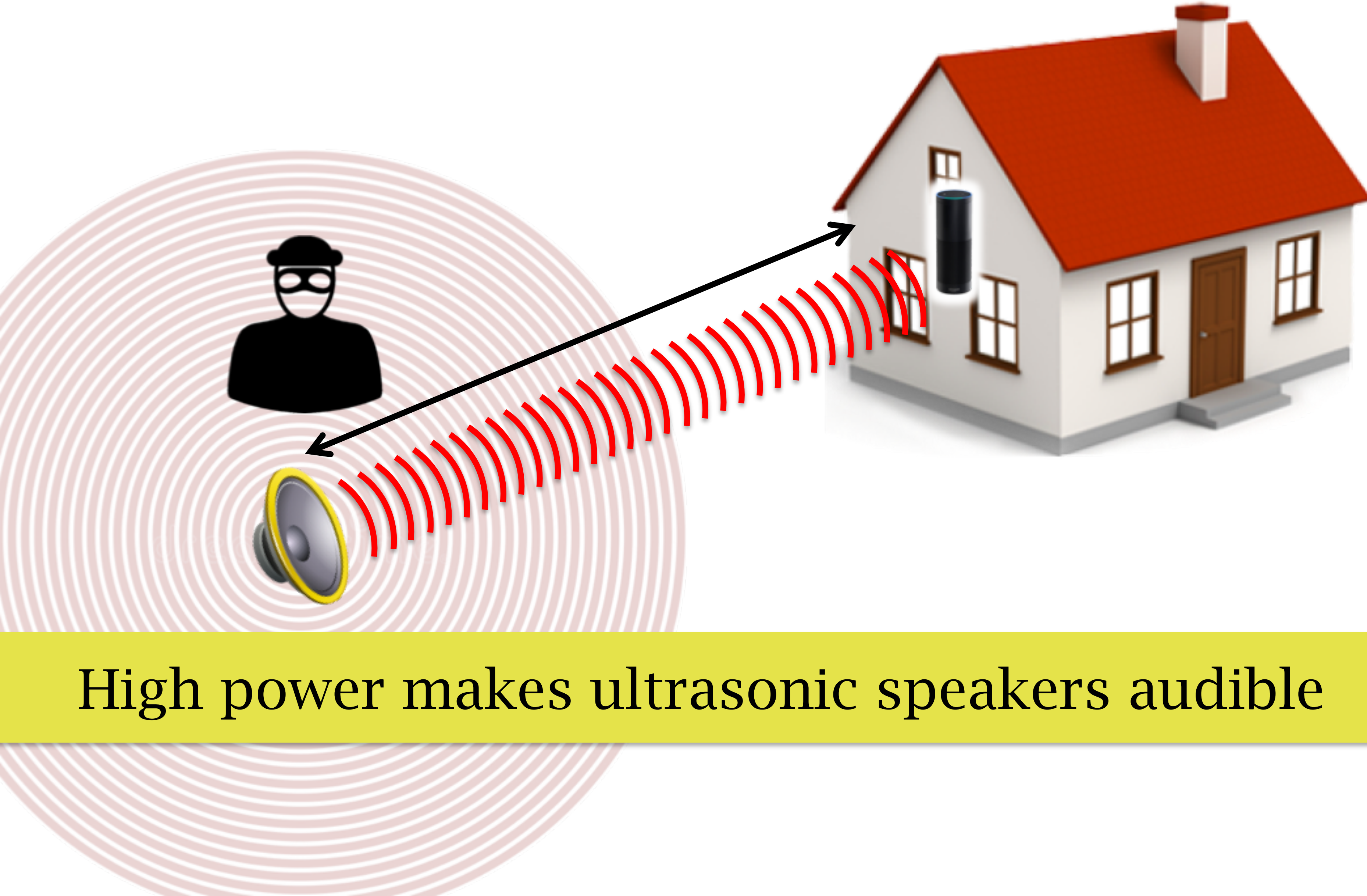

Audible
Inaudible
Amplitude
F2
Microphone
filter
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
0 1 1 0 0 1 0 1 0 1 0 1 0 1 1


Audible
Inaudible
Amplitude
Microphone filter
F2
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency


Audible
Inaudible
Amplitude
Microphone filter
F2
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
Ok
V(t) = "Alexa, open the garage door!"


3-5 ft

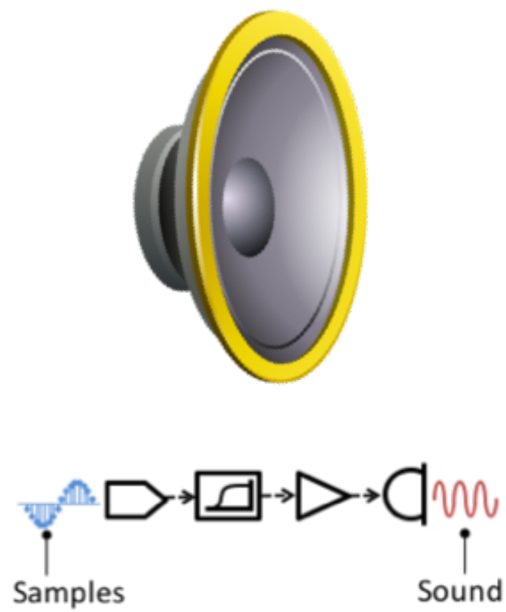
Can someone attack from a longer range?

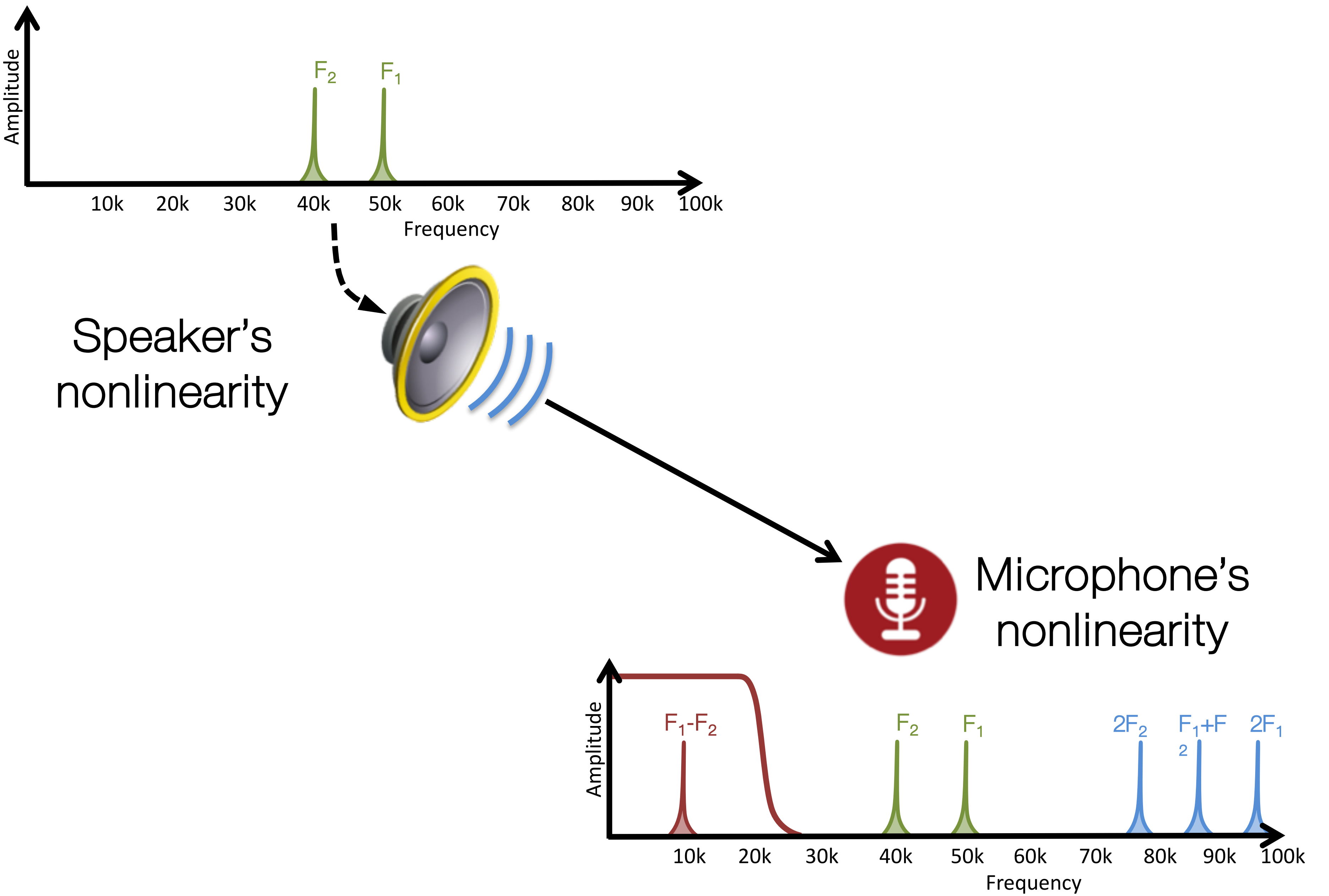


High power makes ultrasonic speakers audible

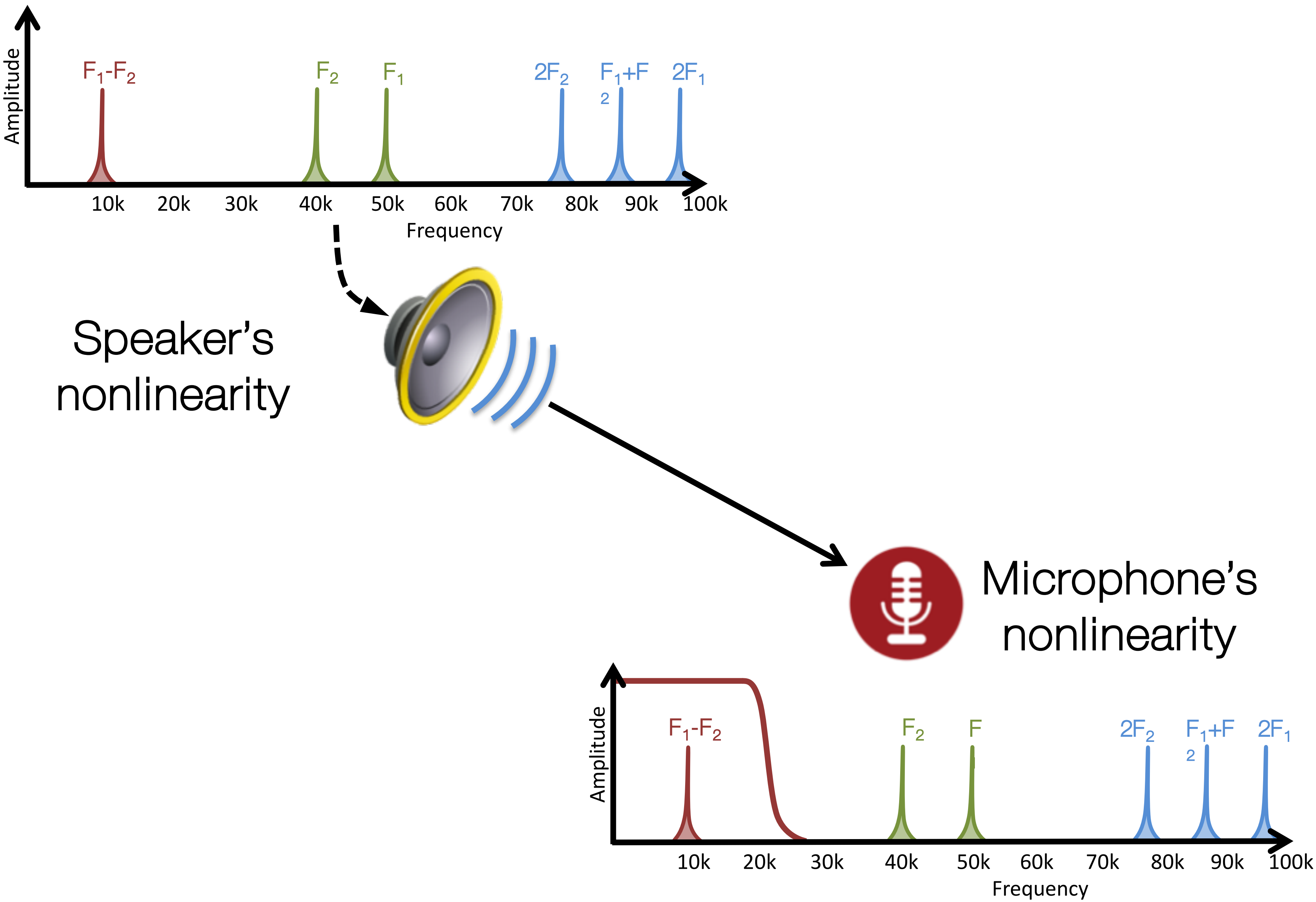

Sound
Samples

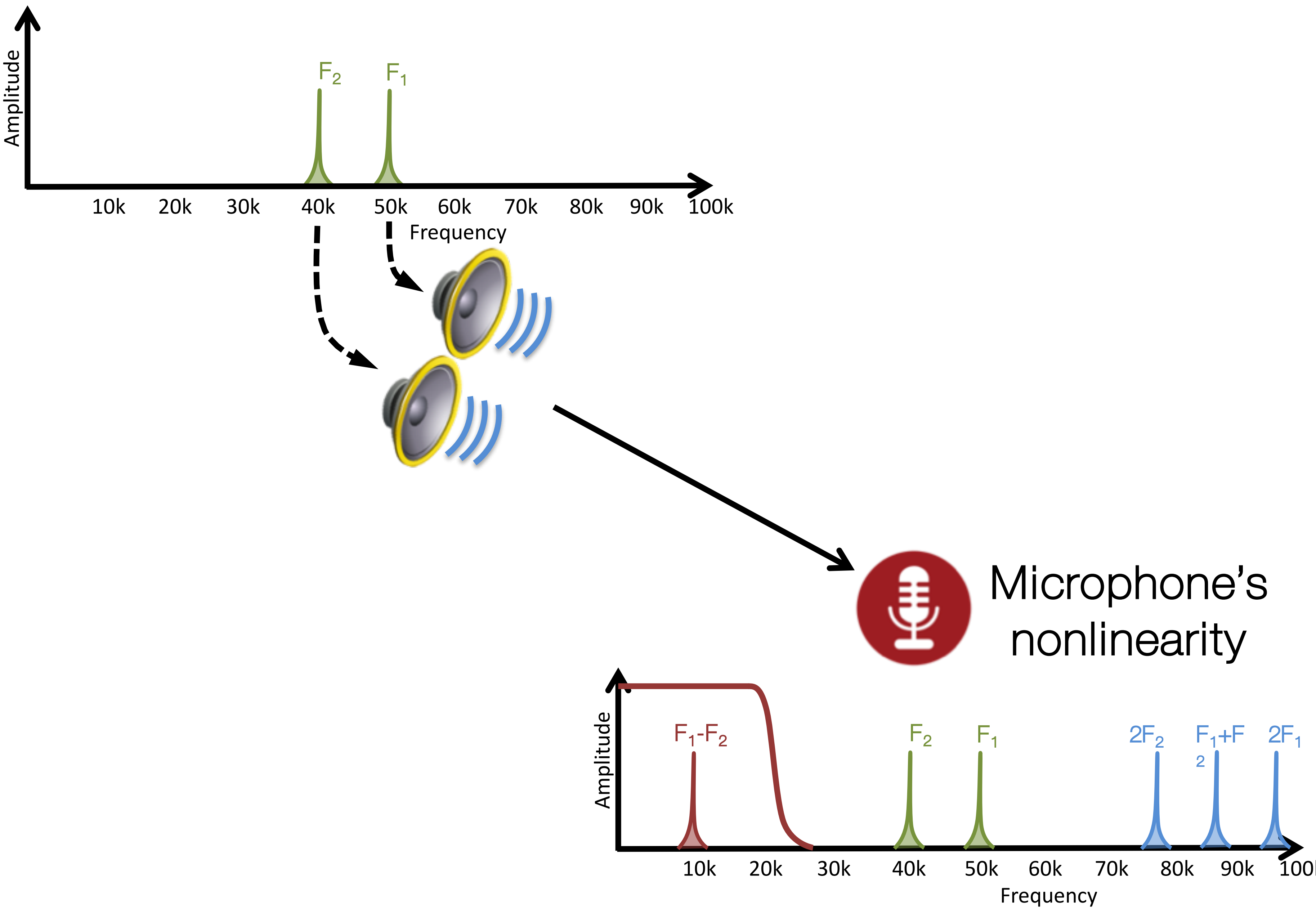
≈


Samples
Sound

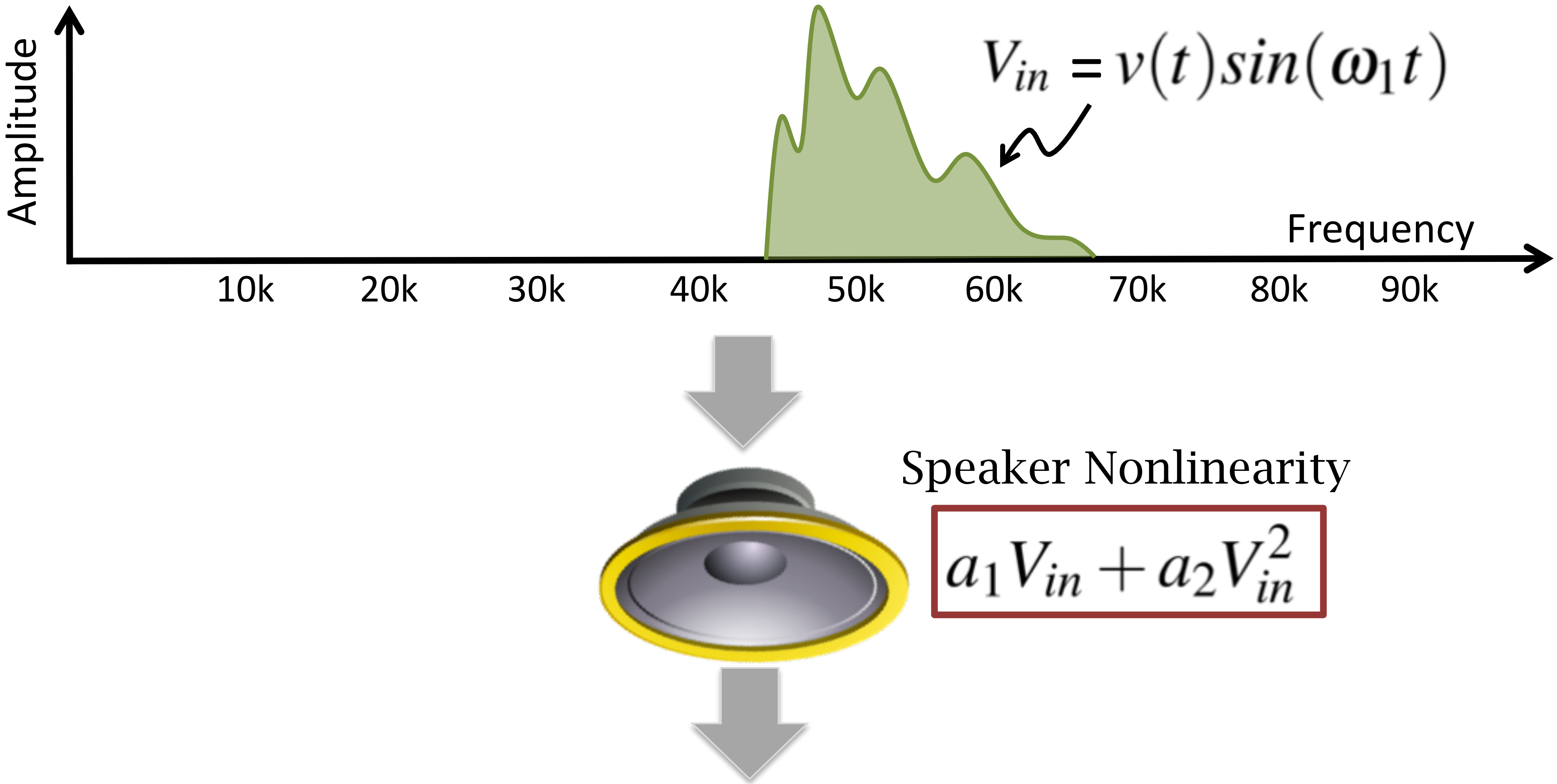


Amplitude
F2
F1
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
Speaker's nonlinearity
Microphone's nonlinearity
Amplitude
F1-F2
F2
F1
2F2
F1+F2
2F1
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency

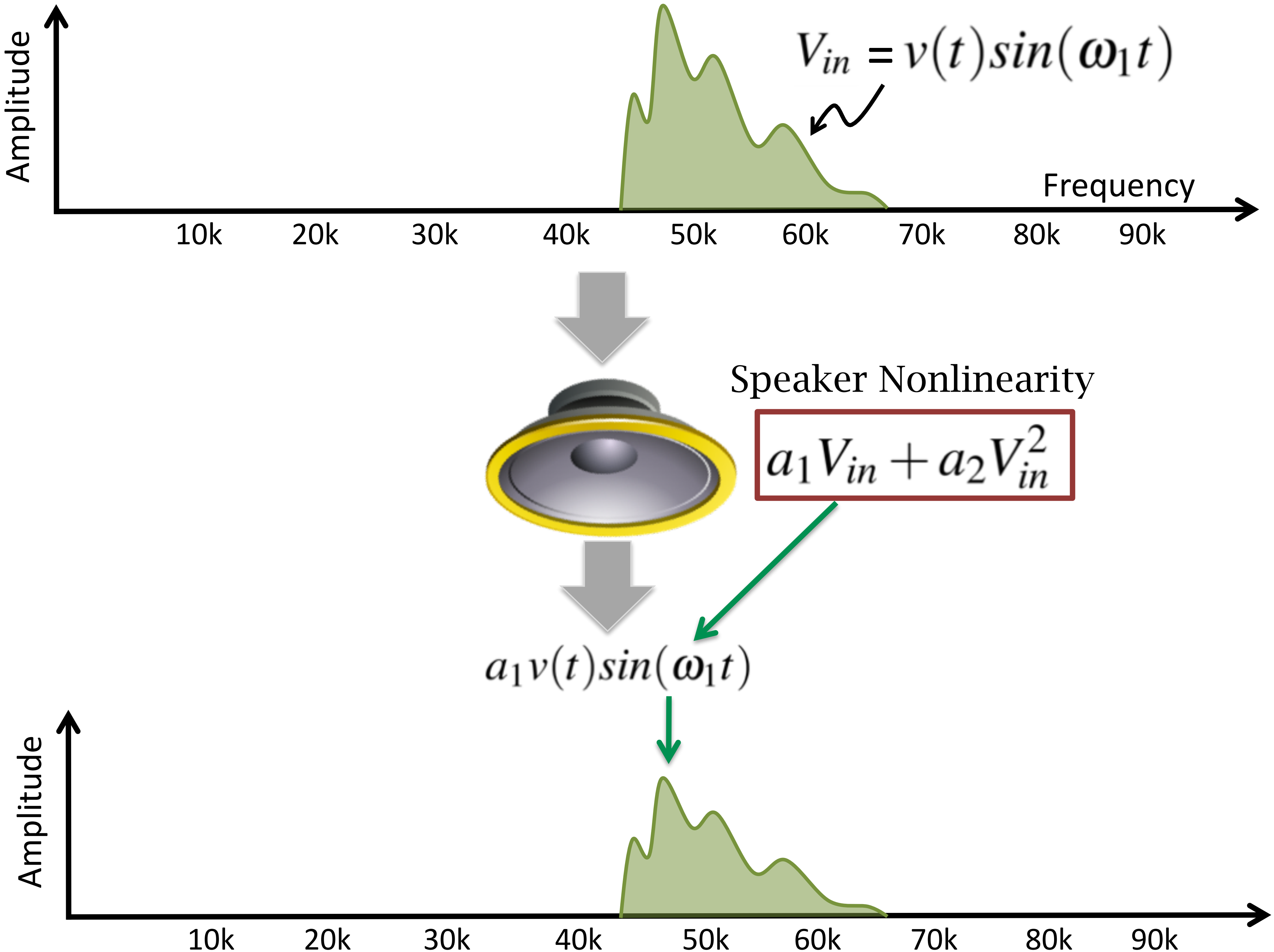

Amplitude
F_1-F_2
F_2
F_1
$2F_2$
F_1+F
2
$2F_1$
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency
Speaker's nonlinearity
Microphone's nonlinearity
Amplitude
F_1-F_2
F_2
F
$2F_2$
F_1+F
2
$2F_1$
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Frequency



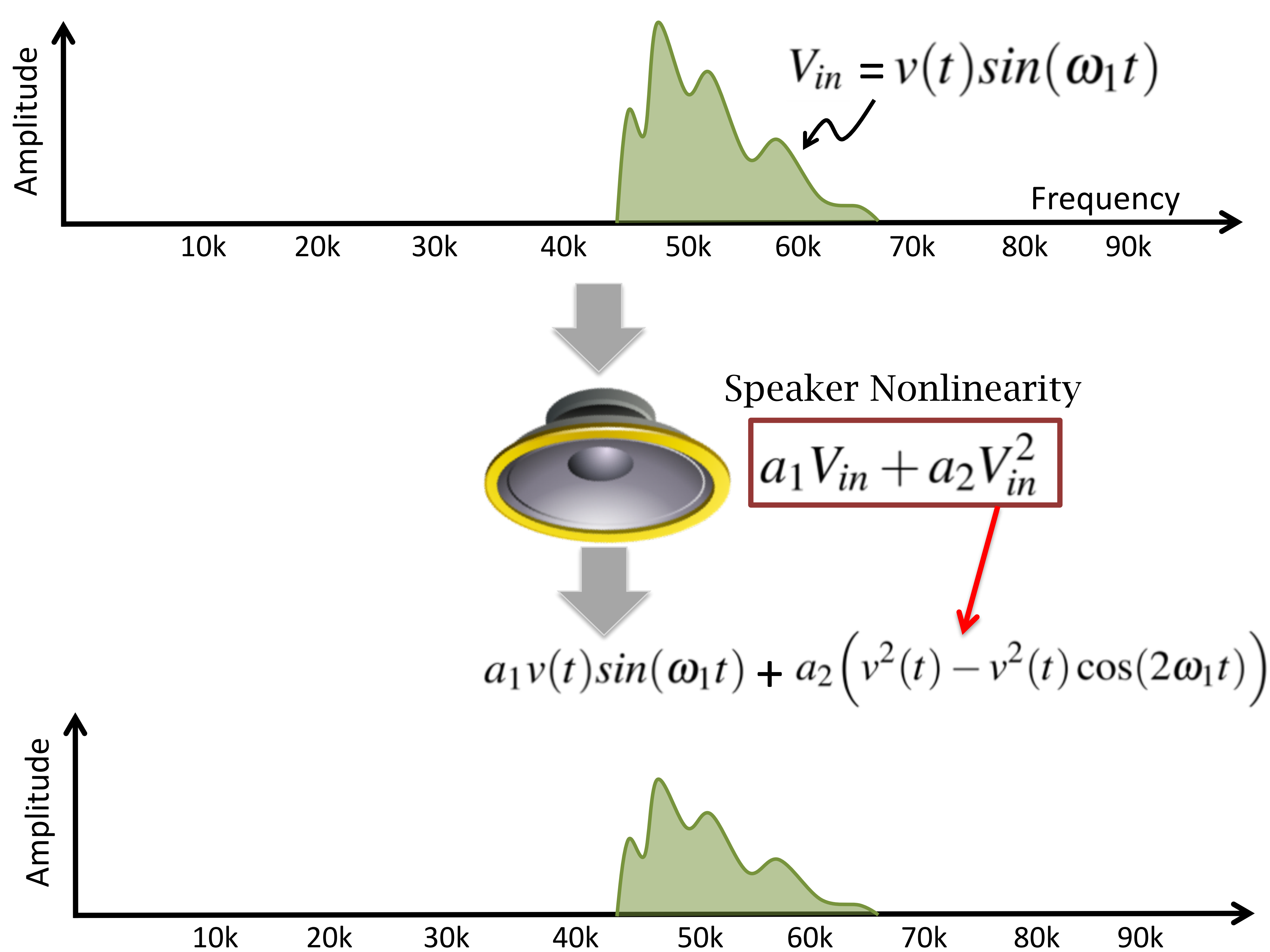
Amplitude
F2
F1
10k 20k 30k 40k 50k 60k 70k 80k 90k 100k
Frequency
Microphone's nonlinearity
Amplitude
F1-F2
F2
F1
2F2
F1+F2
2F1
10k 20k 30k 40k 50k 60k 70k 80k 90k 100k
Frequency

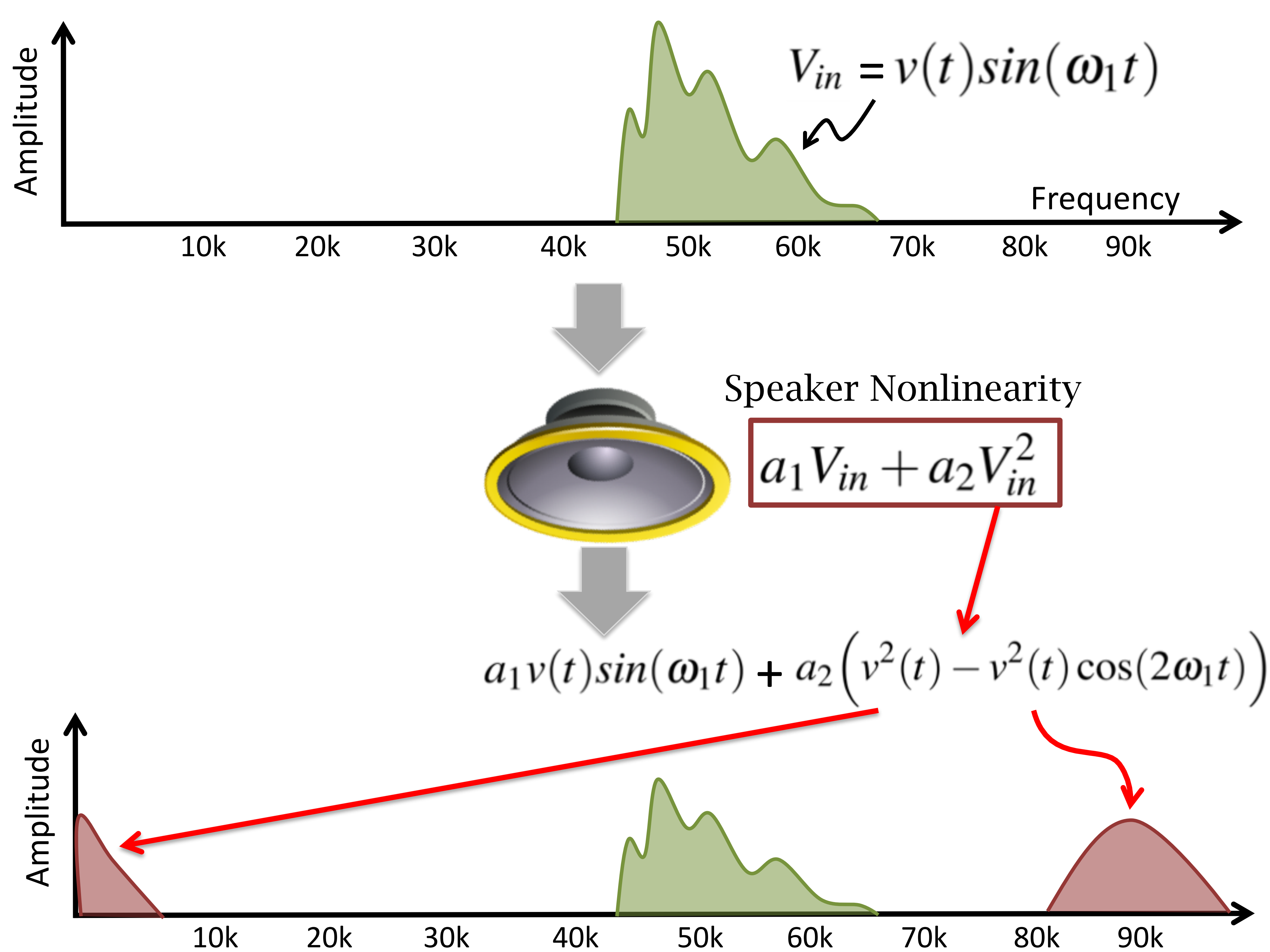


Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
$V_{in} = v(t)sin(\omega_1 t)$
Speaker Nonlinearity
$a_1 V_{in} + a_2 V_{in}^2$

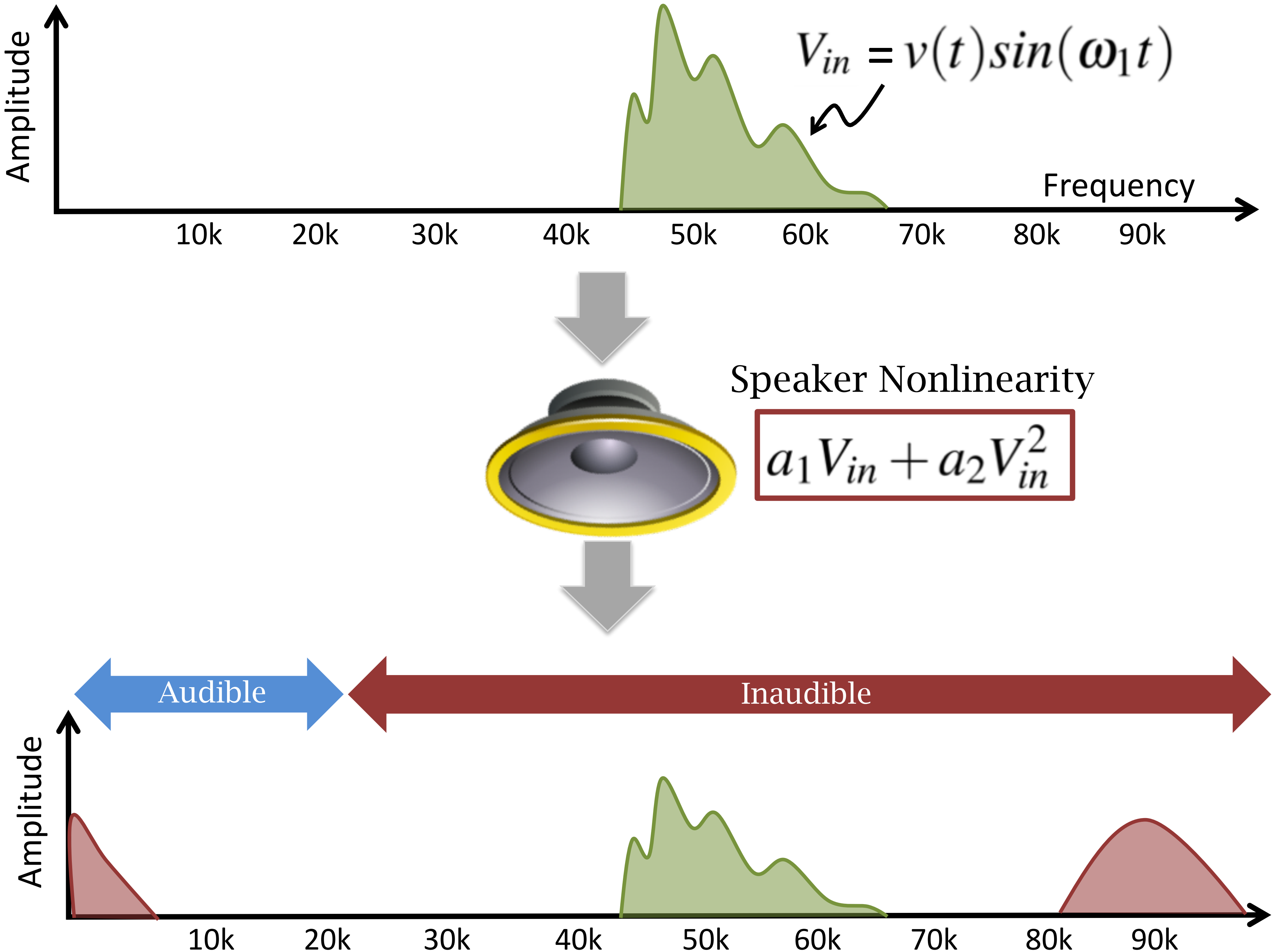


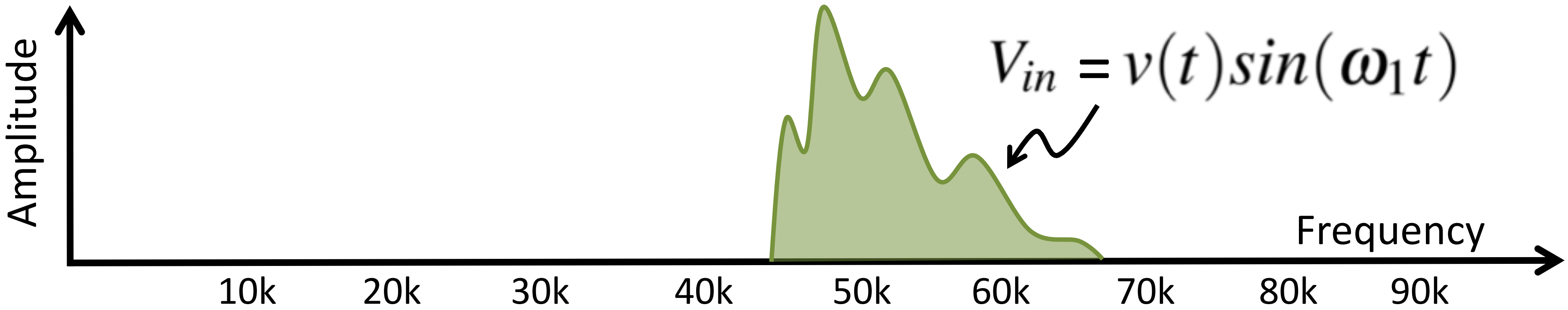
Amplitude
$V_{in} = v(t)sin(\omega_1 t)$
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
Speaker Nonlinearity
$a_1 V_{in} + a_2 V_{in}^2$
$a_1 v(t)sin(\omega_1 t)$
Amplitude
10k
20k
30k
40k
50k
60k
70k
80k
90k


Amplitude
$V_{in} = v(t)sin(\omega_1 t)$
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
Speaker Nonlinearity
$a_1 V_{in} + a_2 V_{in}^2$
$a_1 v(t)sin(\omega_1 t) + a_2\left(v^2(t) - v^2(t)\cos(2\omega_1 t)\right)$
Amplitude
10k
20k
30k
40k
50k
60k
70k
80k
90k

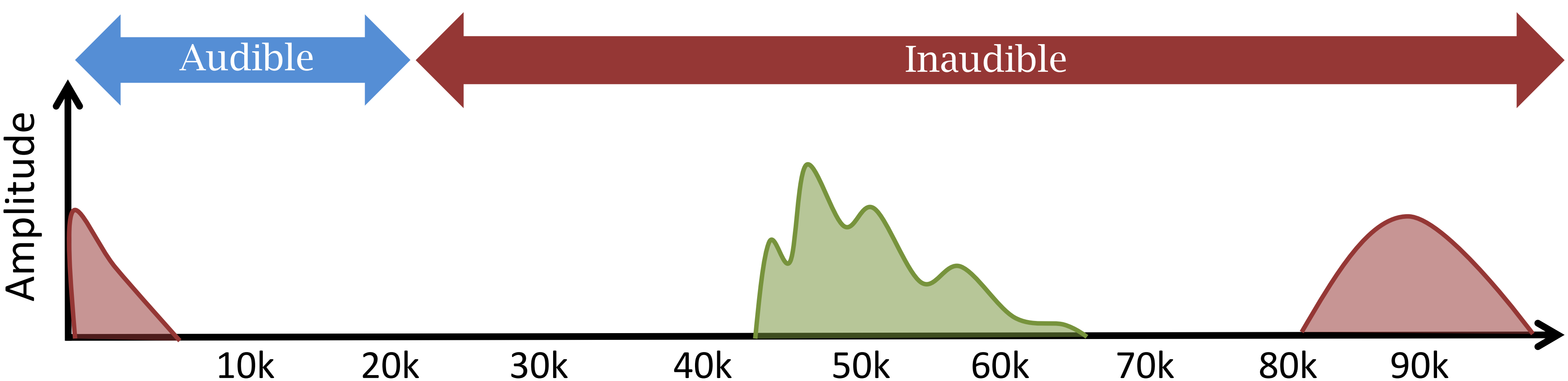


Amplitude
$V_{in} = v(t)sin(\omega_1 t)$
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
Speaker Nonlinearity
$a_1 V_{in} + a_2 V_{in}^2$
$a_1 v(t) sin(\omega_1 t) + a_2 \left(v^2(t) - v^2(t) \cos(2\omega_1 t) \right)$
Amplitude
10k
20k
30k
40k
50k
60k
70k
80k
90k


Amplitude
$V_{in} = v(t)\sin(\omega_1 t)$
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
Speaker Nonlinearity
$a_1 V_{in} + a_2 V_{in}^2$
Audible
Inaudible
Amplitude
10k
20k
30k
40k
50k
60k
70k
80k
90k



Why not modify V_{in} to cancel out audible leakage?



Nonlinearity
+
Channel memory
+
Power dependency
+
Thermal noise

$$a_1 V_{in} + a_2 V_{in}^2$$

Non-invertible channel

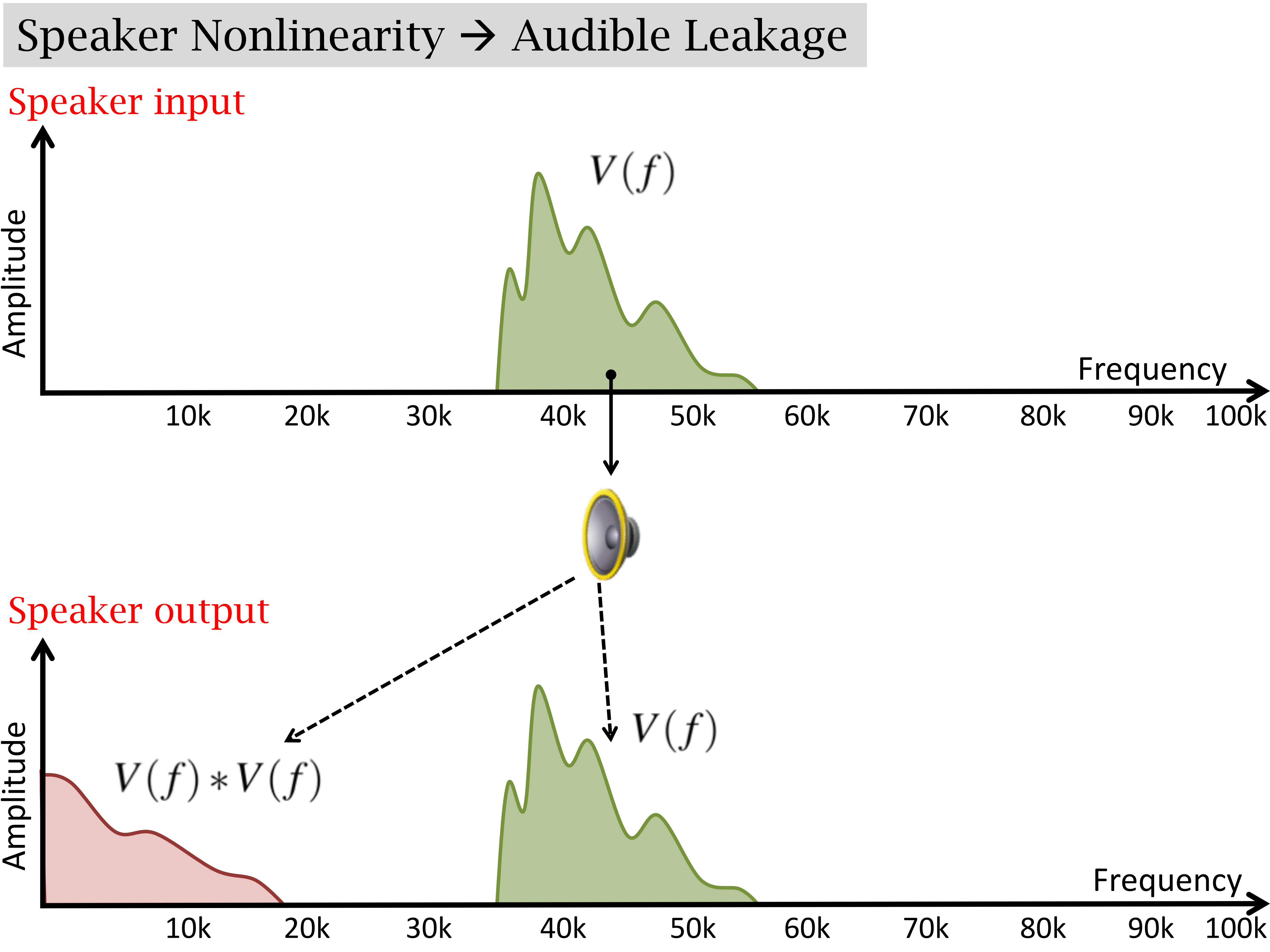
Nonlinearity
+
Channel memory
+
Power dependency
+
Thermal noise

$$a_1 V_{in} + a_2 V_{in}^2$$

Non-invertible channel

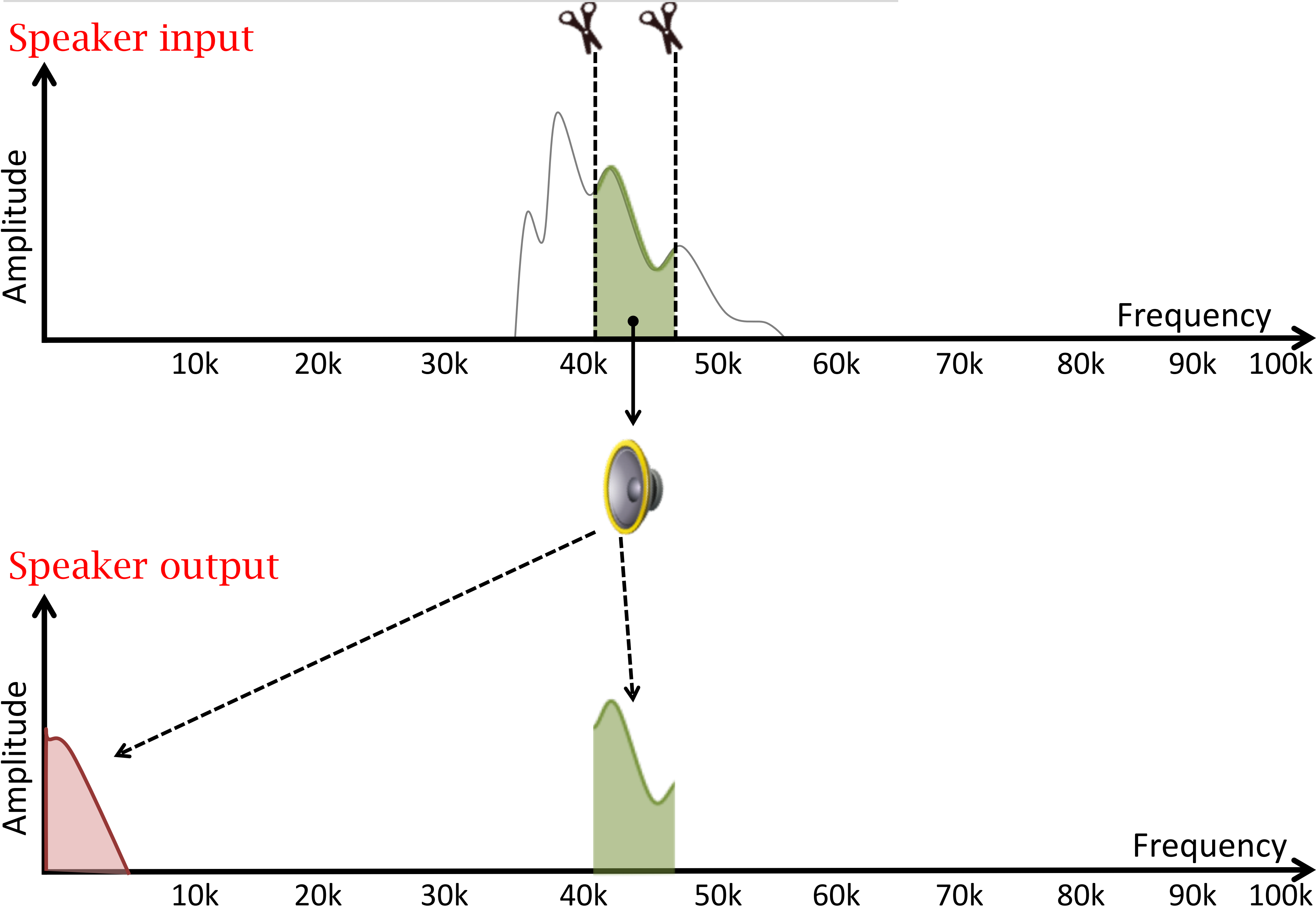
Speaker leakage is unavoidable…

… motivating “Leakage Optimization”

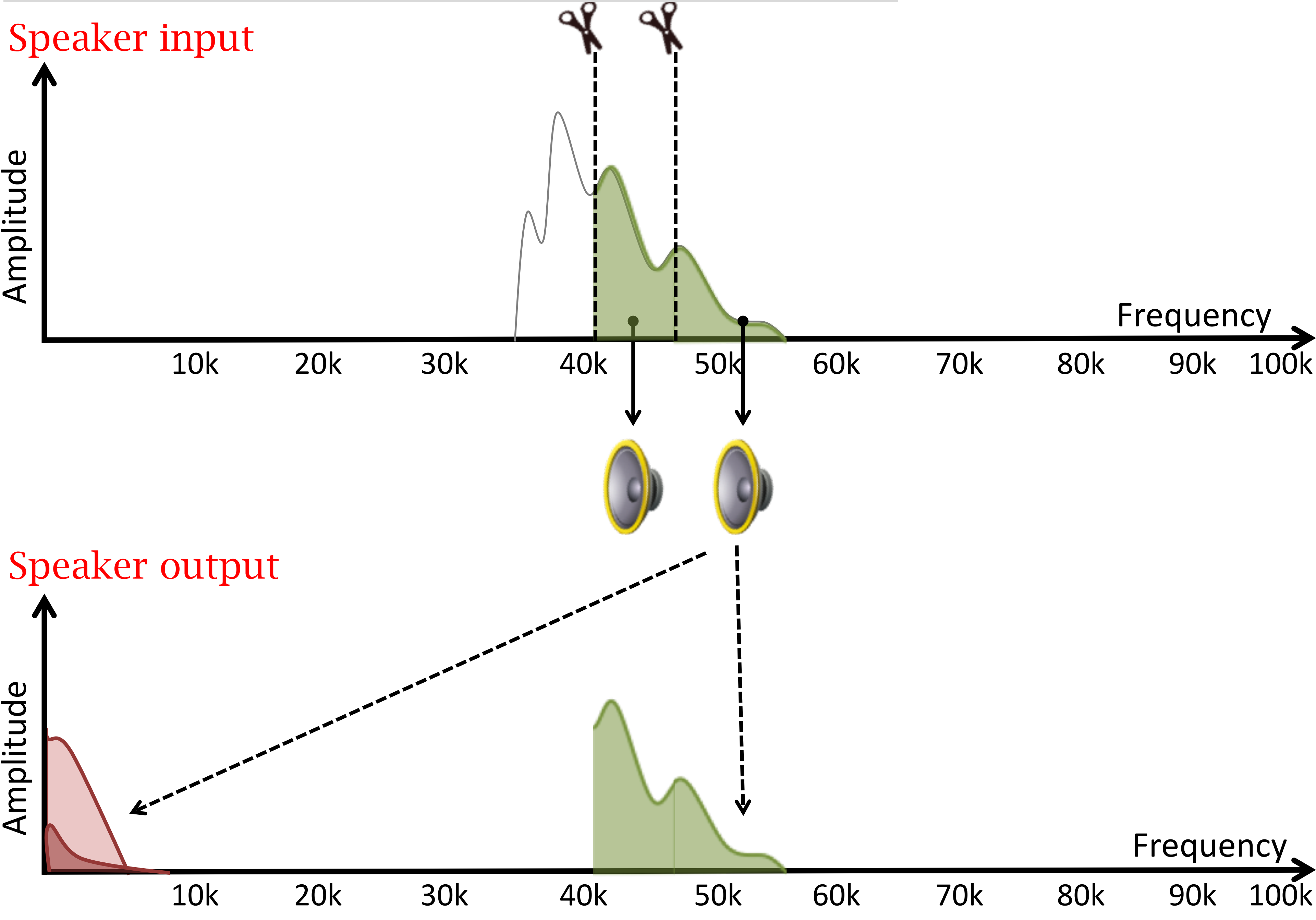


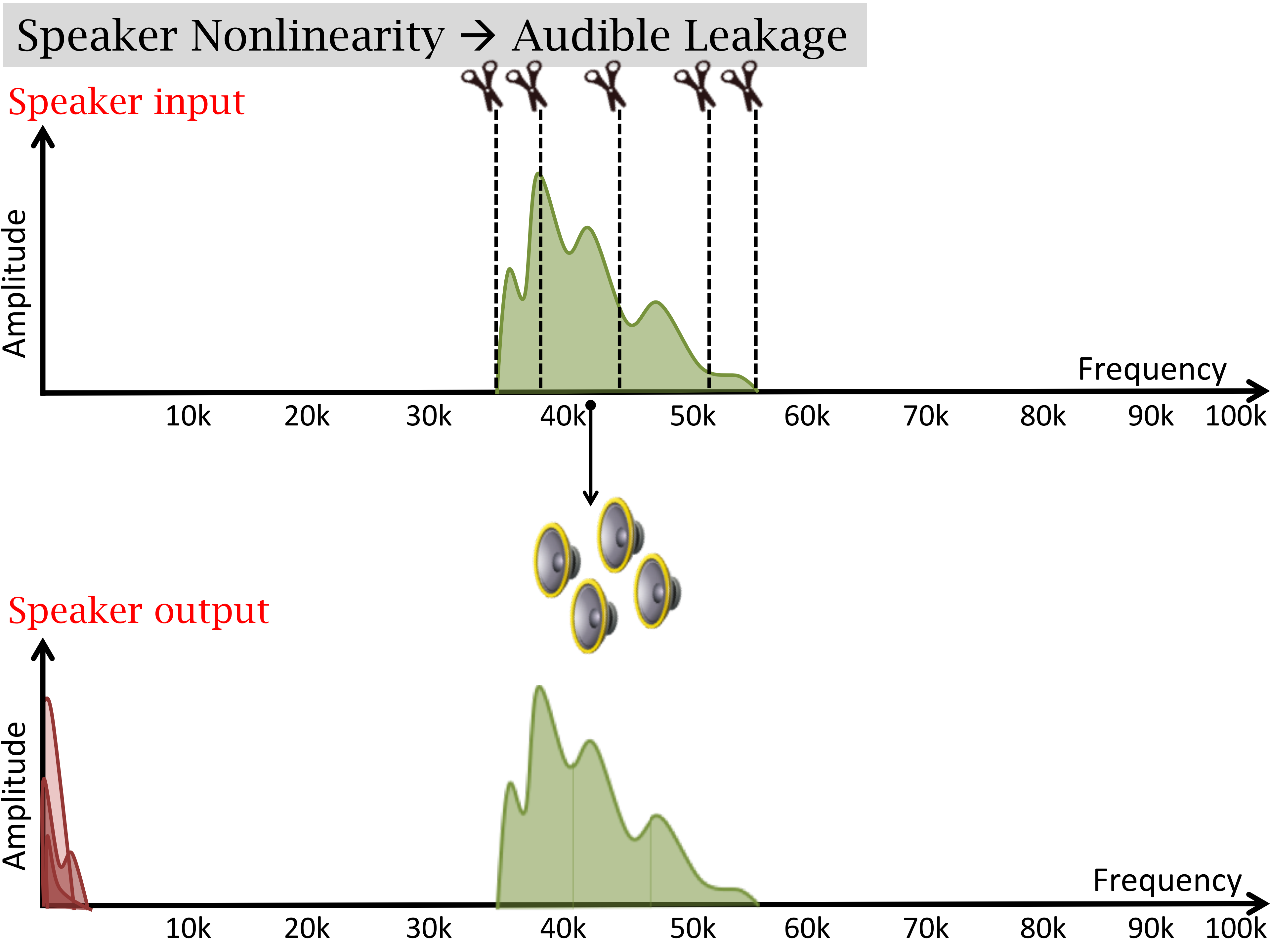
Speaker Nonlinearity → Audible Leakage
Speaker input
Amplitude
V(f)
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Speaker output
Amplitude
V(f) * V(f)
V(f)
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k

Speaker Nonlinearity → Audible Leakage



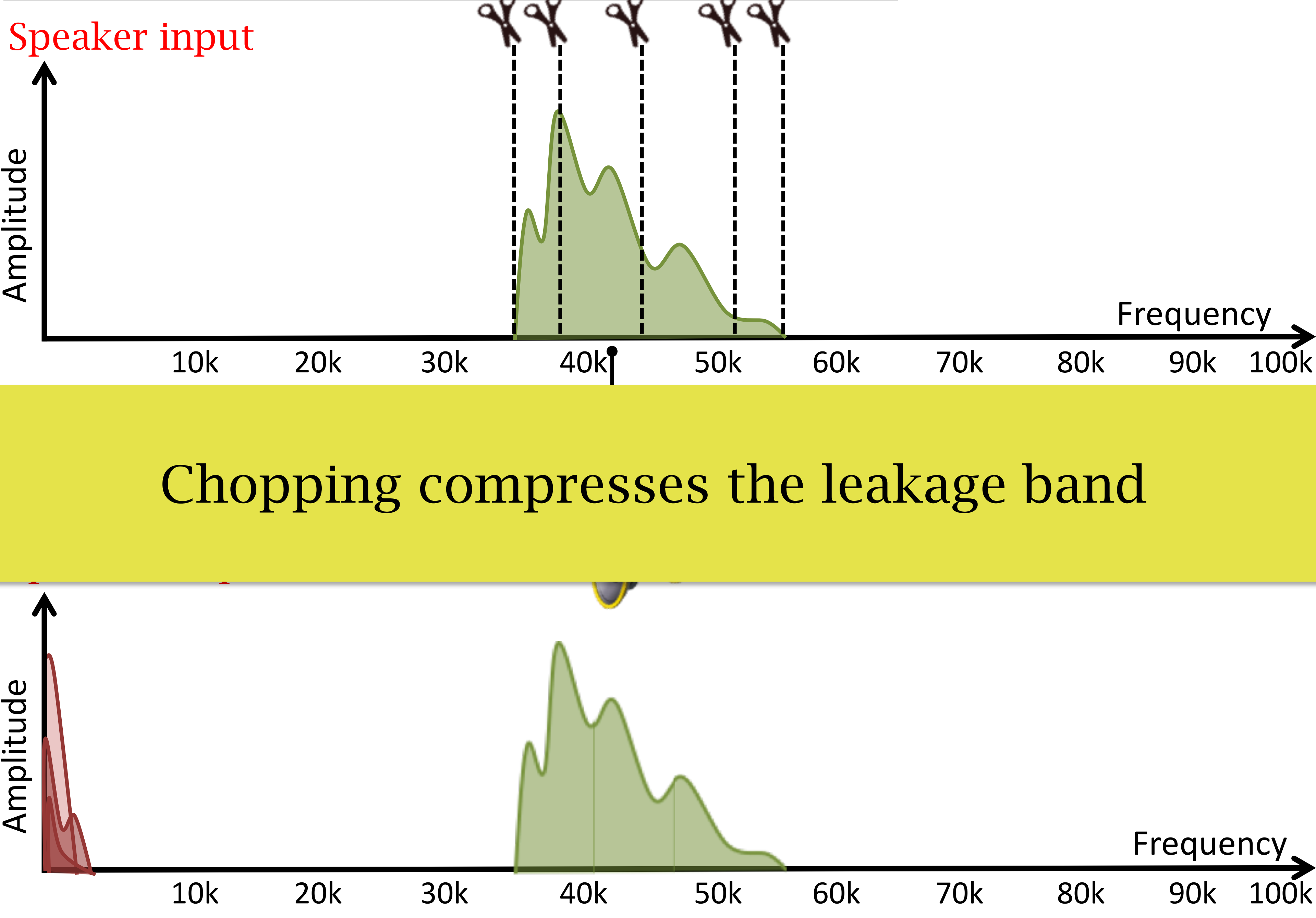
Speaker Nonlinearity → Audible Leakage



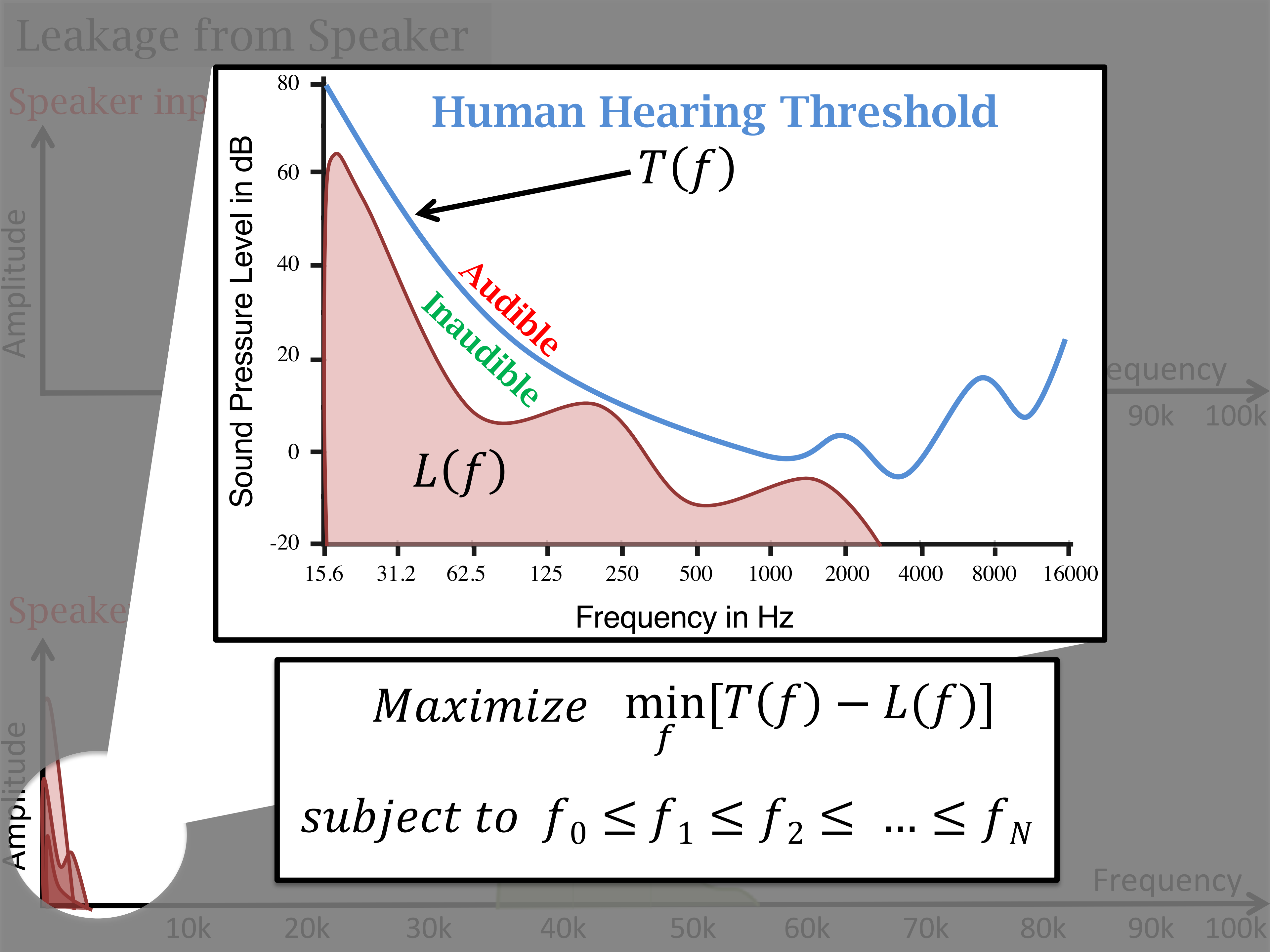


Speaker Nonlinearity → Audible Leakage
Speaker input
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Speaker output
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k

Speaker Nonlinearity → Audible Leakage

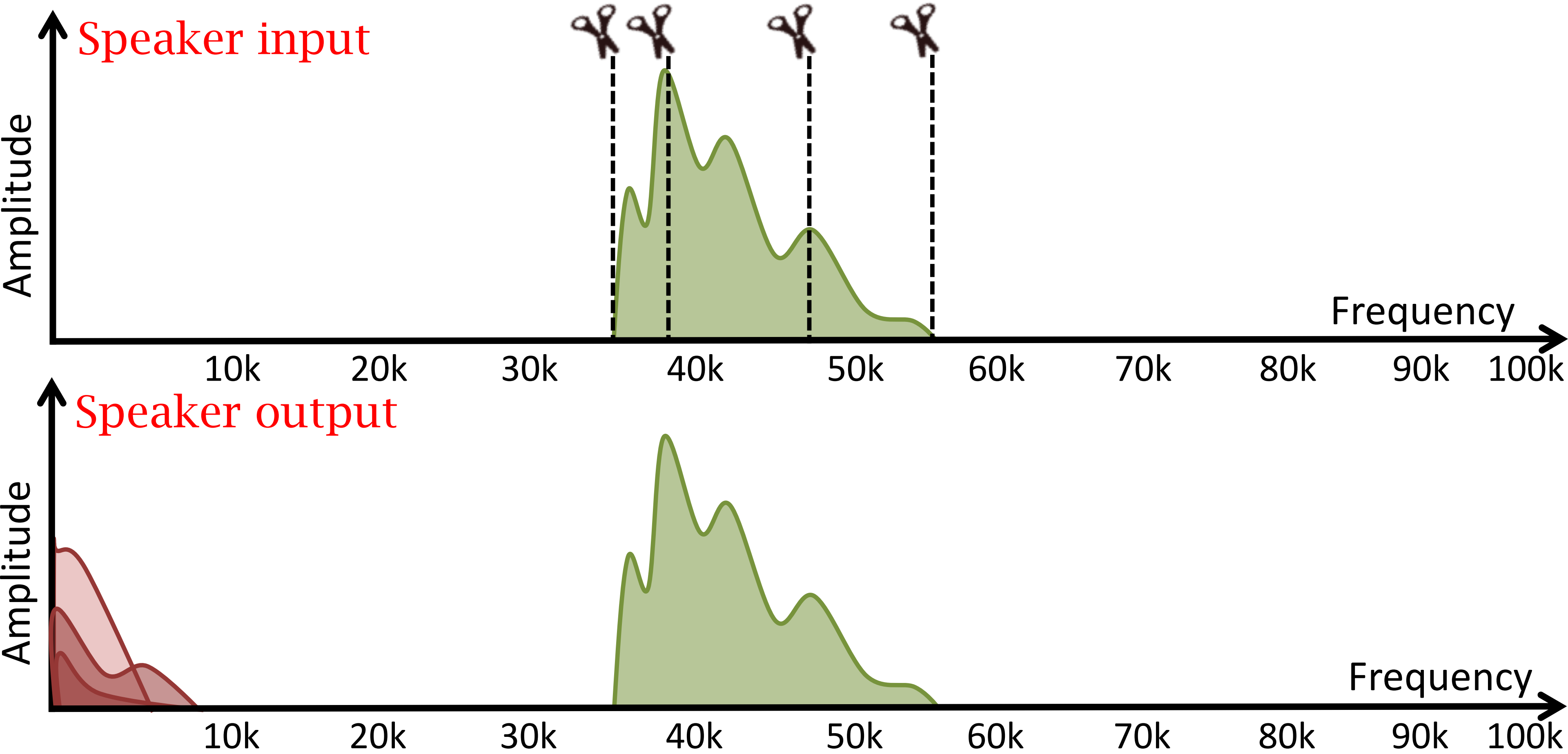


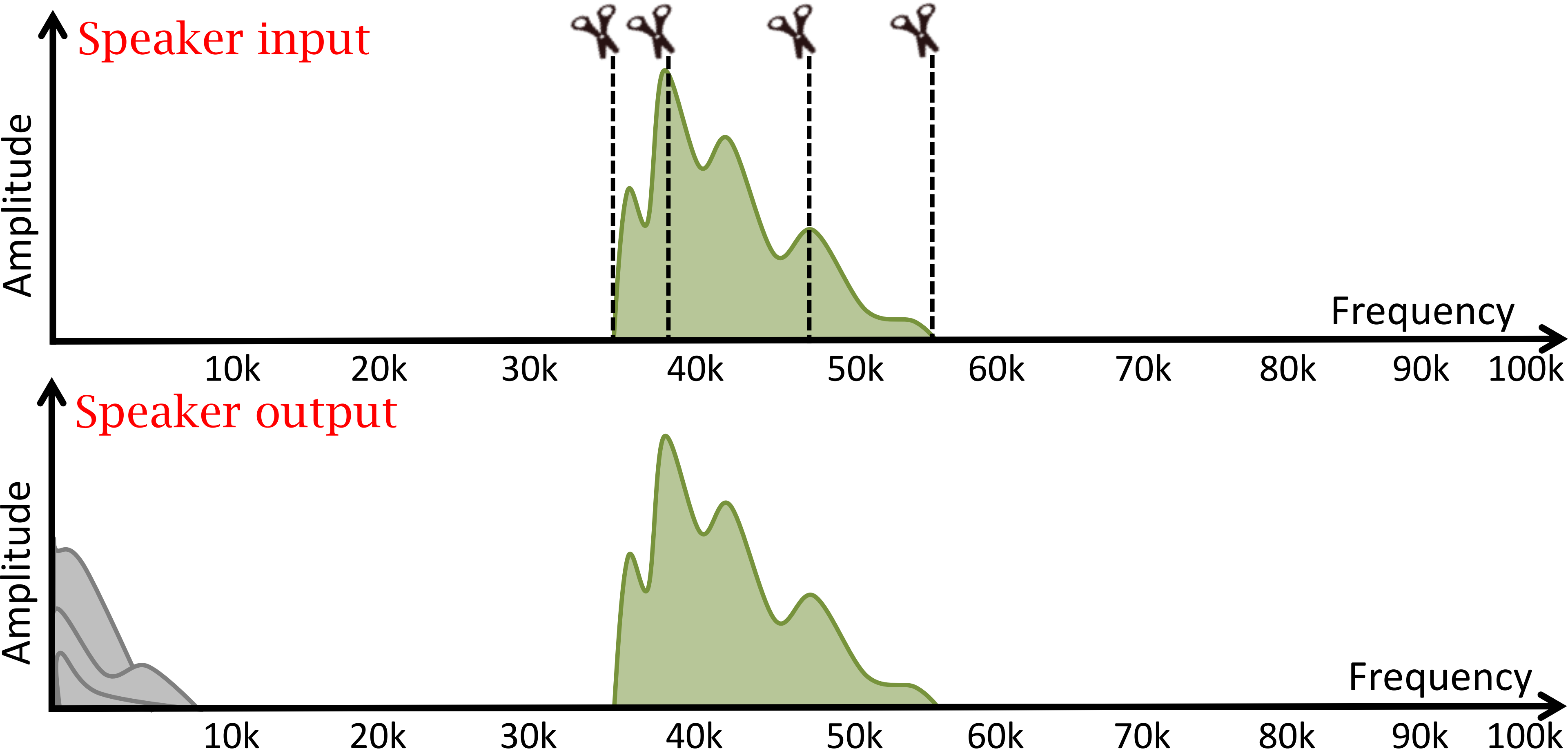
Chopping compresses the leakage band



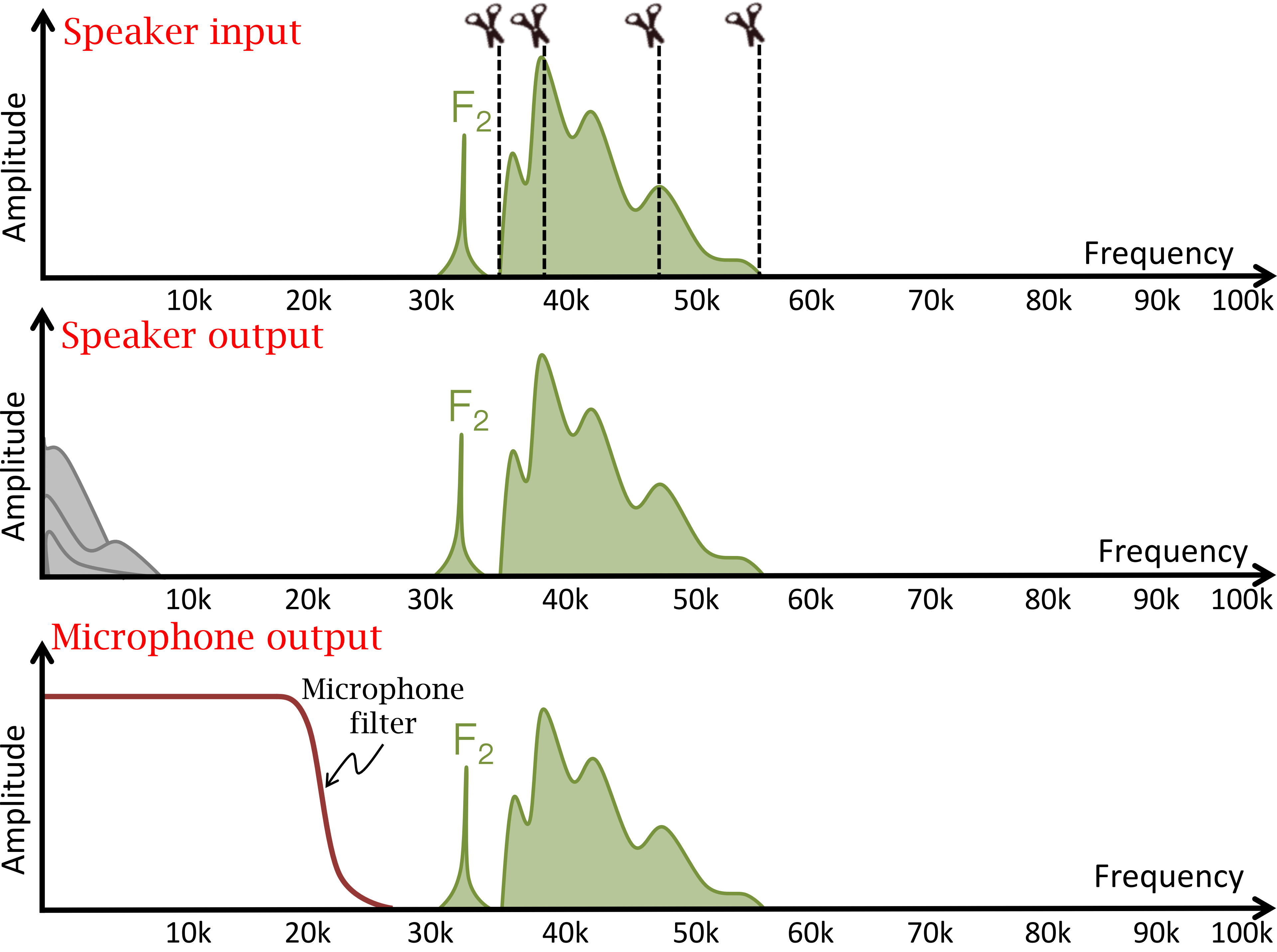
$$Maximize \quad \min_{f}[T(f) - L(f)]$$

$$subject\ to \quad f_0 \leq f_1 \leq f_2 \leq \ \ldots \leq f_N$$


Speaker input
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Speaker output
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k



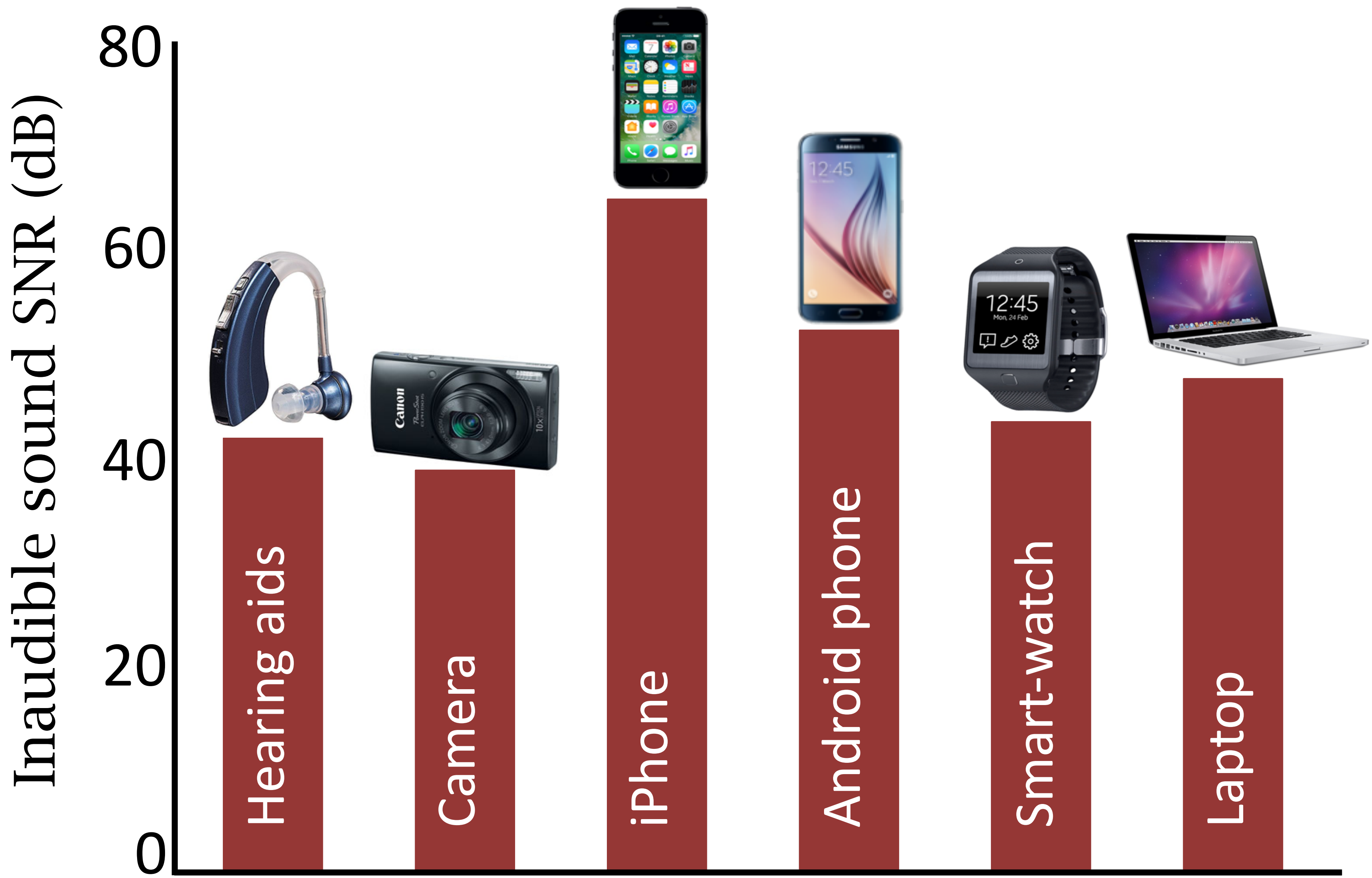
Speaker input
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Speaker output
Amplitude
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k



Speaker input
Amplitude
F2
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Speaker output
Amplitude
F2
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k
Microphone output
Amplitude
Microphone filter
F2
Frequency
10k
20k
30k
40k
50k
60k
70k
80k
90k
100k

Inaudible voice commands: Long range

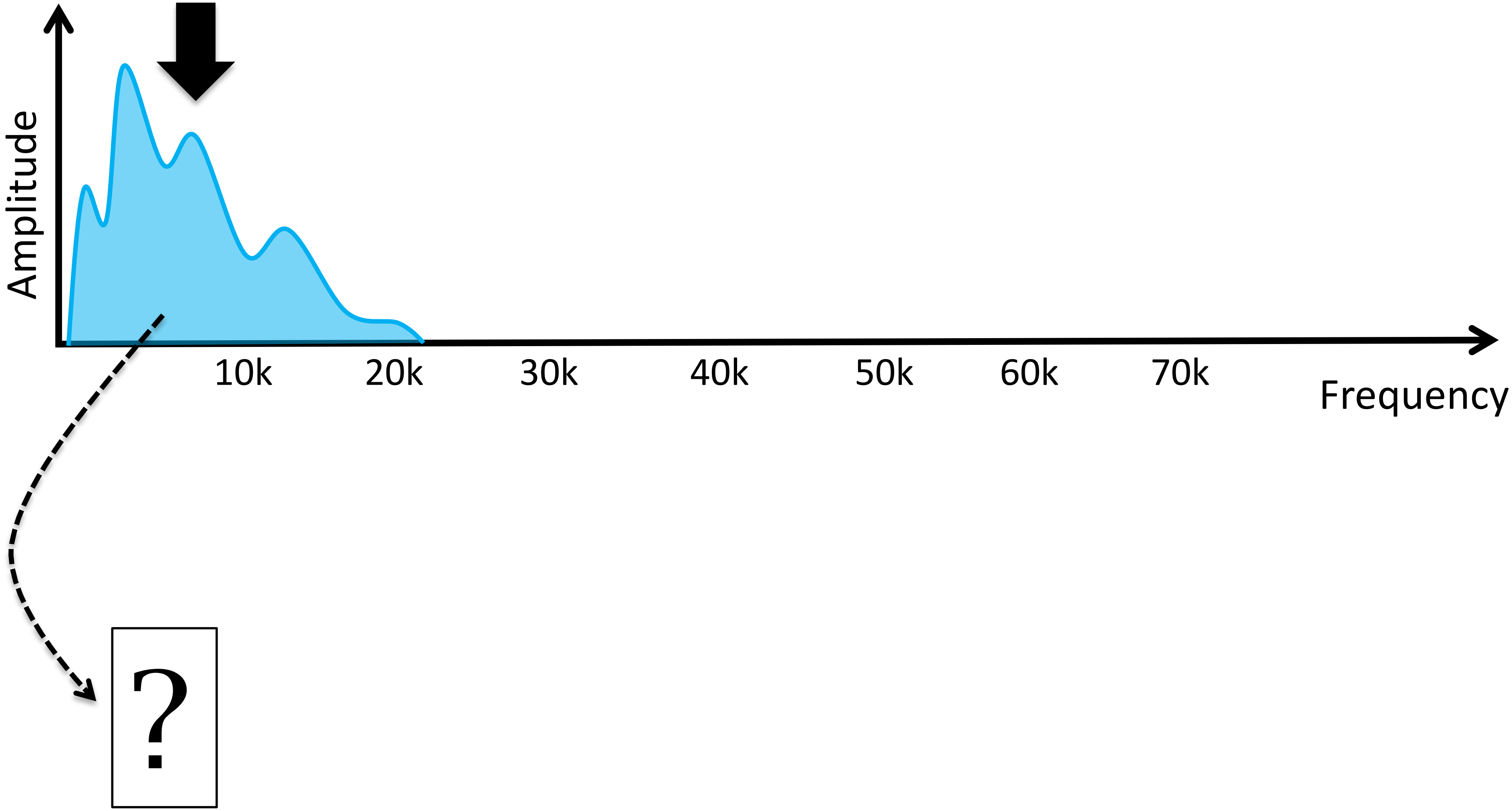
Applies to all microphone-enabled devices



Defending Inaudible Attacks

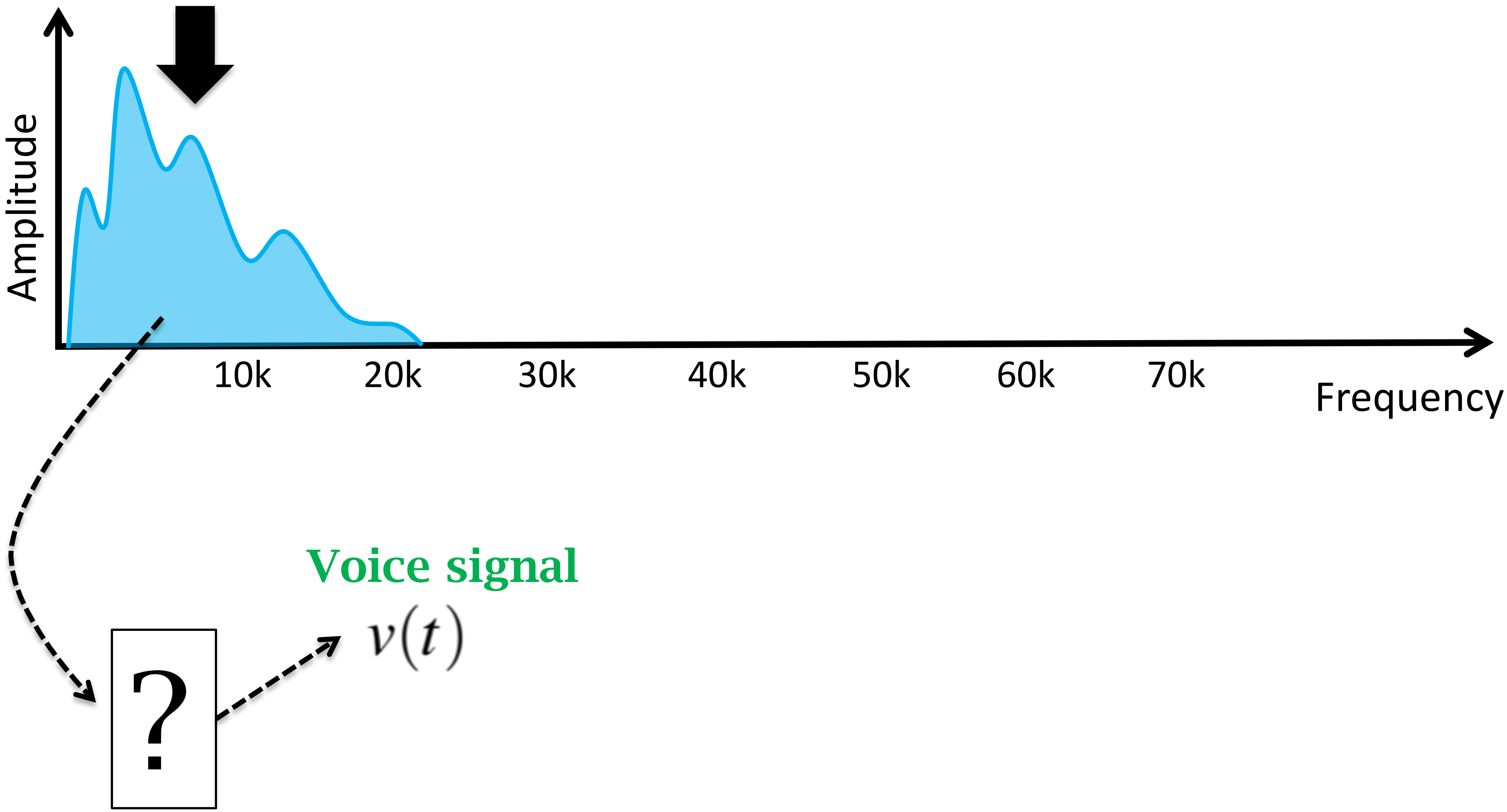
Core Question:

Is this a "non-linear signal" or normally recorded signal



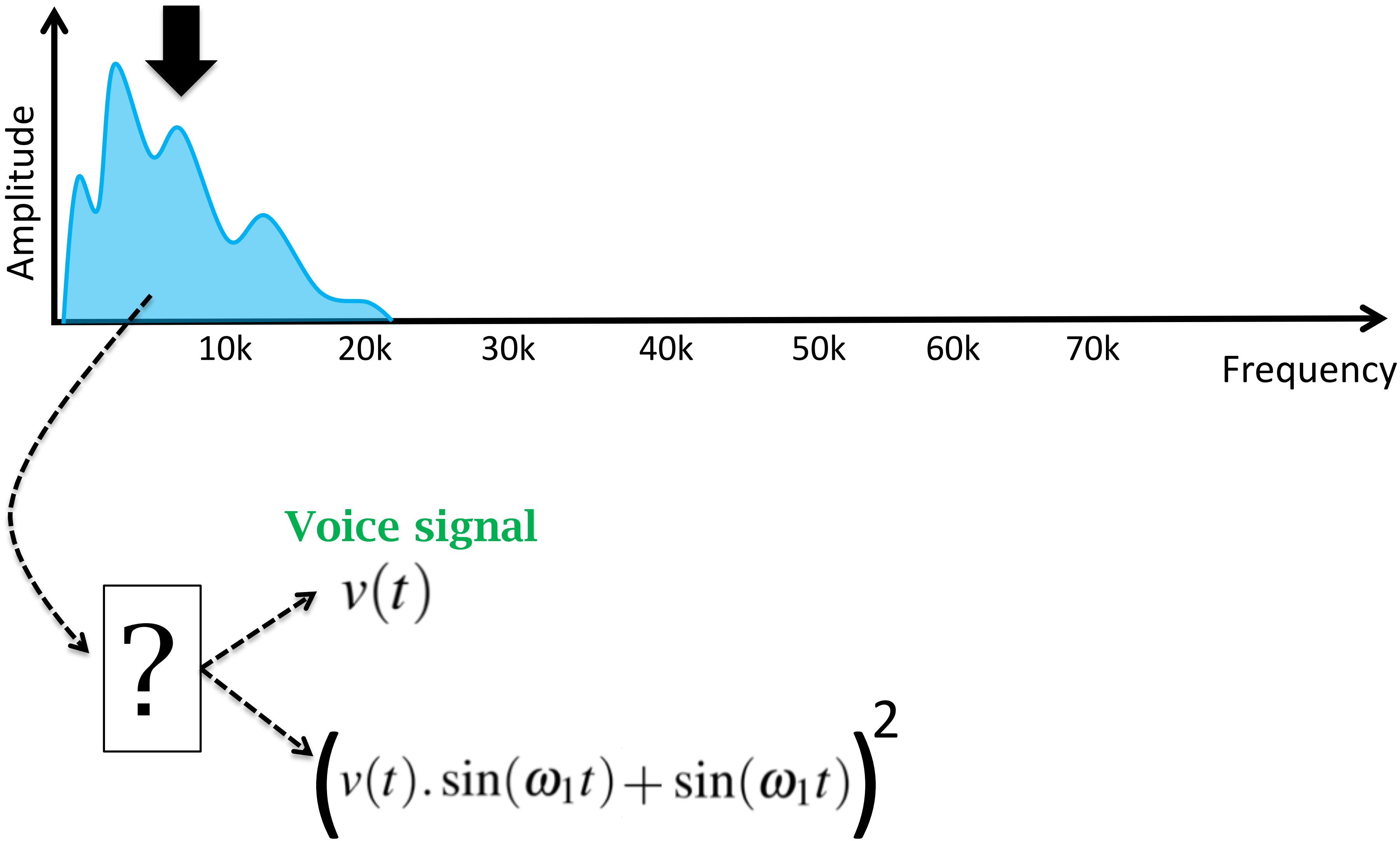
Core Question:

Is this a “non-linear signal” or normally recorded signal



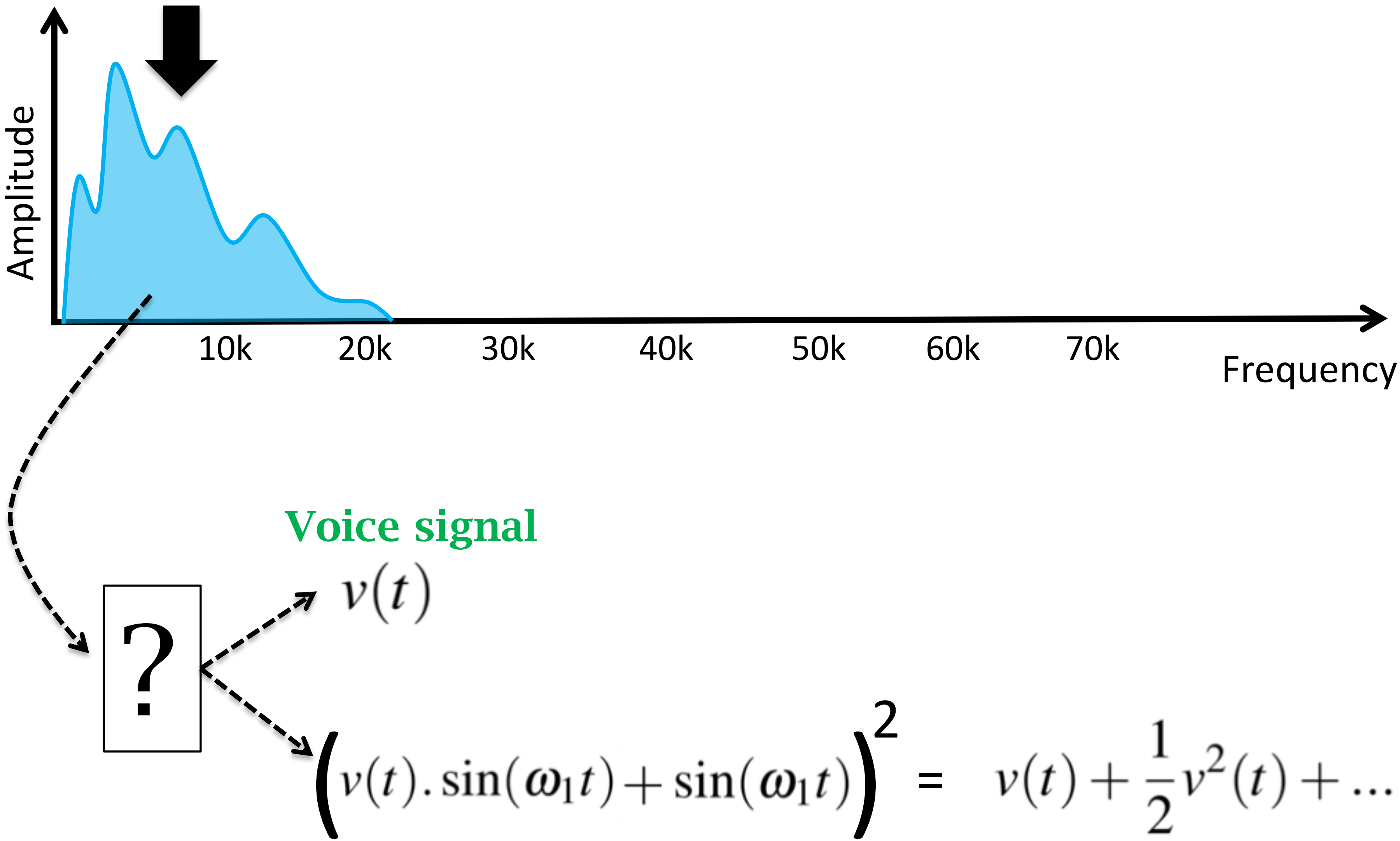
Core Question:

Is this a “non-linear signal” or normally recorded signal



Core Question:

Is this a “non-linear signal” or normally recorded signal

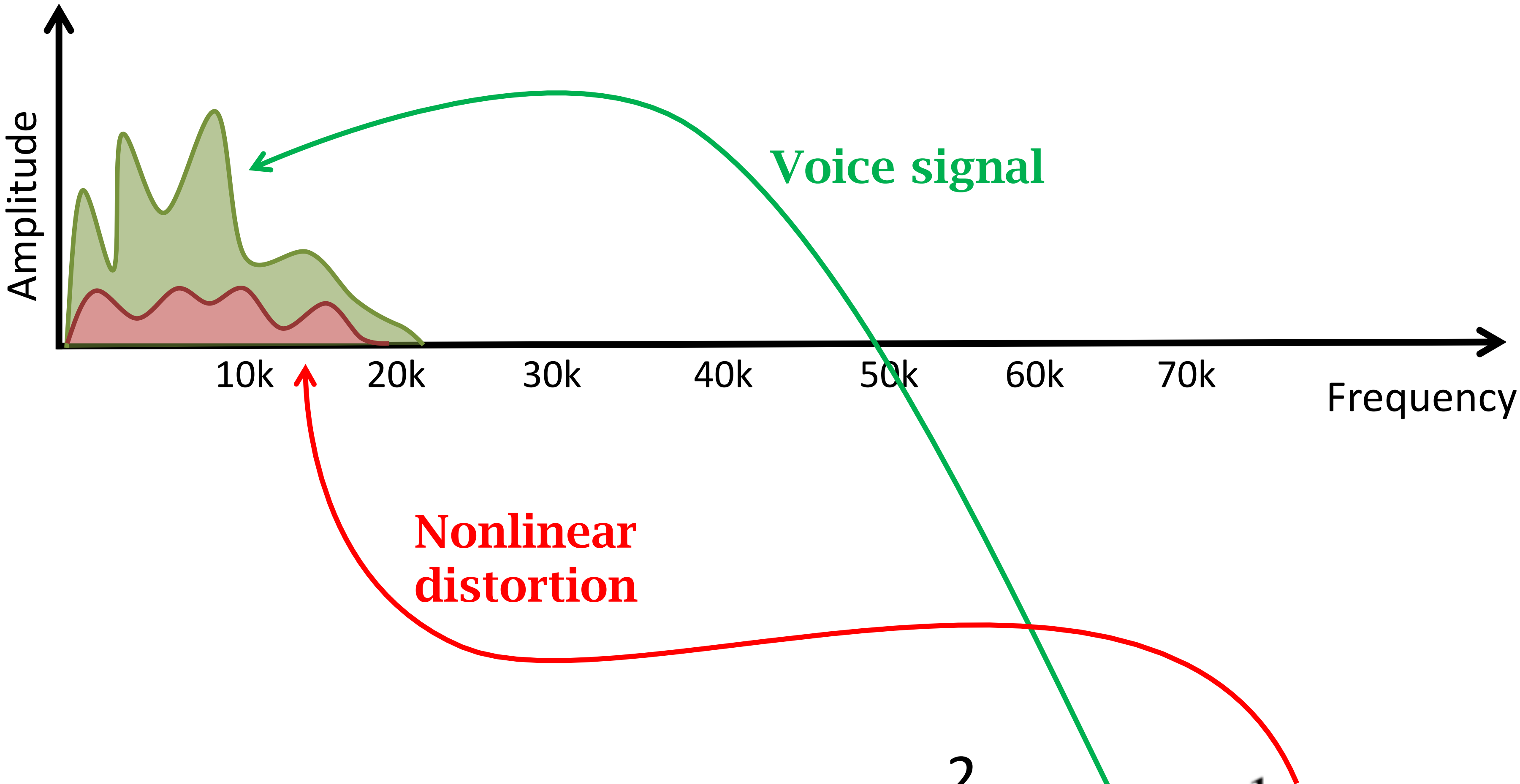


$$v(t)$$

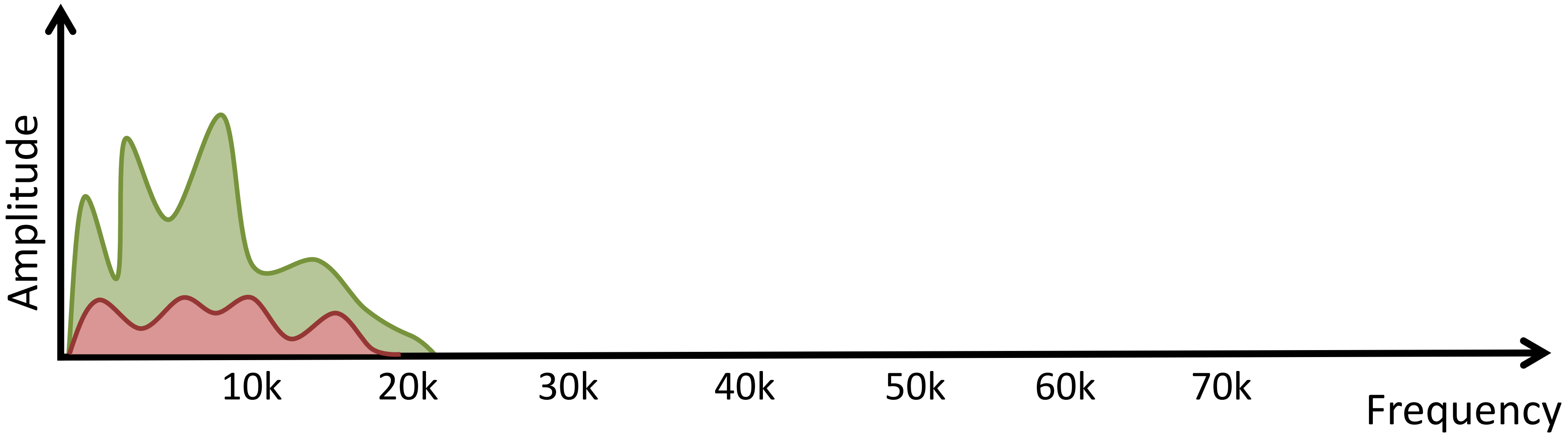
$$\left(v(t).\sin(\omega_1 t)+\sin(\omega_1 t)\right)^2 = v(t)+\frac{1}{2}v^2(t)+\ldots$$

Core Question:

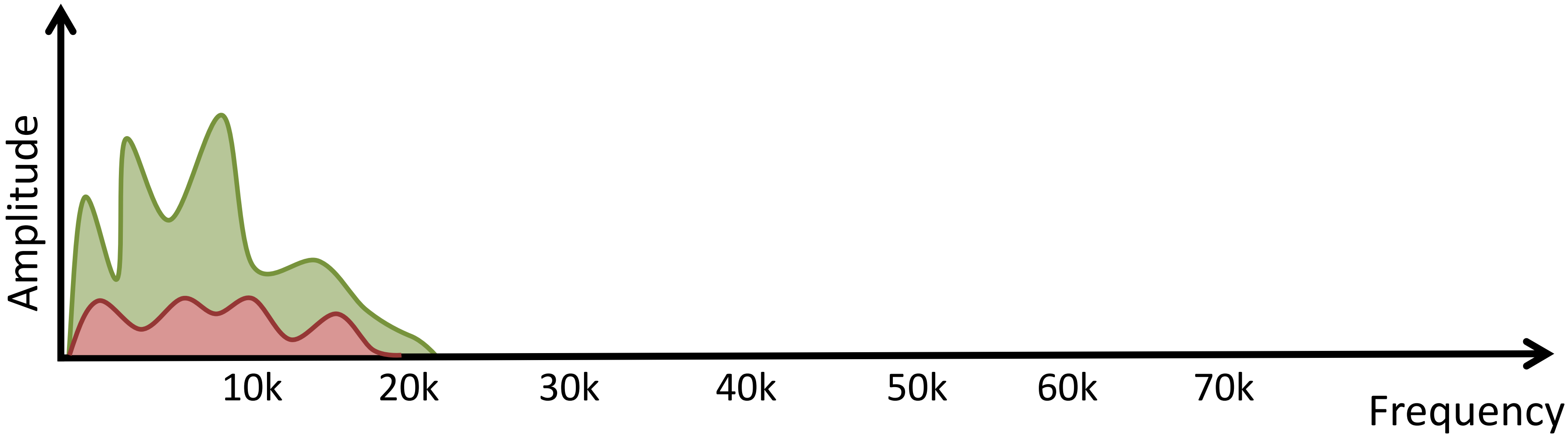
Is this a “non-linear signal” or normally recorded signal



$$\left(v(t).\sin(\omega_1 t)+\sin(\omega_1 t)\right)^2 = v(t)+\frac{1}{2}v^2(t)+\ldots$$



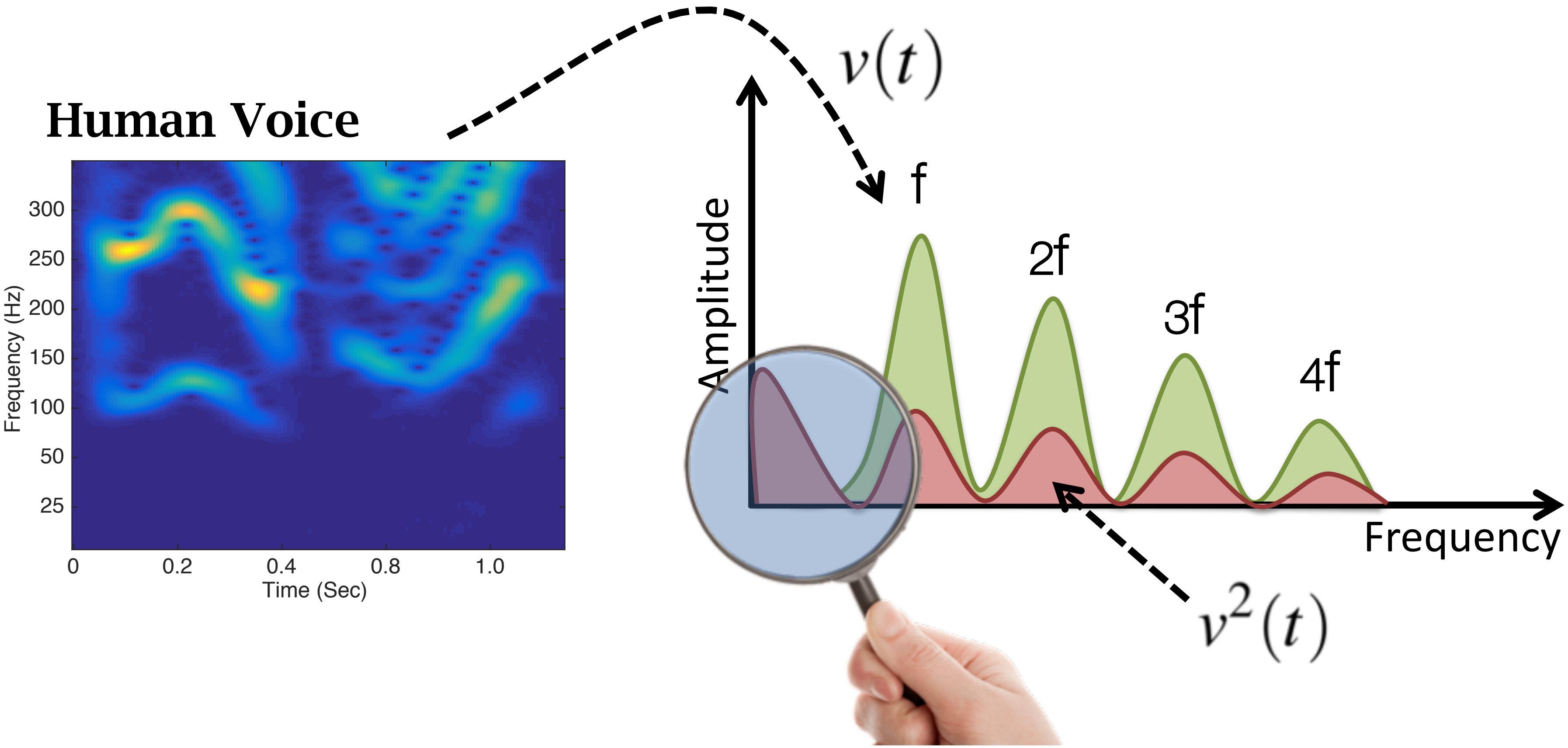
Do not know the original signal …
Hence, difficult to differentiate "non-linear signal"
from a noisy but legitimate sound signal



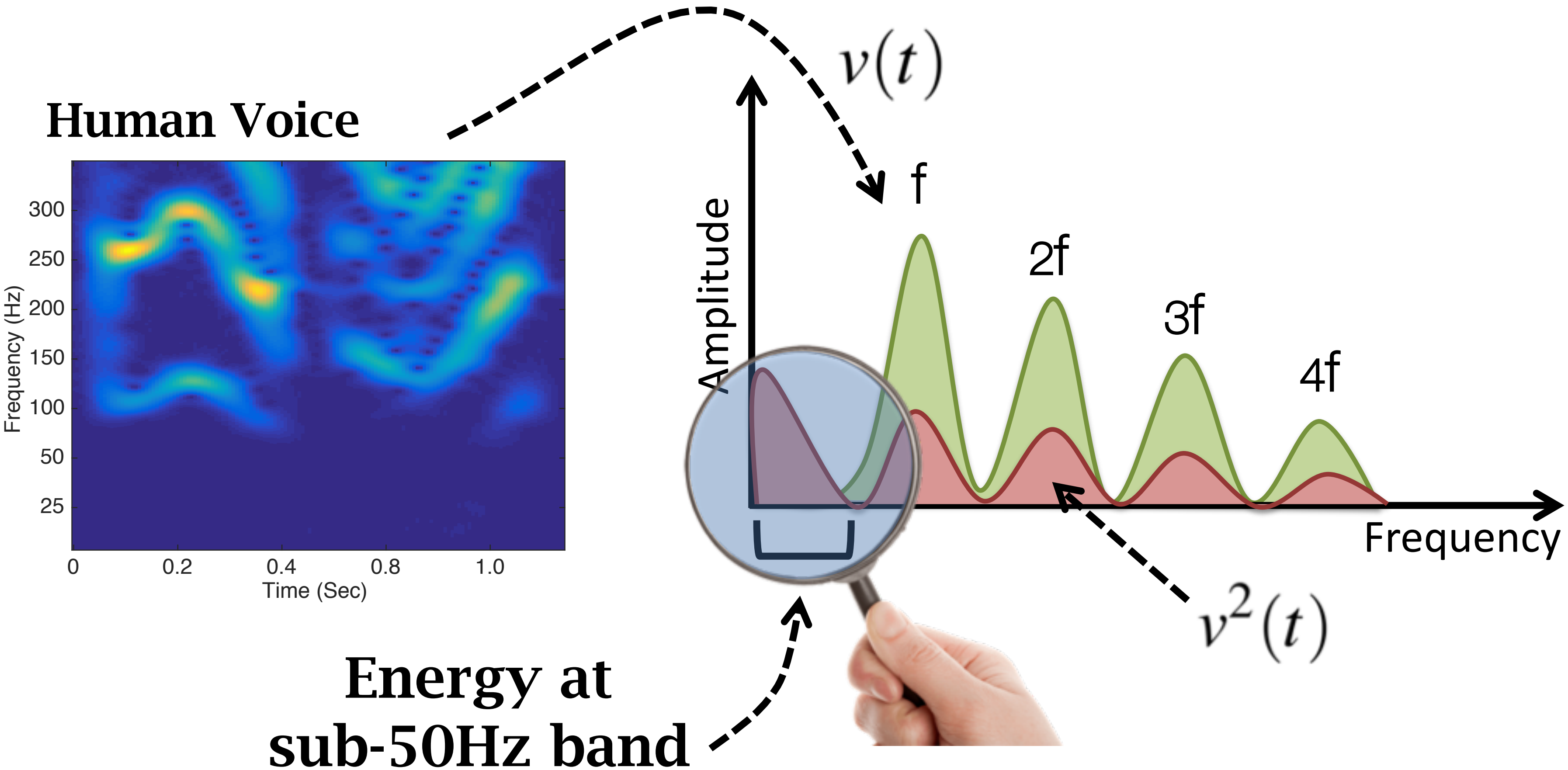
Do not know the original signal …
Hence, difficult to differentiate “non-linear signal” from a noisy but legitimate sound signal

However, voice signals present opportunities …

Opportunity: Harmonic Structure in Voice > 50 Hz

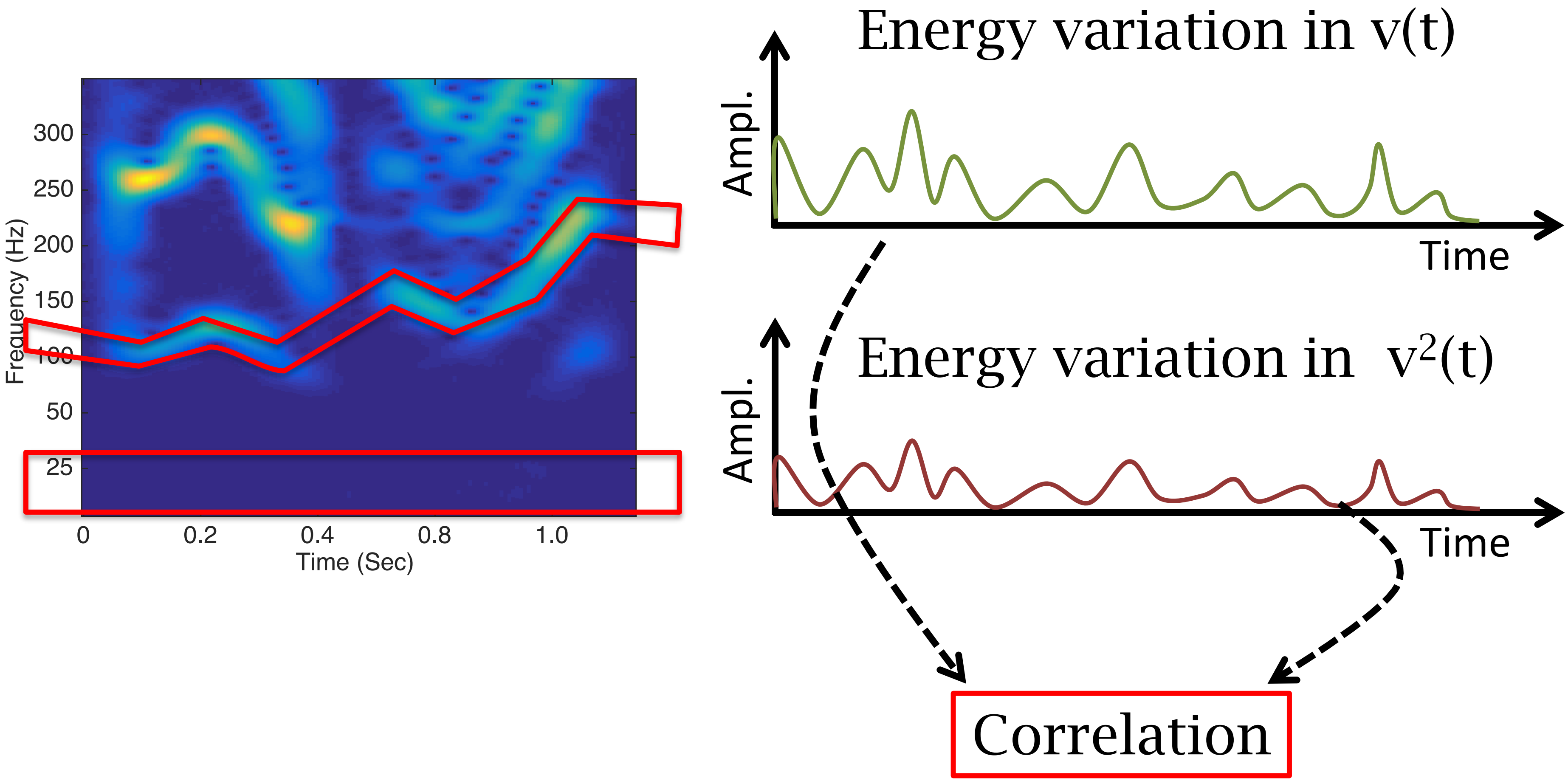


Opportunity: Harmonic Structure in Voice > 50 Hz

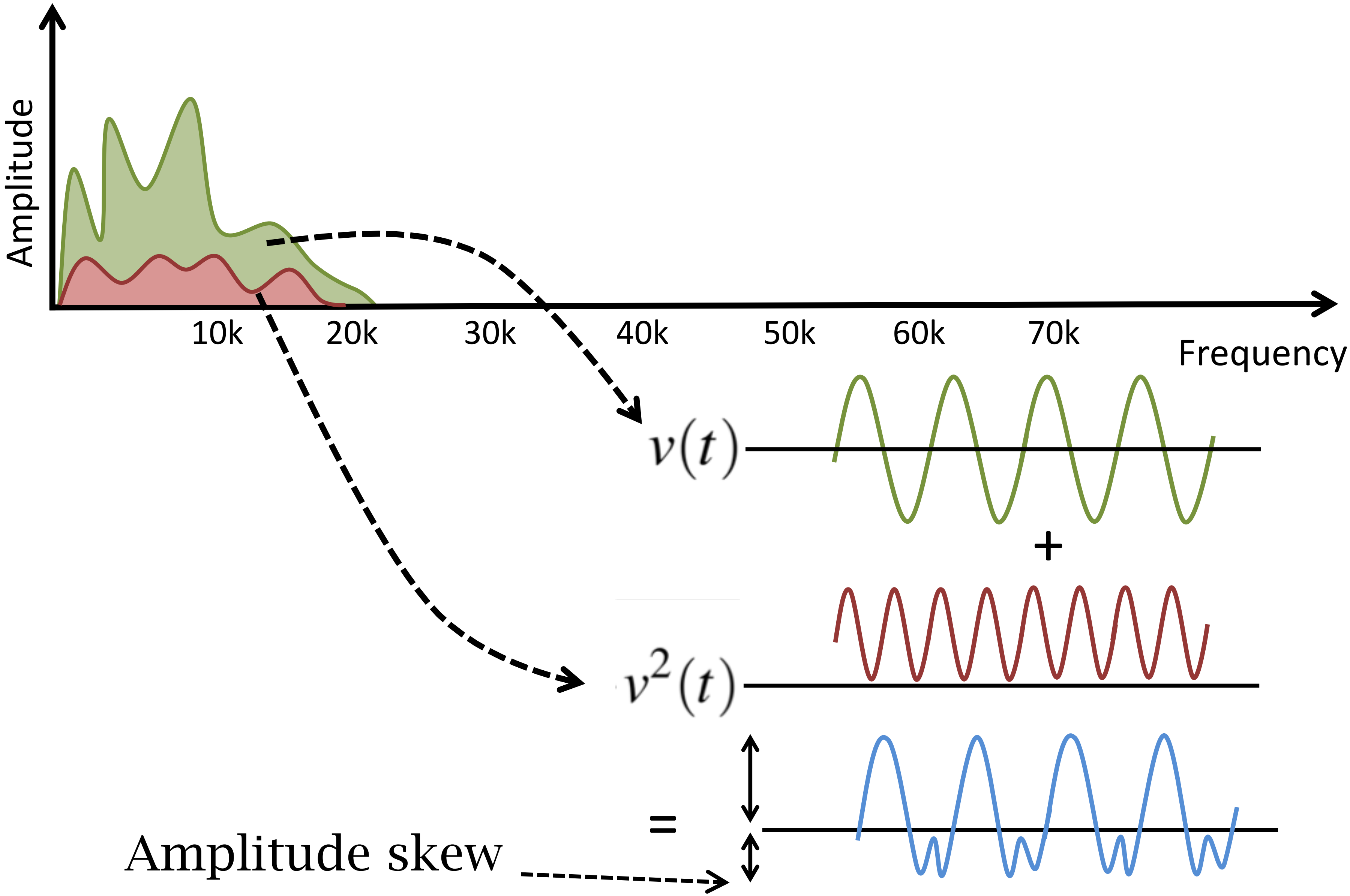


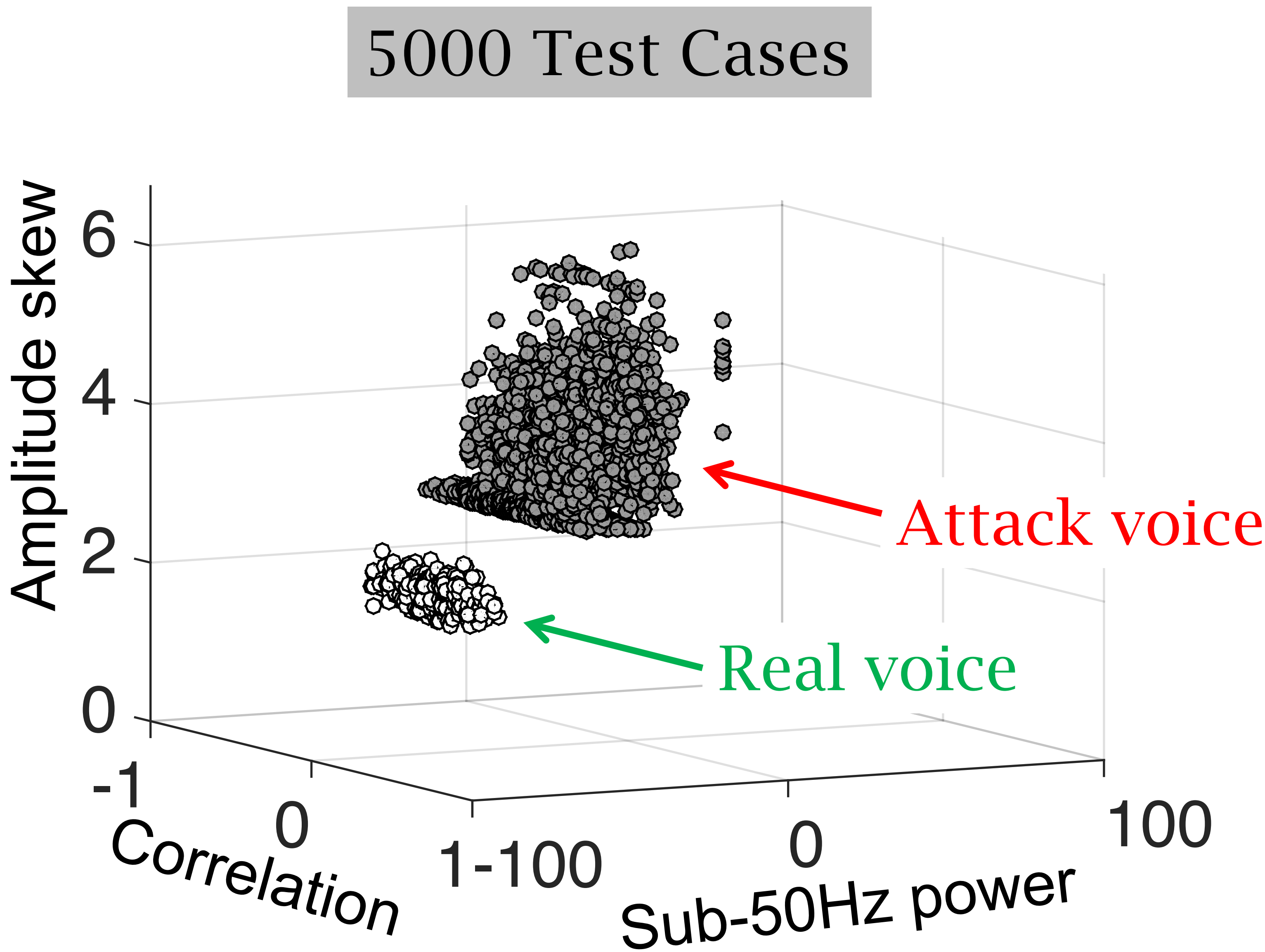
Can sub-50 Hz signals reveal non-linearity ?

Sub-50 Hz signals preserves harmonics …



Leakage signal component is (+)ve amplitude





5000 Test Cases
Amplitude skew
0
2
4
6
Attack voice
Real voice
-1
0
1
Correlation
-100
0
100
Sub-50Hz power

To summarize…

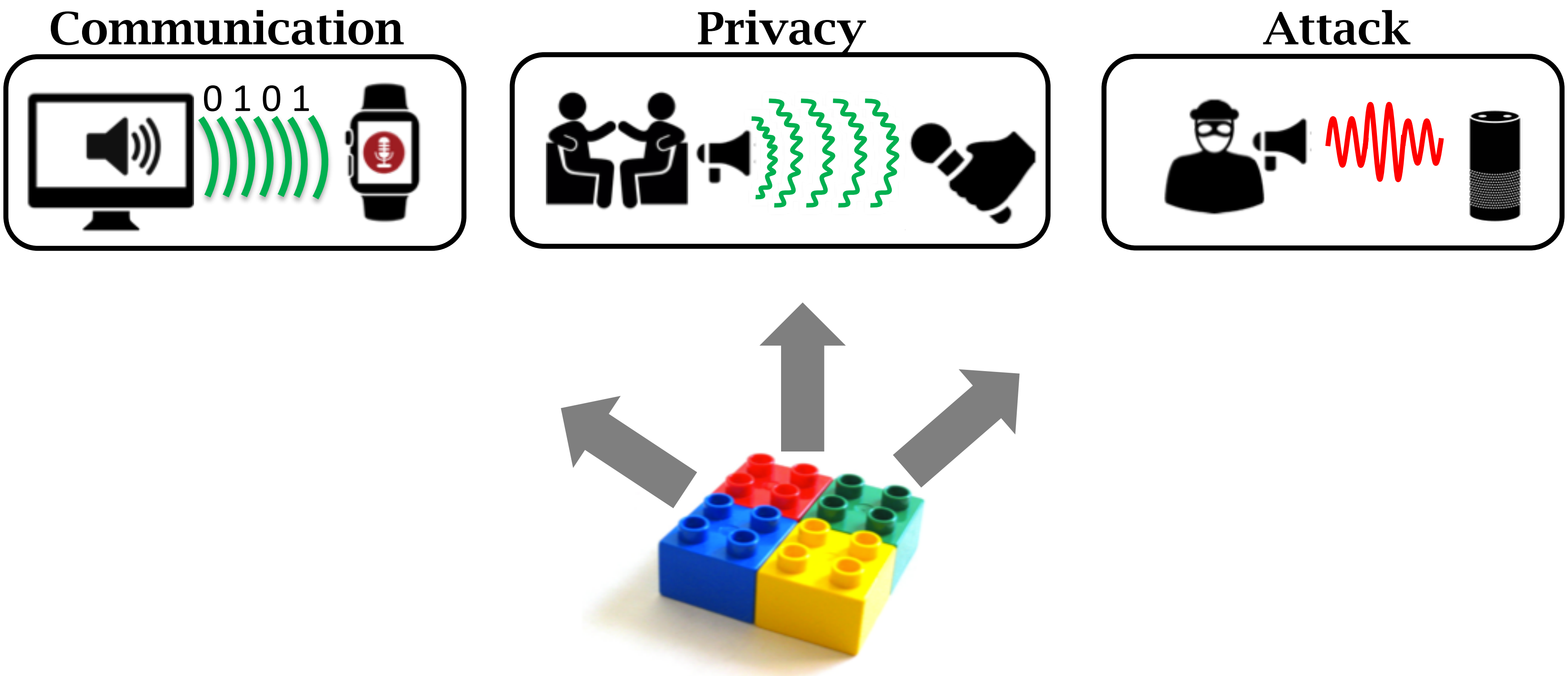
Inaudible Acoustics is a new primitive …

To summarize…

Inaudible Acoustics is a new primitive …

that makes **inaudible** ultrasound **audible** to microphones

To summarize…



Inaudible Acoustics is a new primitive …

that makes **inaudible** ultrasound **audible** to microphones

Underpinning a wide range of IoAT applications …