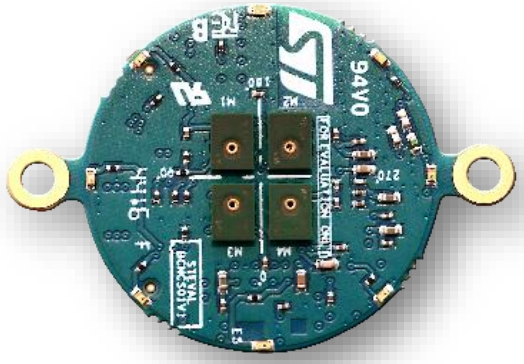
MATLAB EXPO 2018

Designing and Testing Voice Interfaces through Microphone Array Modeling, Audio Prototyping, and Text Analytics

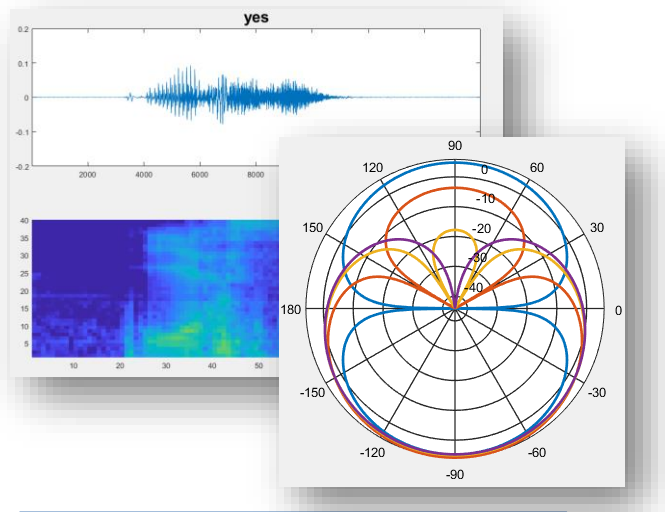
Vidya Viswanathan

What Device Is This?

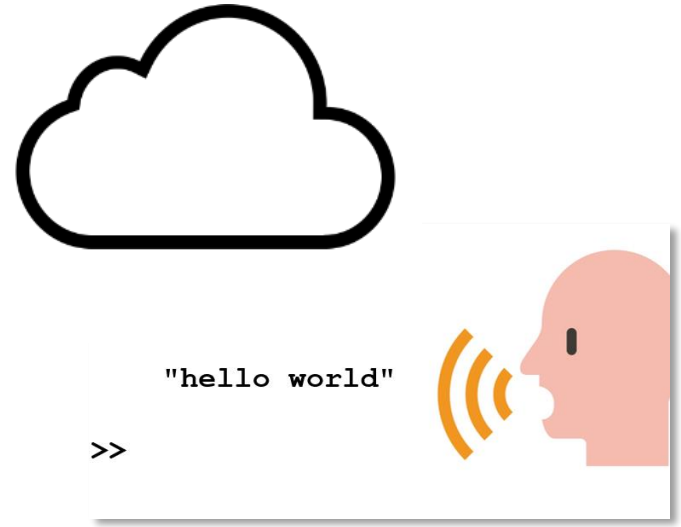
Overview of the System



Audio Acquisition

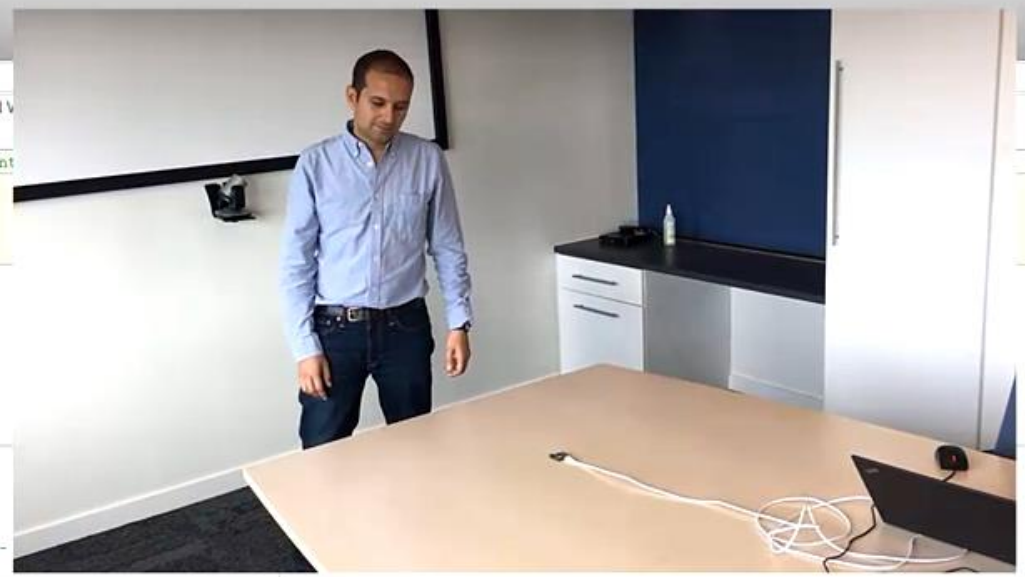


Speech Preprocessing



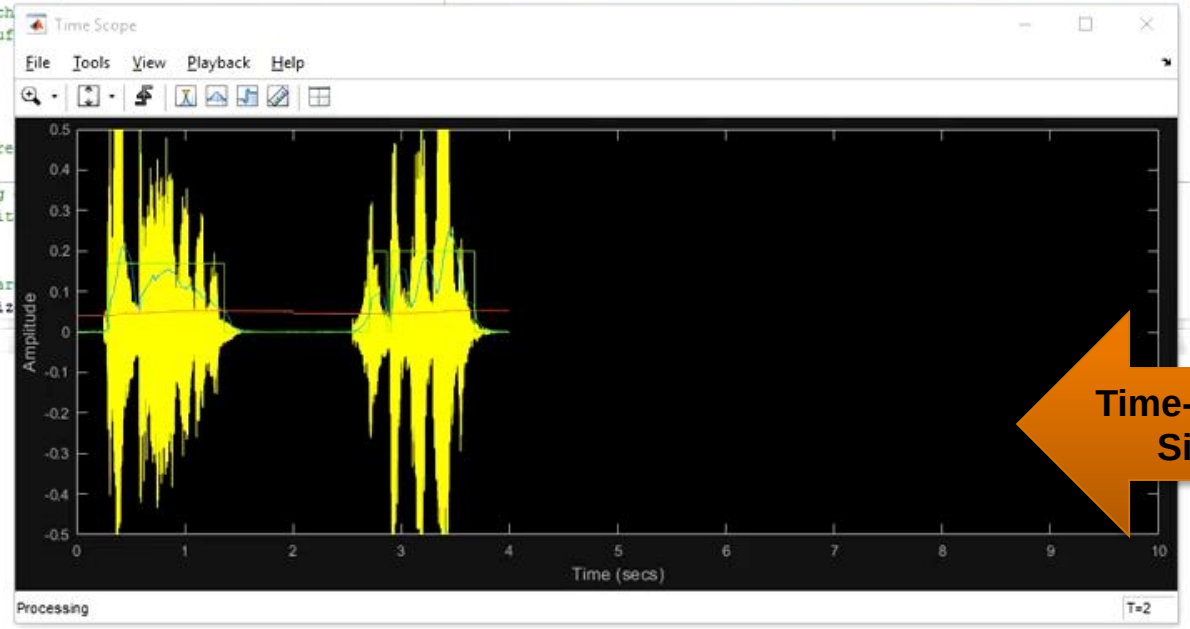
Post Processing

Beam Steering



Detected Speech

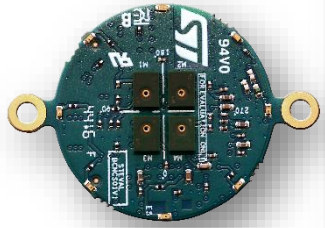
good afternoon everybody
do you speak Matlab

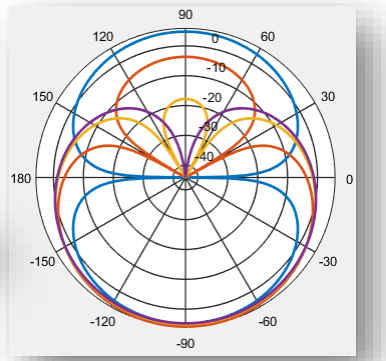


Time-Domain Signal

How Can I…

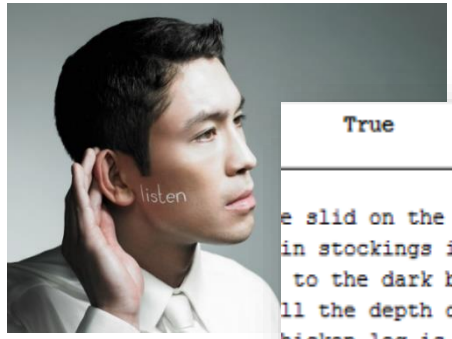
1. **Design a voice interface?**





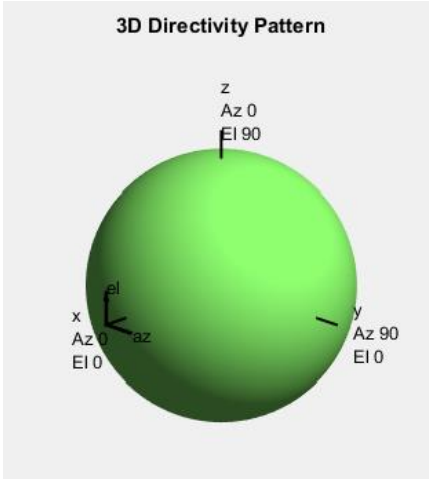
2. Validate if my voice interface can work in real-life scenarios?

3. Test the performance of my system?

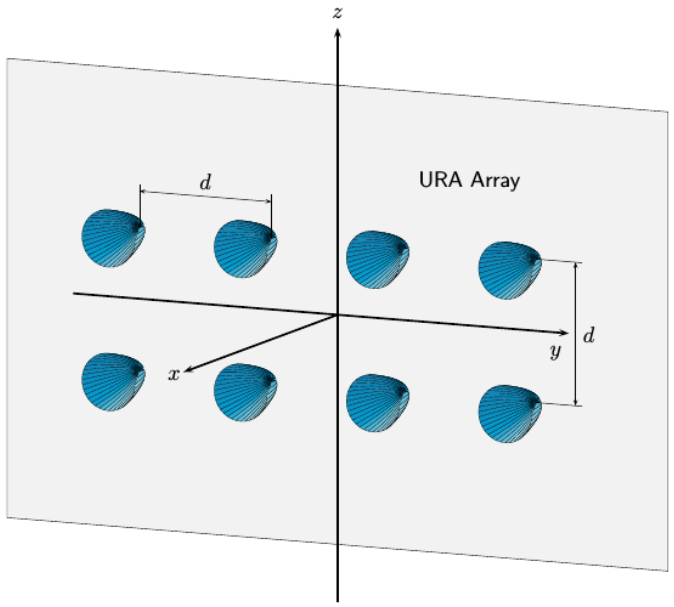


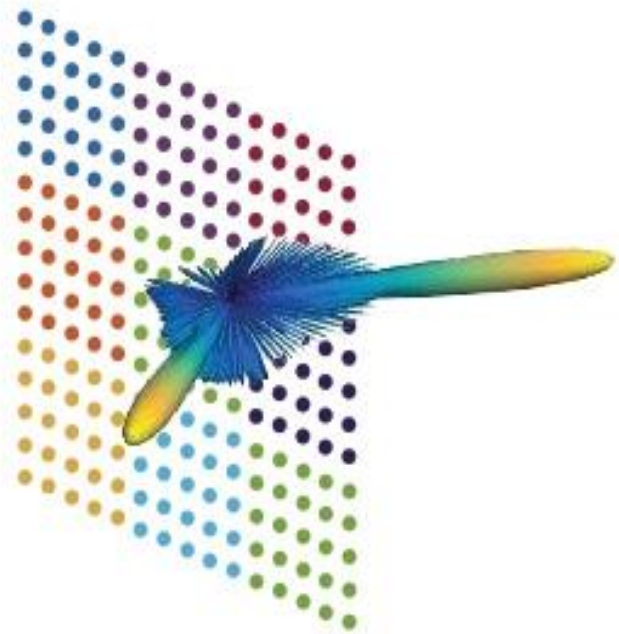
True	Guessed
e slid on the smooth planks"	"the birth canal we sl
in stockings is hard to sell"	"a large size in stock
to the dark blue background"	"blue the sheet to the
ll the depth of a well"	"it's easy to tell the
hicken leg is a rare dish"	"these days a chicken
served in round bowls"	"rice is often served

Choice of Microphones



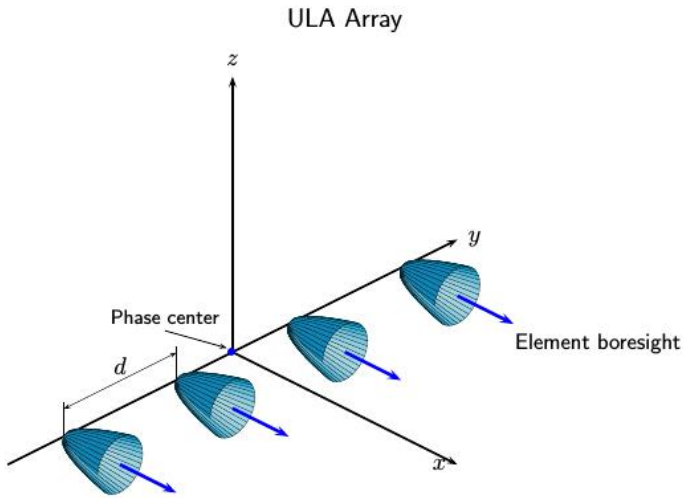
- Single Microphone
  - Fixed directivity in a given direction
  - Noise cancellation is non-trivial



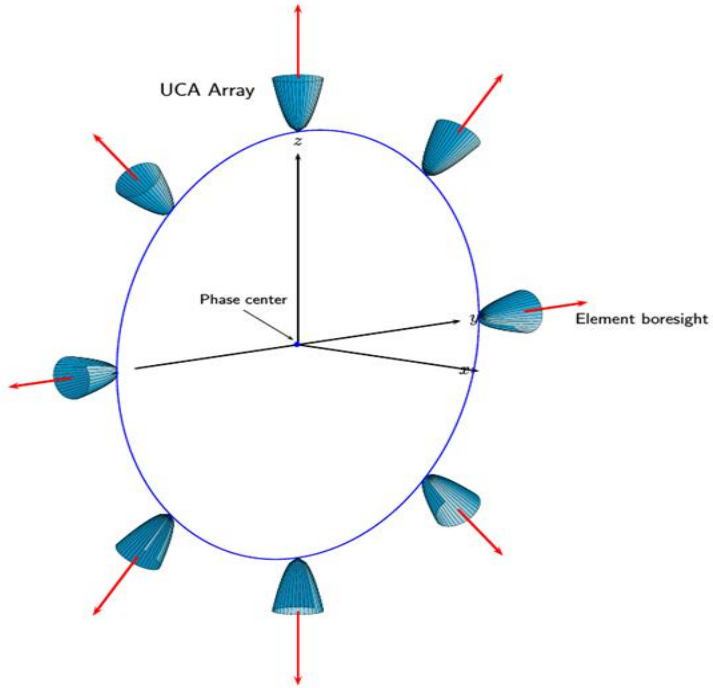


- Array of Microphones
  - Can control the directivity for a specific direction
  - Noise cancellation is easier

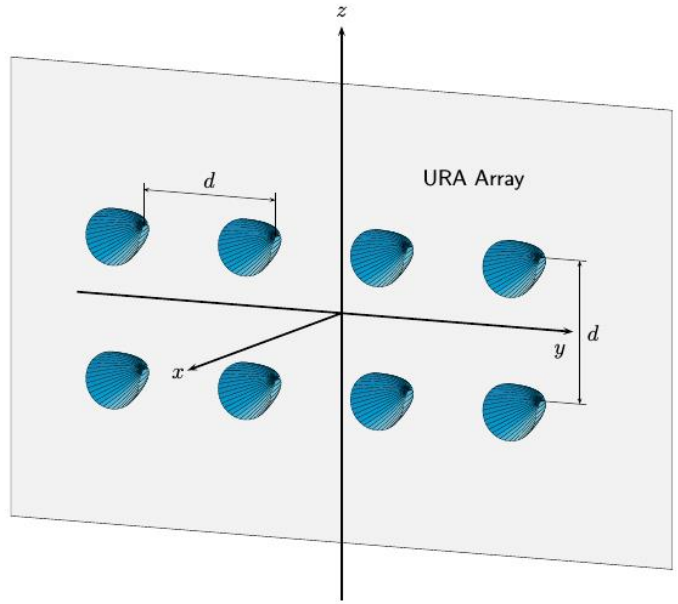
Microphone Array Geometries



Uniform Linear Array



Uniform Circular Array



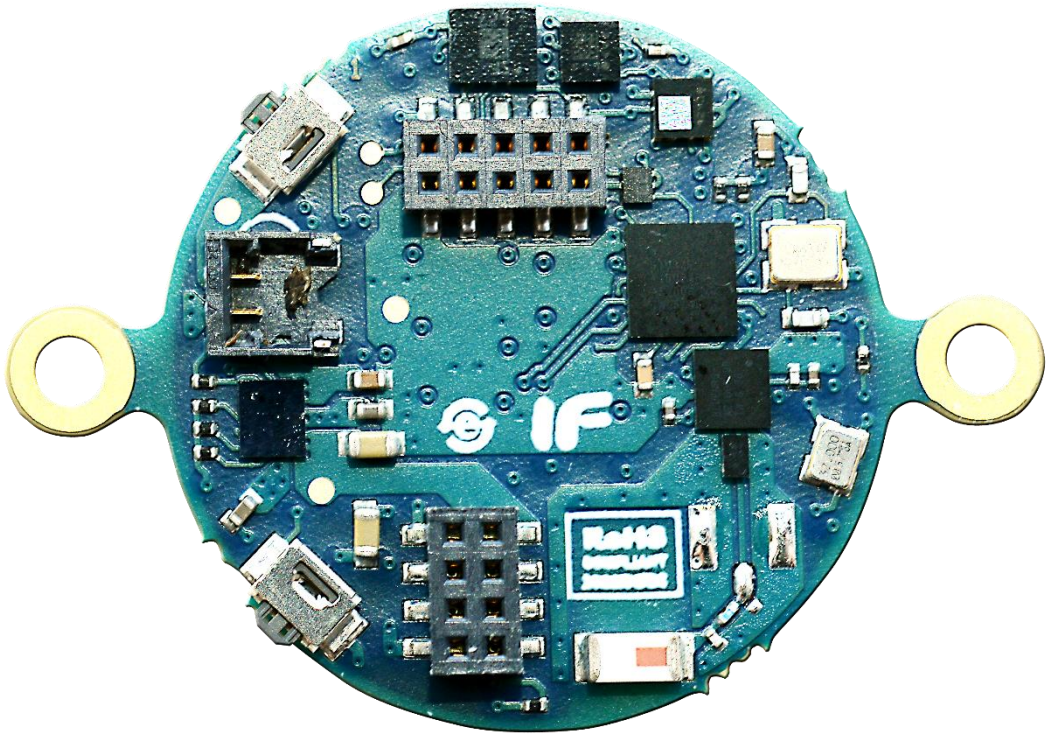
Uniform Rectangular Array

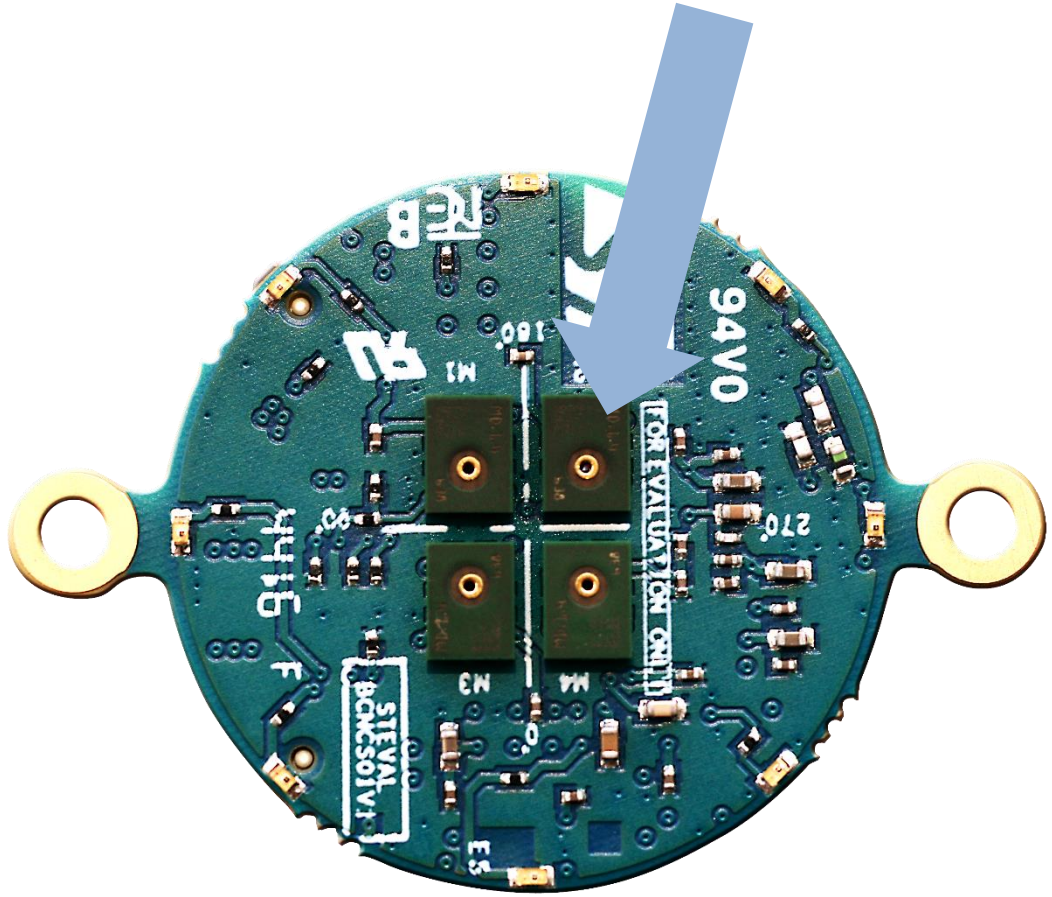
and many more…

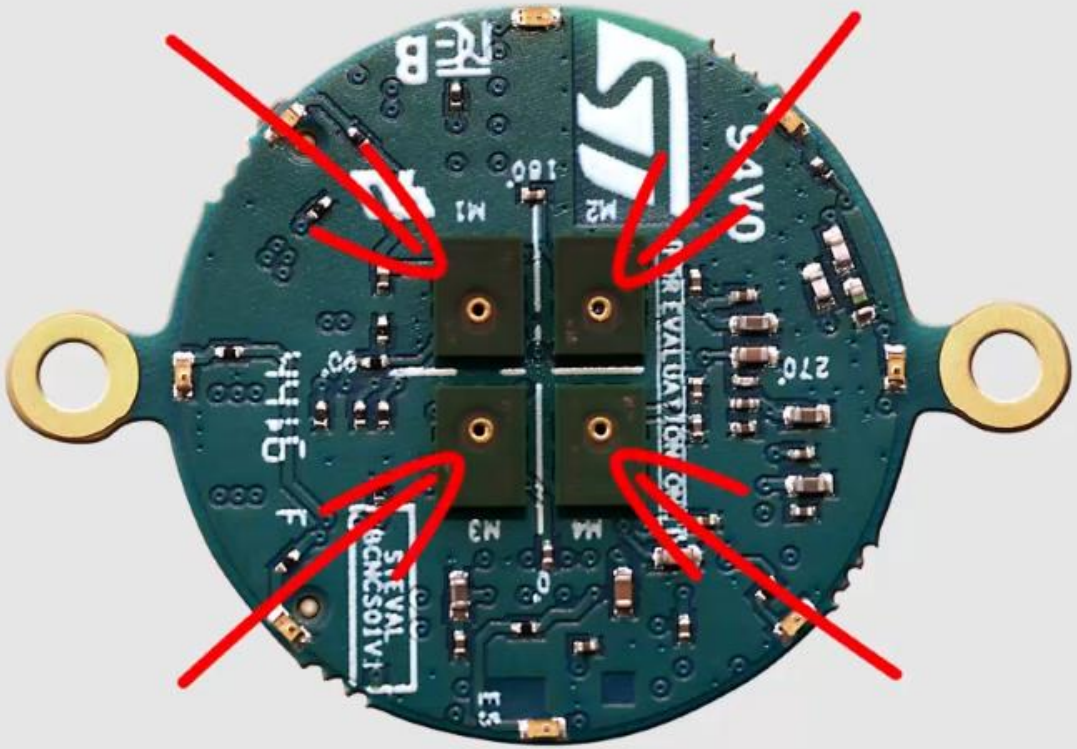
Design a Microphone Array System

…starting from a given array hardware

- BlueCoin from ST Microelectronics







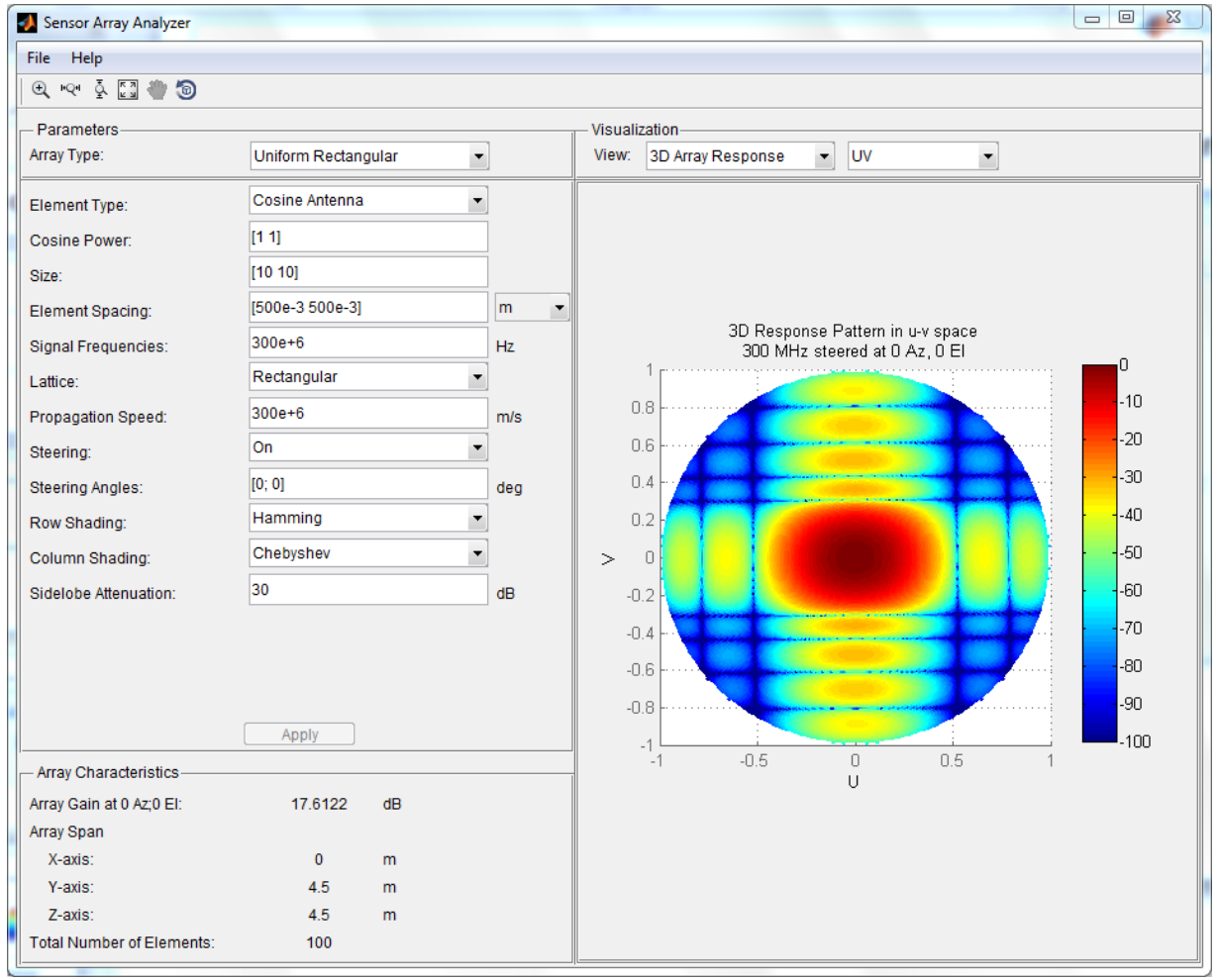
4 microphones

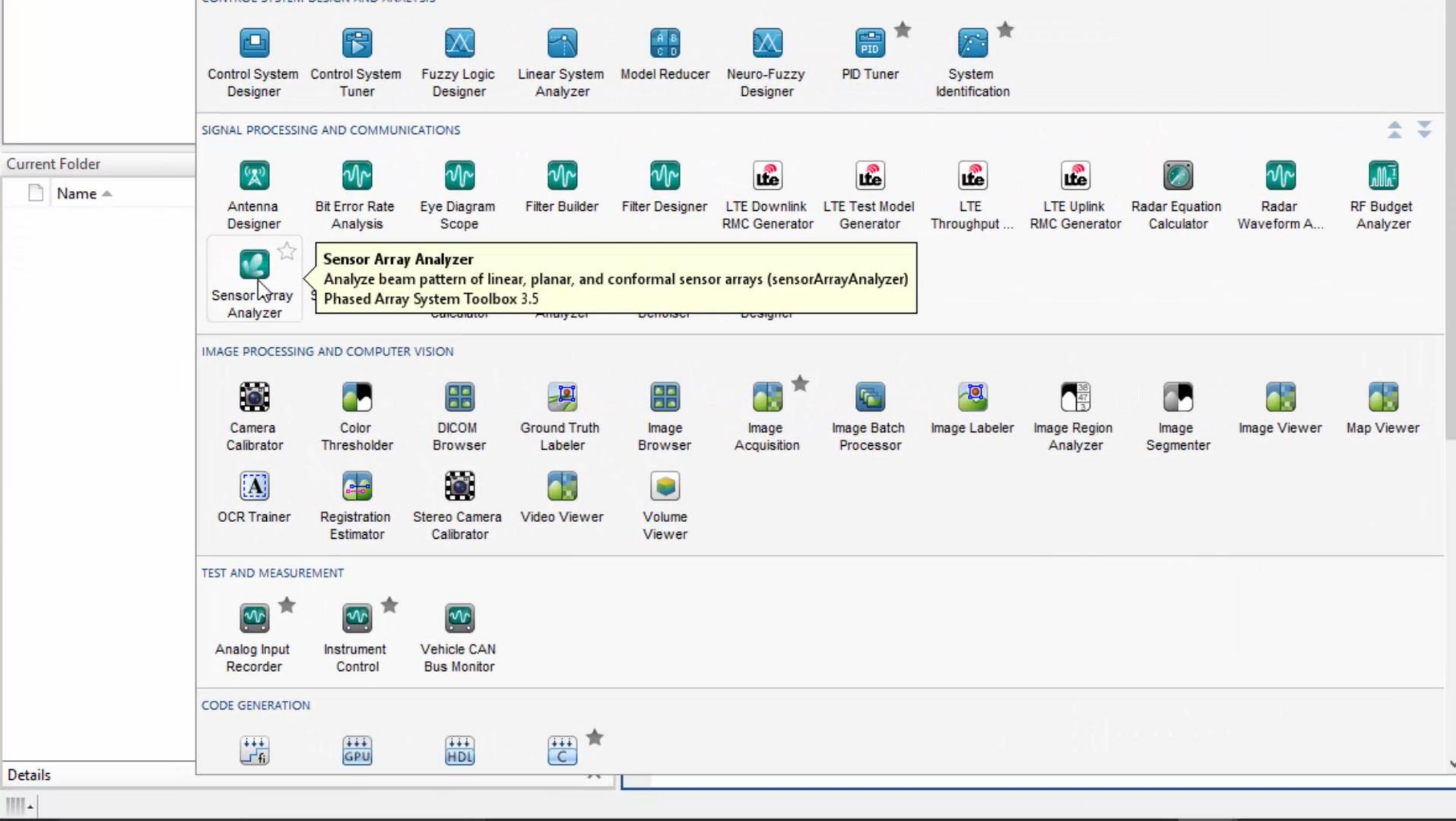
Microphone Array Design- Where to Start?

Array Design and Analysis

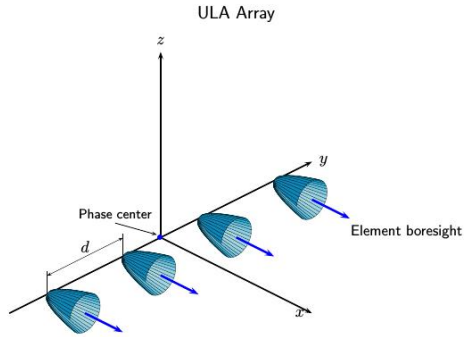
Sensor Array Analyzer

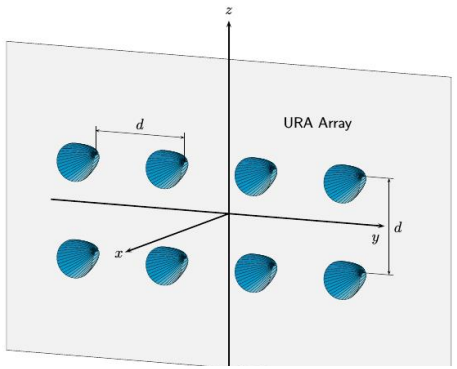
- Element definition
- Array geometry definition
- Array shading/tapers
- Array steering
- Radiation patterns in different forms

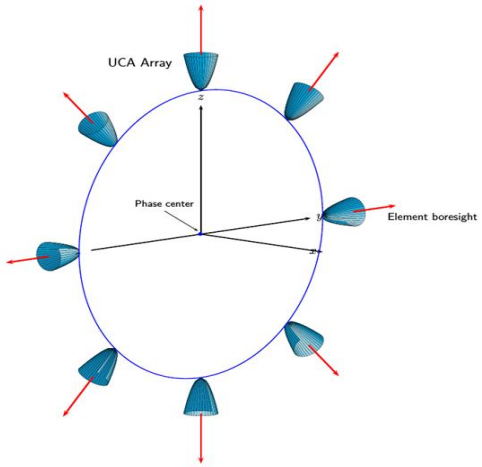


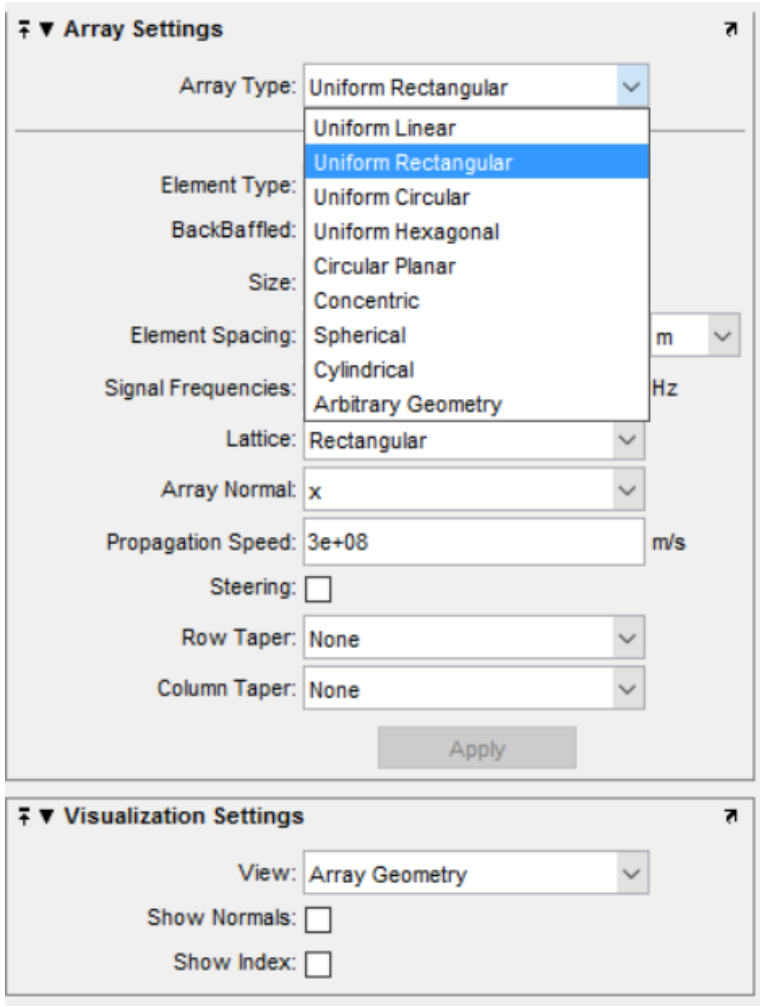

Control System Designer
Control System Tuner
Fuzzy Logic Designer
Linear System Analyzer
Model Reducer
Neuro-Fuzzy Designer
PID Tuner
System Identification
SIGNAL PROCESSING AND COMMUNICATIONS
Antenna Designer
Bit Error Rate Analysis
Eye Diagram Scope
Filter Builder
Filter Designer
LTE Downlink RMC Generator
LTE Test Model Generator
LTE Throughput ...
LTE Uplink RMC Generator
Radar Equation Calculator
Radar Waveform A...
RF Budget Analyzer
Sensor Array Analyzer
Sensor Array Analyzer
Analyze beam pattern of linear, planar, and conformal sensor arrays (sensorArrayAnalyzer)
Phased Array System Toolbox 3.5
IMAGE PROCESSING AND COMPUTER VISION
Camera Calibrator
Color Thresholder
DICOM Browser
Ground Truth Labeler
Image Browser
Image Acquisition
Image Batch Processor
Image Labeler
Image Region Analyzer
Image Segmenter
Image Viewer
Map Viewer
OCR Trainer
Registration Estimator
Stereo Camera Calibrator
Video Viewer
Volume Viewer
TEST AND MEASUREMENT
Analog Input Recorder
Instrument Control
Vehicle CAN Bus Monitor
CODE GENERATION
Current Folder
Name
Details

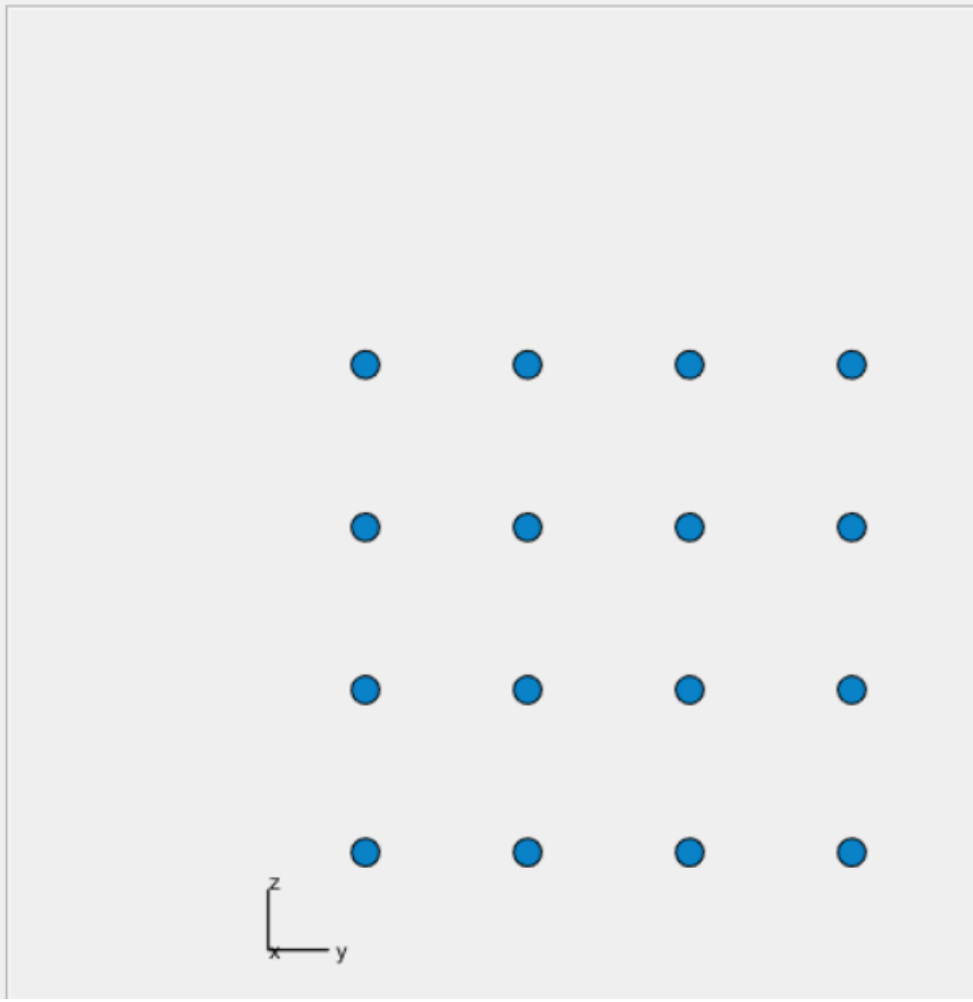
Choosing between different array geometries

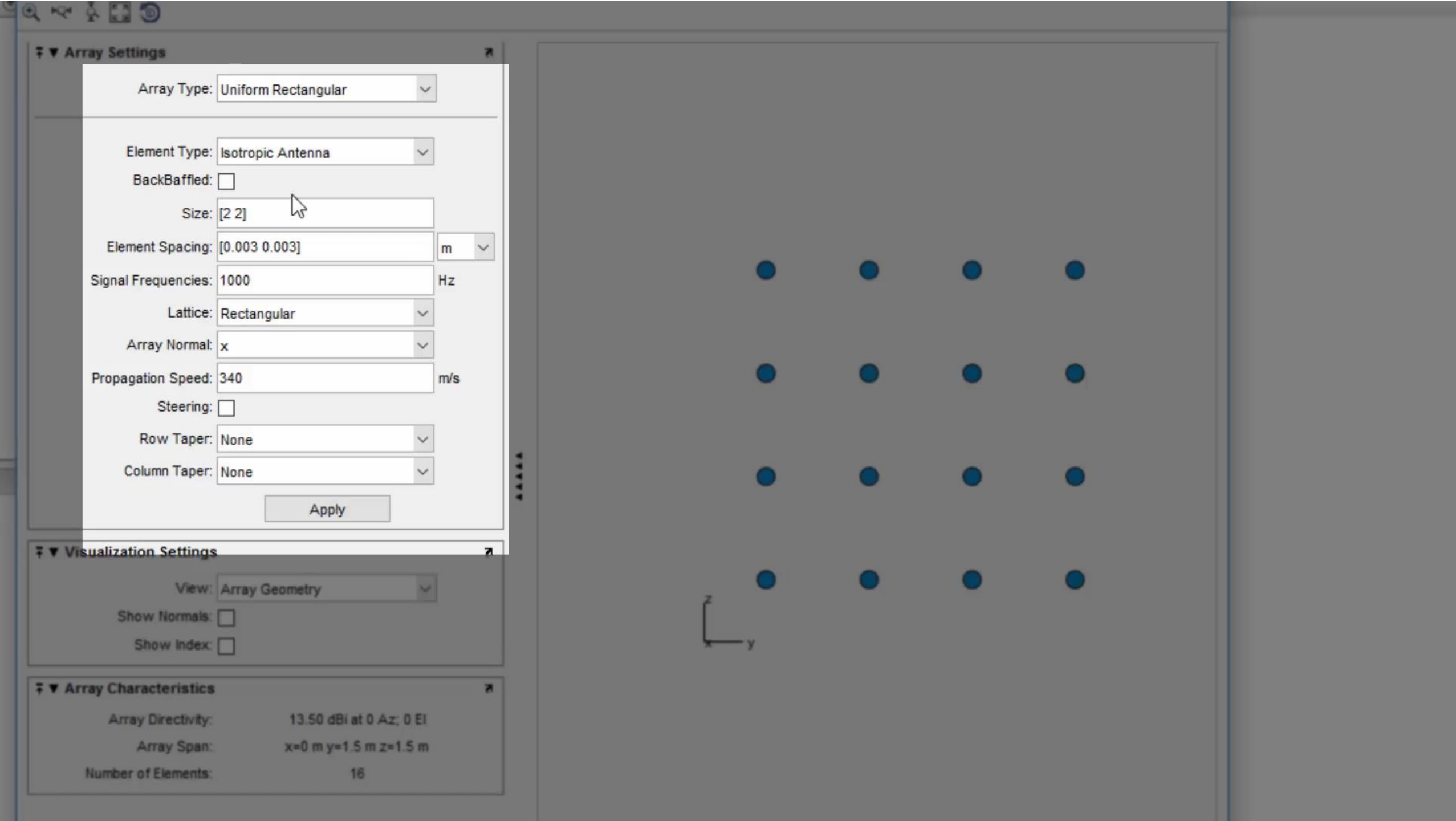











Array Settings
Array Type: Uniform Rectangular
Element Type: Isotropic Antenna
BackBaffled:
Size: [2 2]
Element Spacing: [0.003 0.003] m
Signal Frequencies: 1000 Hz
Lattice: Rectangular
Array Normal: x
Propagation Speed: 340 m/s
Steering:
Row Taper: None
Column Taper: None
Apply
Visualization Settings
View: Array Geometry
Show Normals:
Show Index:
Array Characteristics
Array Directivity: 13.50 dBi at 0 Az; 0 El
Array Span: x=0 m y=1.5 m z=1.5 m
Number of Elements: 16
z
y
x

Array Settings

Array Type: Uniform Rectangular

Element Type: Isotropic Antenna

BackBaffled: ☐

Size: [2 2]

Element Spacing: [0.003 0.003] m

Signal Frequencies: 1000 Hz

Lattice: Rectangular

Array Normal: x

Propagation Speed: 340 m/s

Steering: ☐

Row Taper: None

Column Taper: None

Apply

Visualization Settings

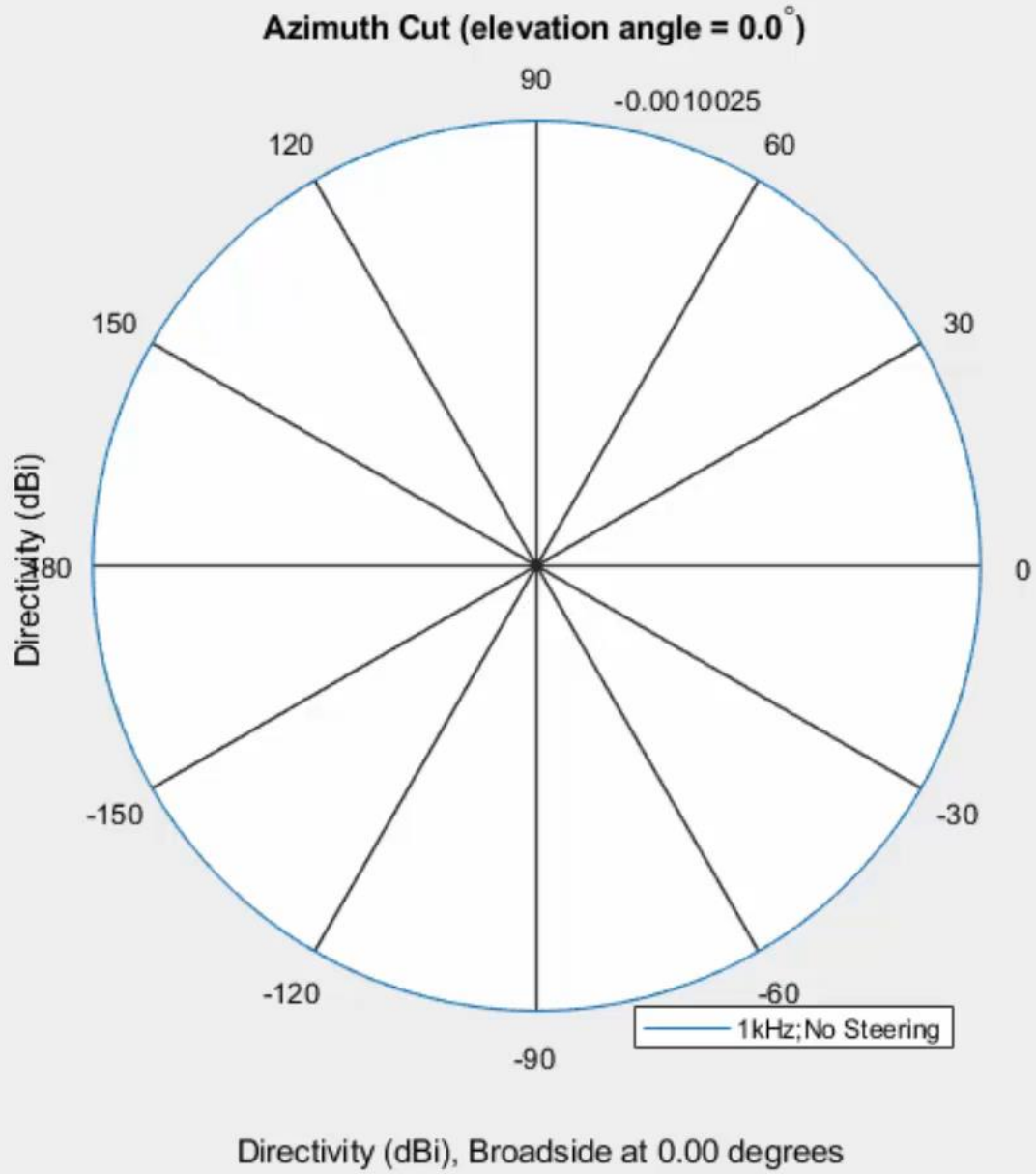
View: 2D Array Directivity

Cut Type: Azimuth Cut-Polar

Cut Value: 0 deg

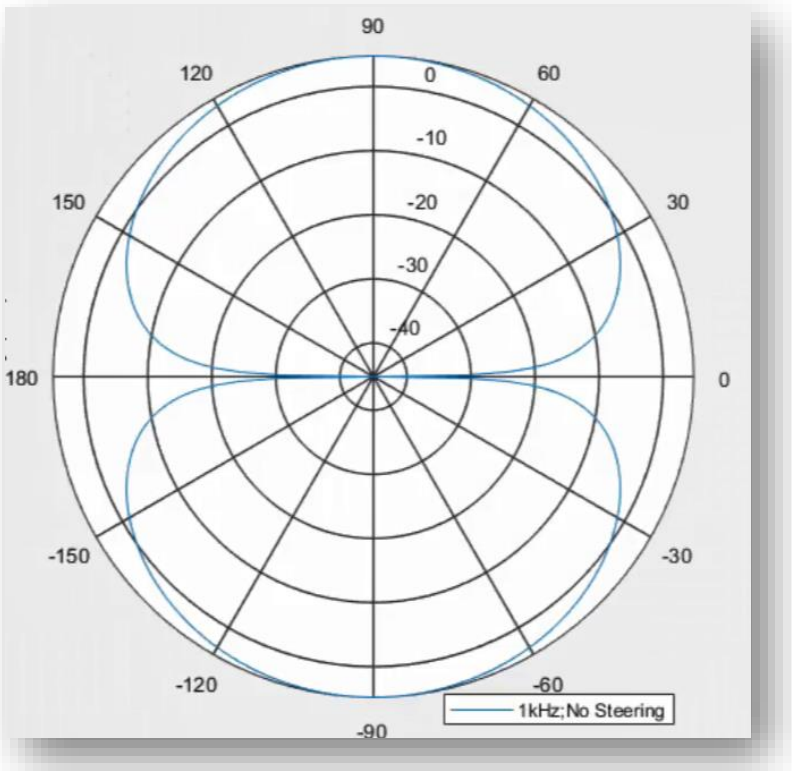
Array Characteristics

Array Directivity:	0.00 dBi at 0 Az; 0 El
Array Span:	x=0 m y=3 mm z=3 mm
Number of Elements:	4

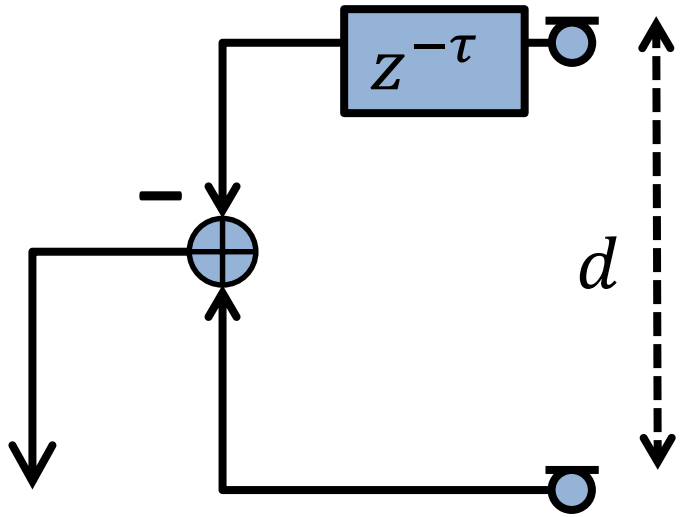


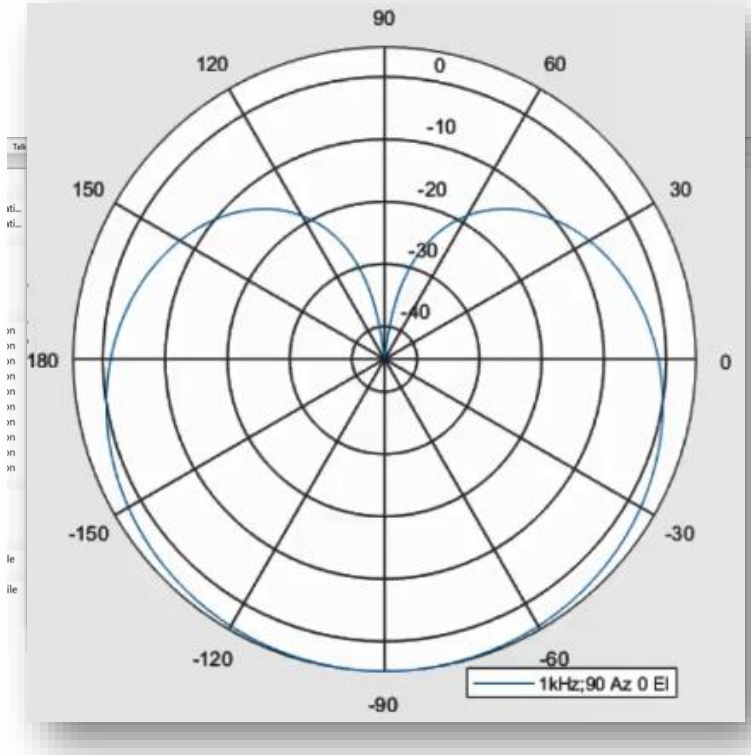
How do you get a Cardioid Pattern?

Differential beamforming

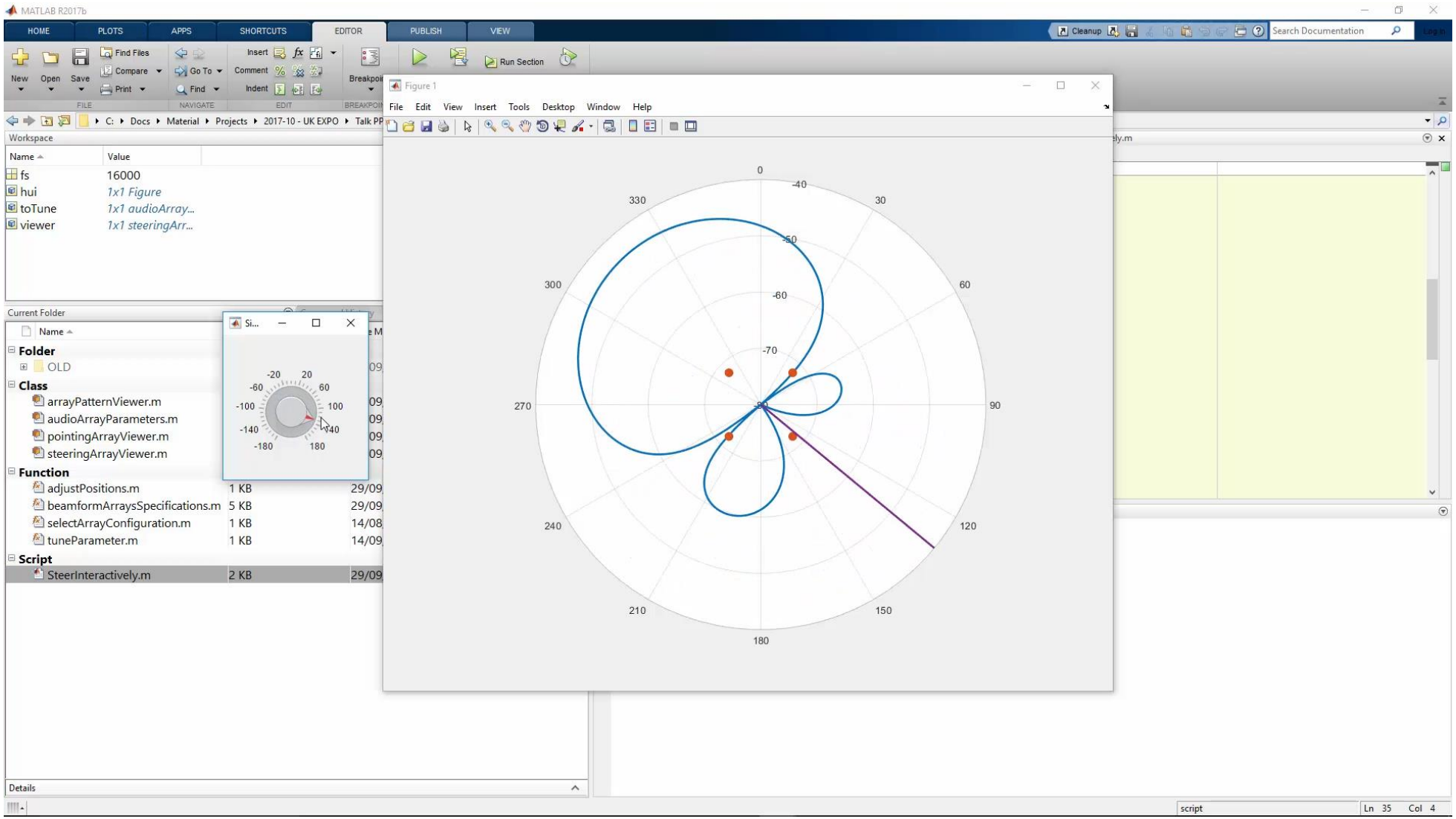


$$\tau = \frac{d}{c}$$

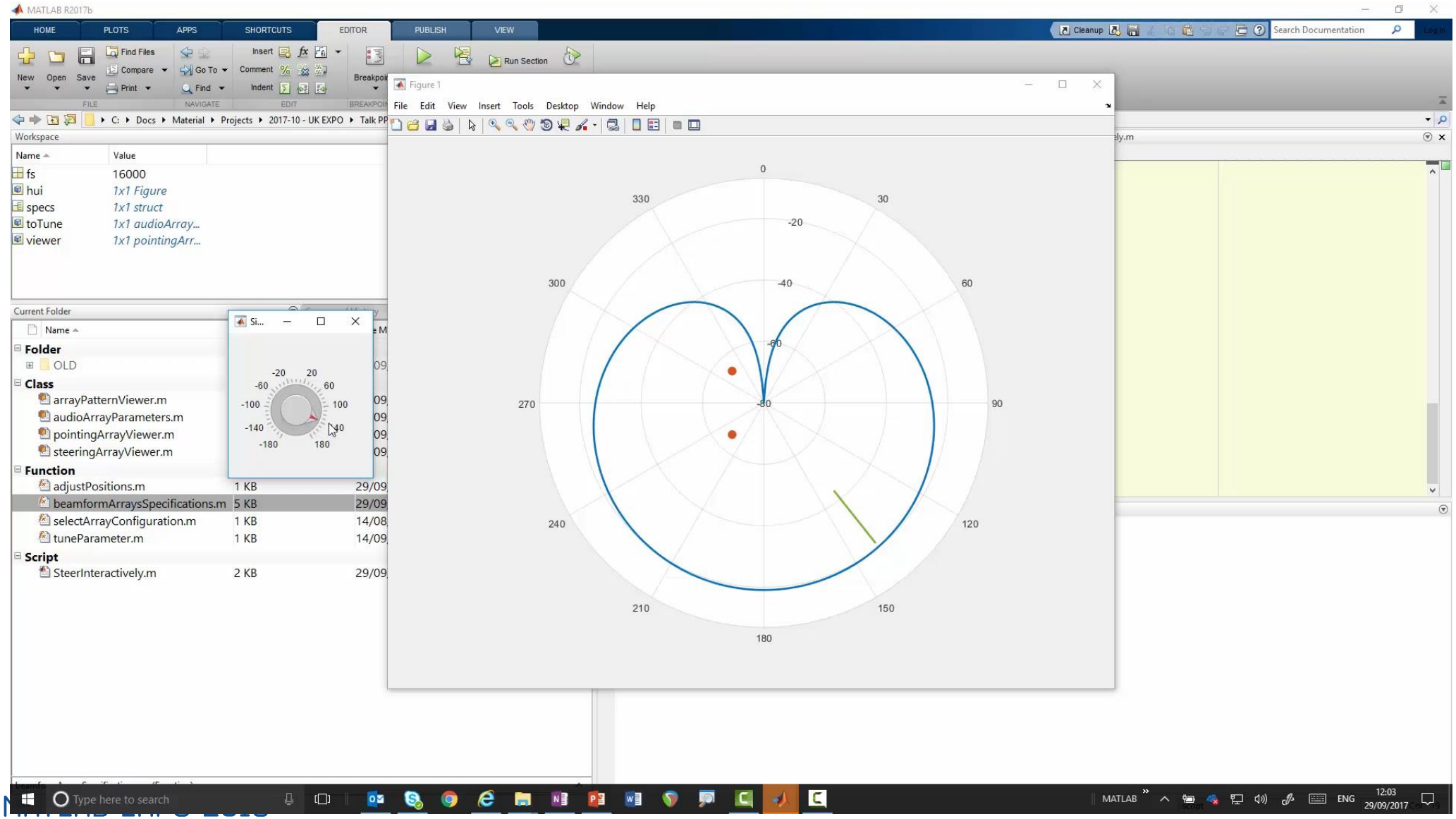




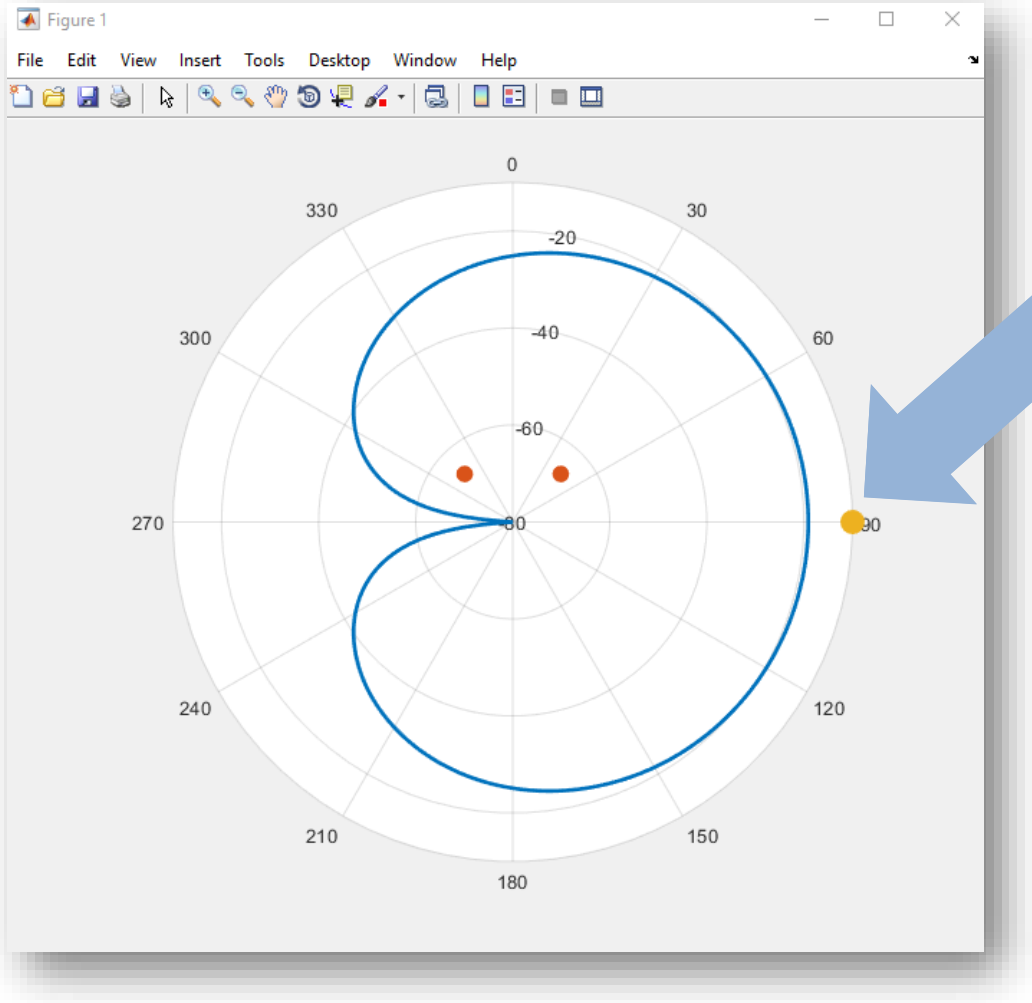
Beam Steering – Does this work?



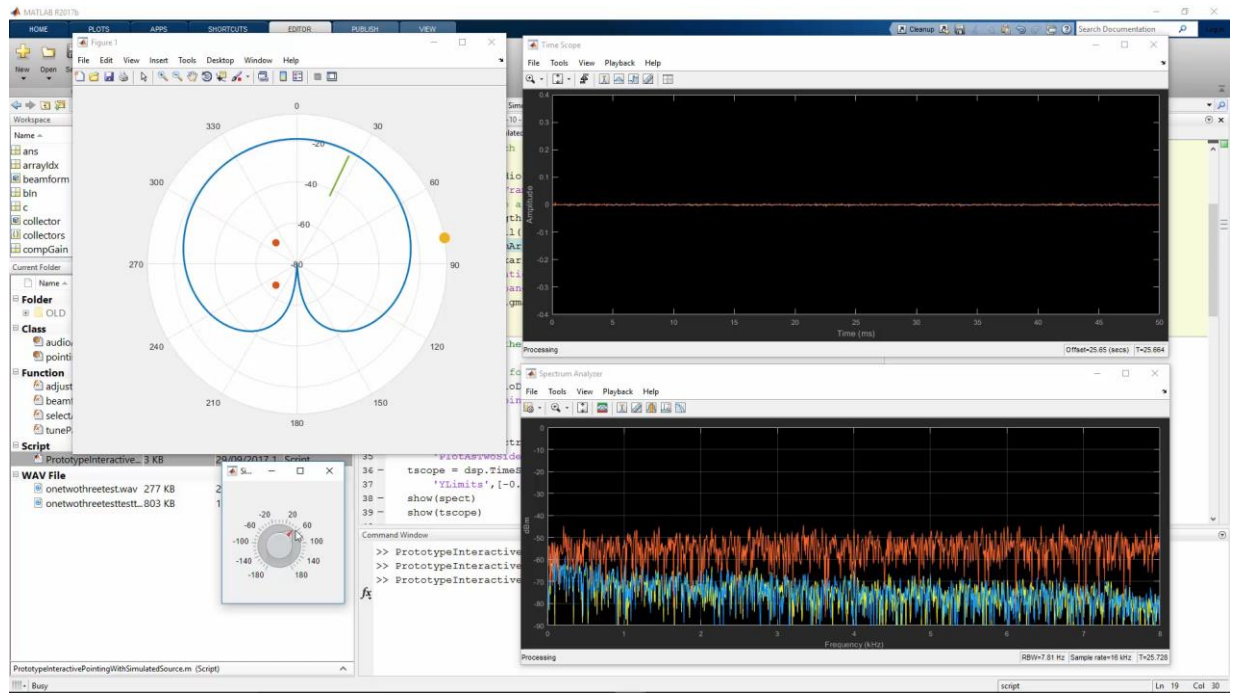
Variable Array Configuration



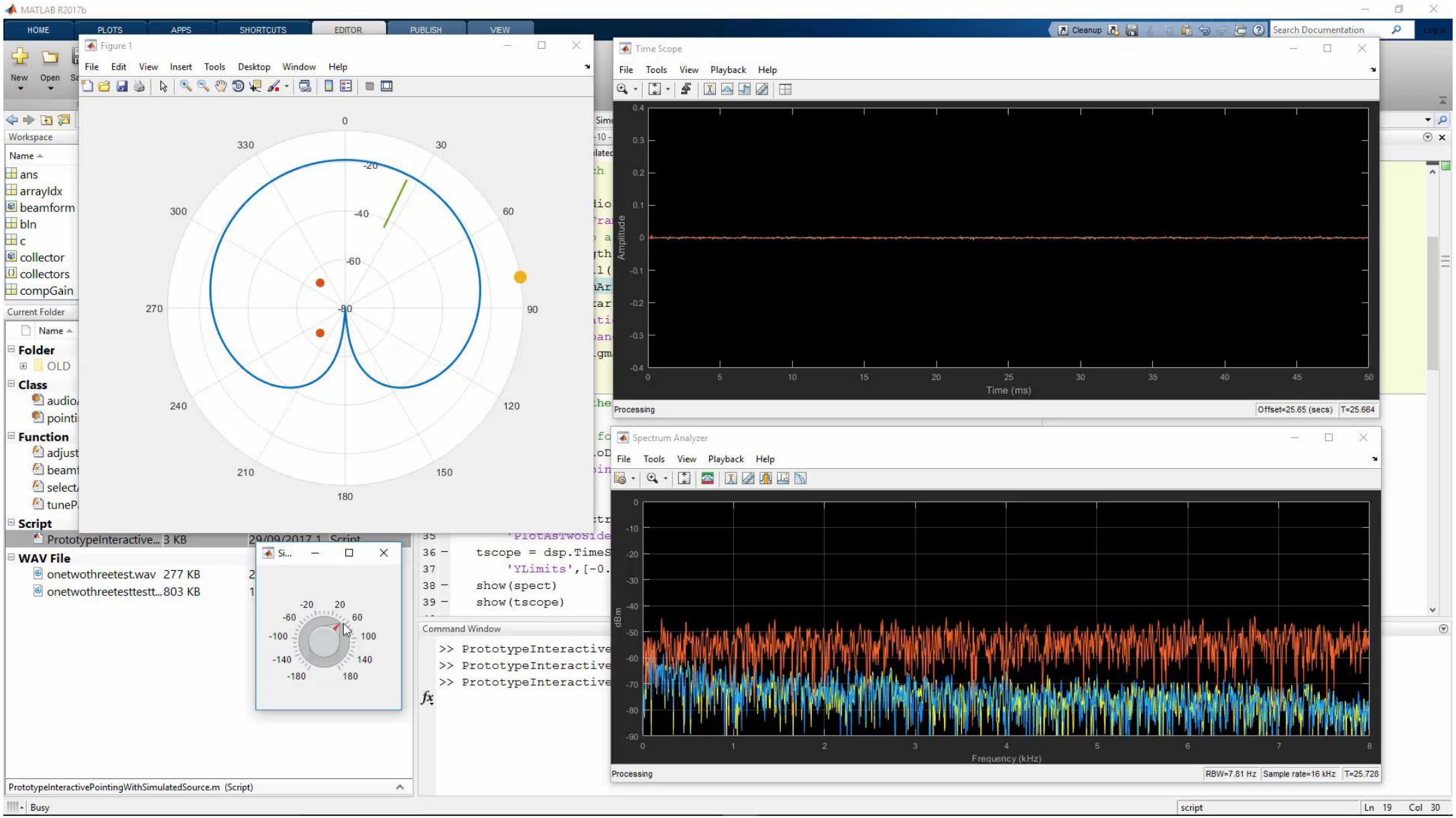
Time Domain Simulation of an Array System



Sound source



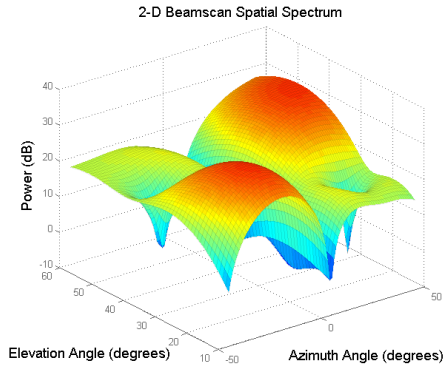
Time Domain Simulation of an Array System

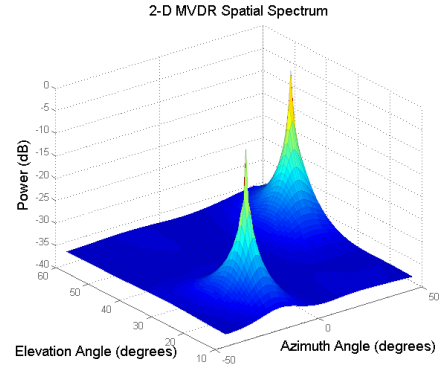


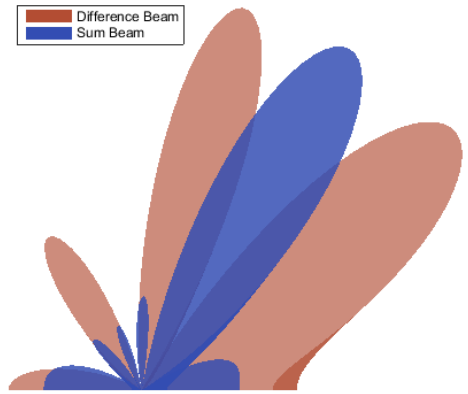
Direction of Arrival Estimation

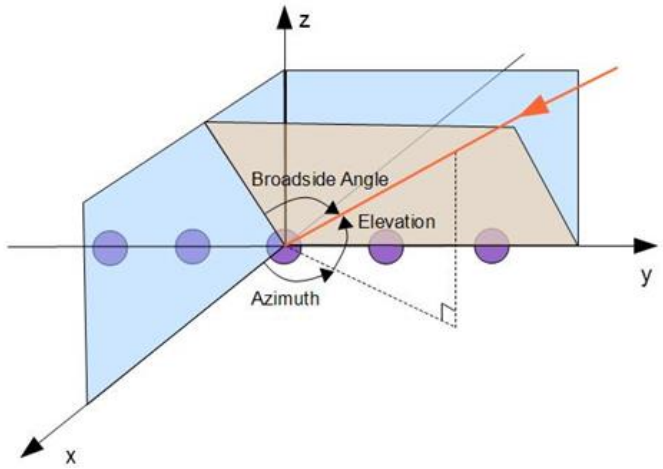
Phased Array System Toolbox

- ULA
  - Sum and difference monopulse
  - Beamscan, MVDR (Capon)
  - High resolution (ESPRIT, Root MUSIC, etc...)
- URA
  - Sum and difference monopulse
  - Beamscan, MVDR (Capon)
- Conformal arrays
  - Beamscan
  - MVDR (Capon)



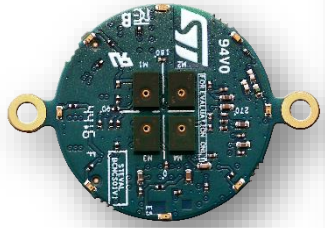


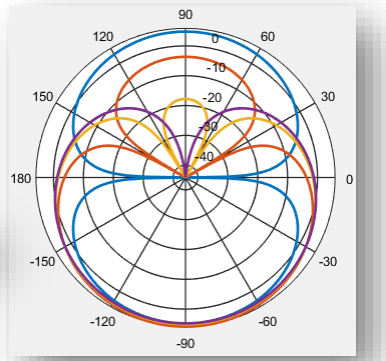




How Can I…

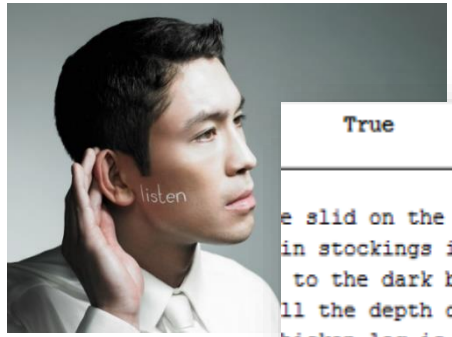
1. Design a voice interface?

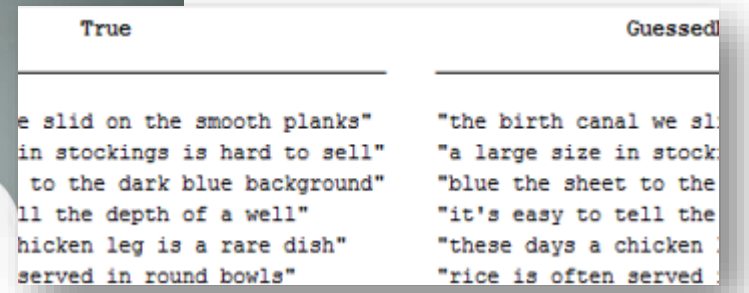




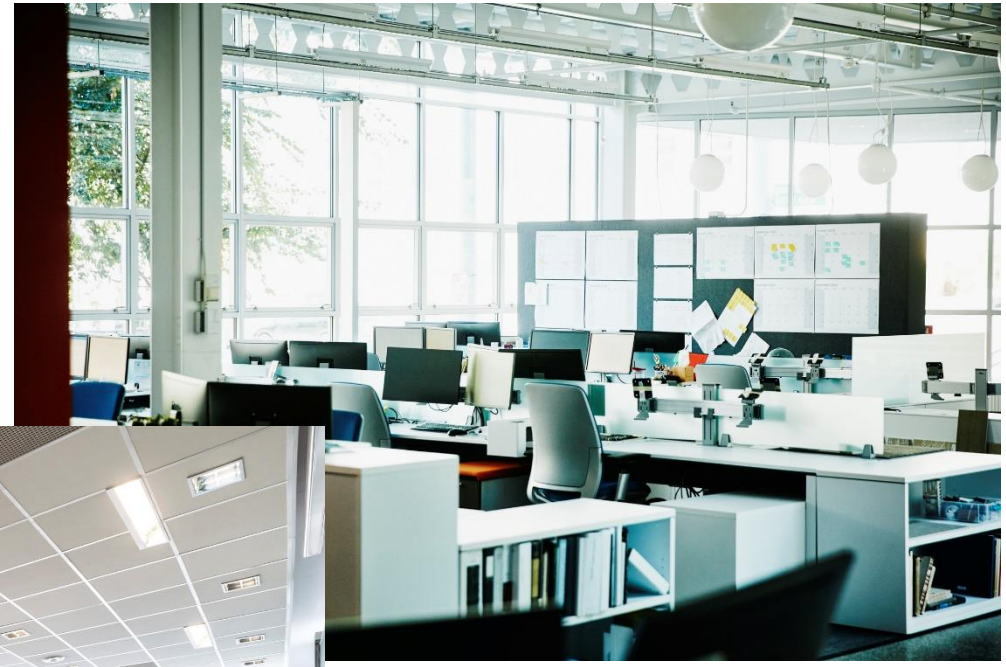
2. **Validate if my voice interface can work in real-life scenarios?**

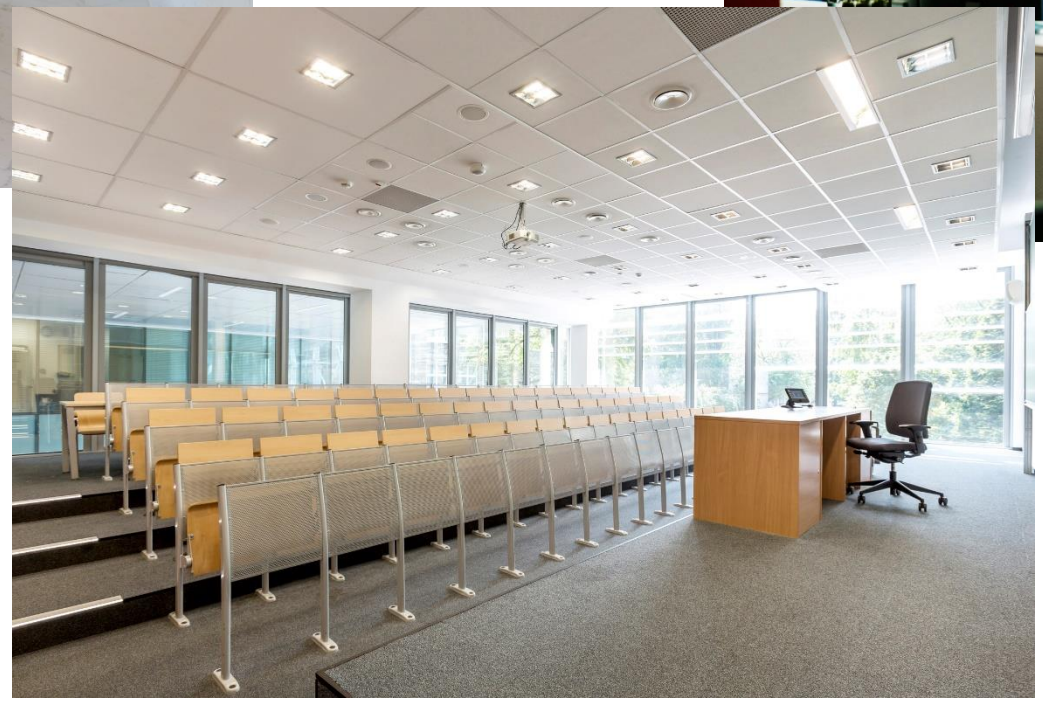
3. Test the performance of my system?





Constrained Simulations vs. Real Life





Validating in Real-Life Scenarios

1. Including the Room Impulse Response

- Record the audio in a constrained environment
- Include the Room Impulse Response (RIR)
  - Generate RIR
  - Measure RIR

2. Live Acquisition

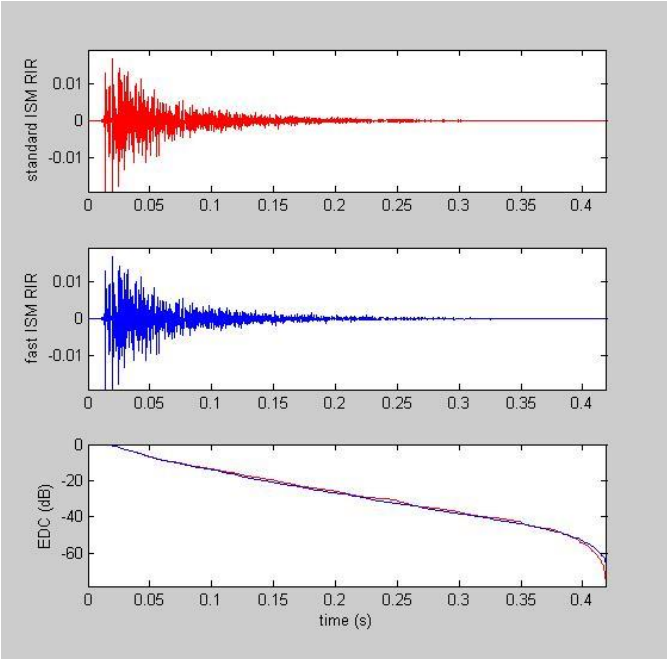
- Connect a microphone array
- Acquire speech signals live in a real-life environment

Integrating the Room Impulse Response

Room Impulse Response Simulation

Image Source Method (ISM) for simulation of Room Impulse Response (RIR) in small-room acoustics

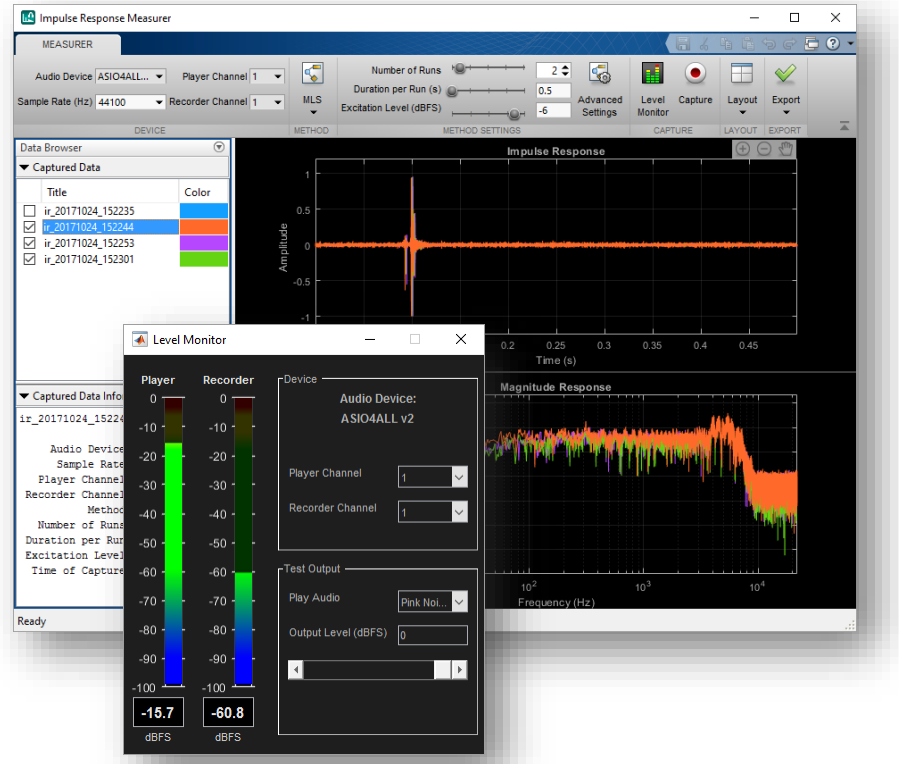
[Link to the code]



Impulse Response Measurer App R2018a

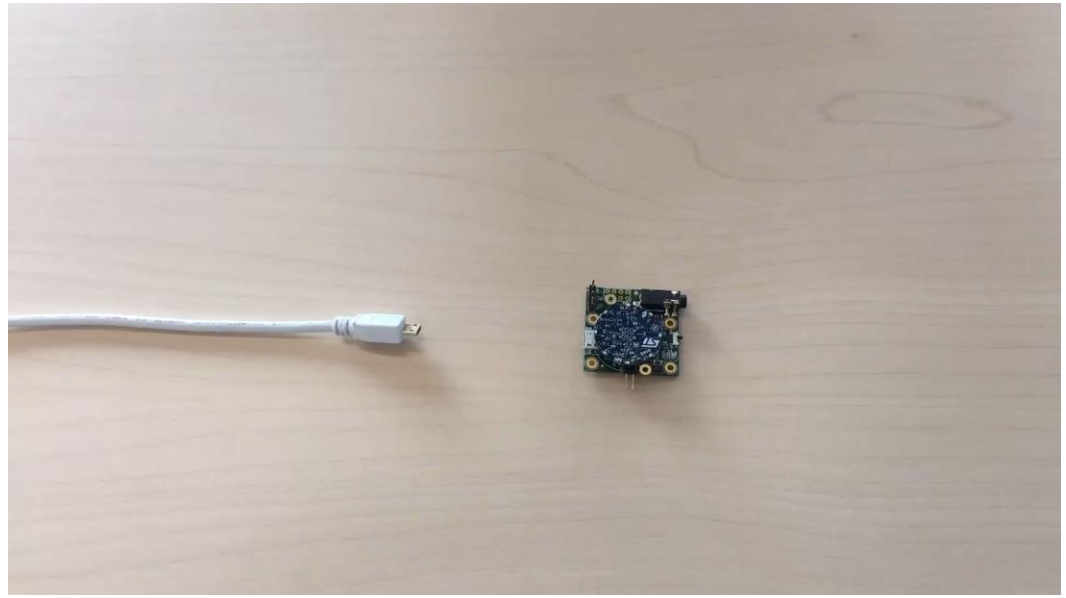
Measure impulse and frequency responses of electrical and acoustic systems with:

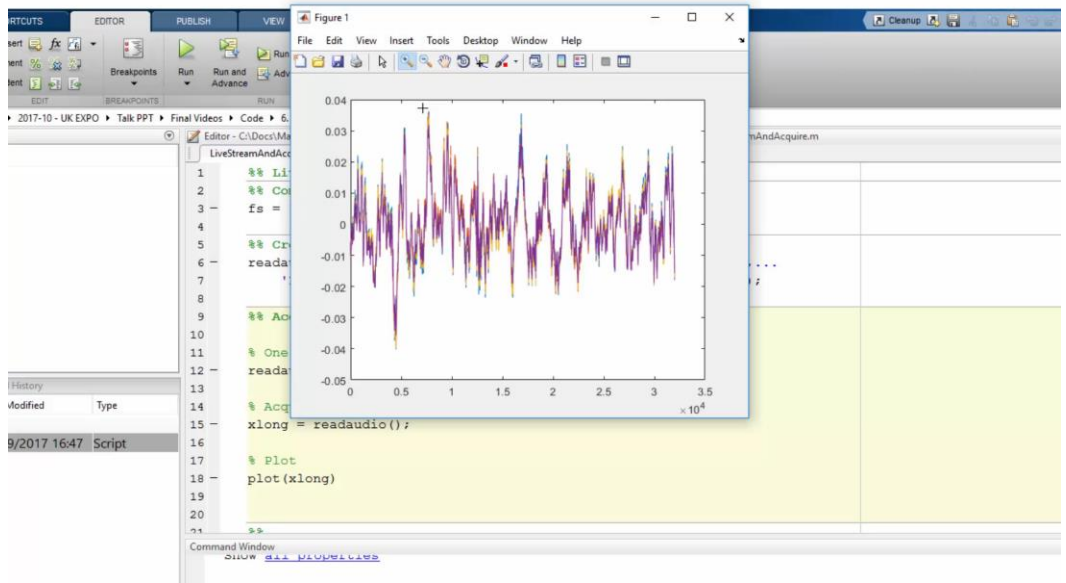
- Maximum-Length Sequences (MLS)
- Exponentially Swept Sinusoids (ESS)



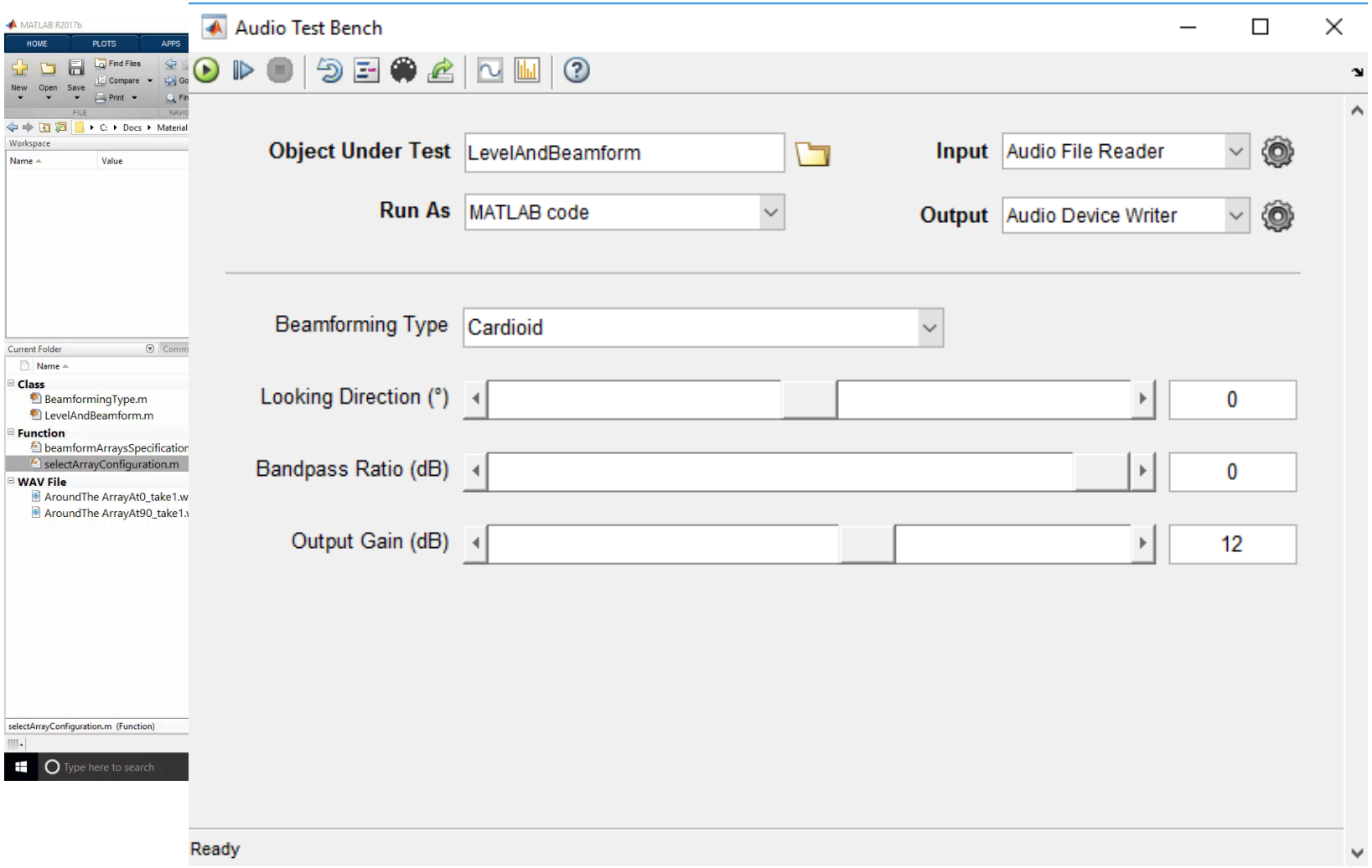
Live Acquisition from Hardware

```
%% Live Audio Acquisition and Streaming
fs = 16000;
tscope = dsp.TimeScope('SampleRate',fs);

readaudio =
audioDeviceReader('SampleRate',fs,'NumChann
els',4,...
    'Device', 'Microphone (STM32 AUDIO
Streaming in FS Mode)');
% Set block duration
readaudio.SamplesPerFrame = 1024;
while isVisible(tscope)
    % Acquire live
    in = readaudio();
    % Visualize in real-time
    tscope(in);
end

release(readaudio)
```





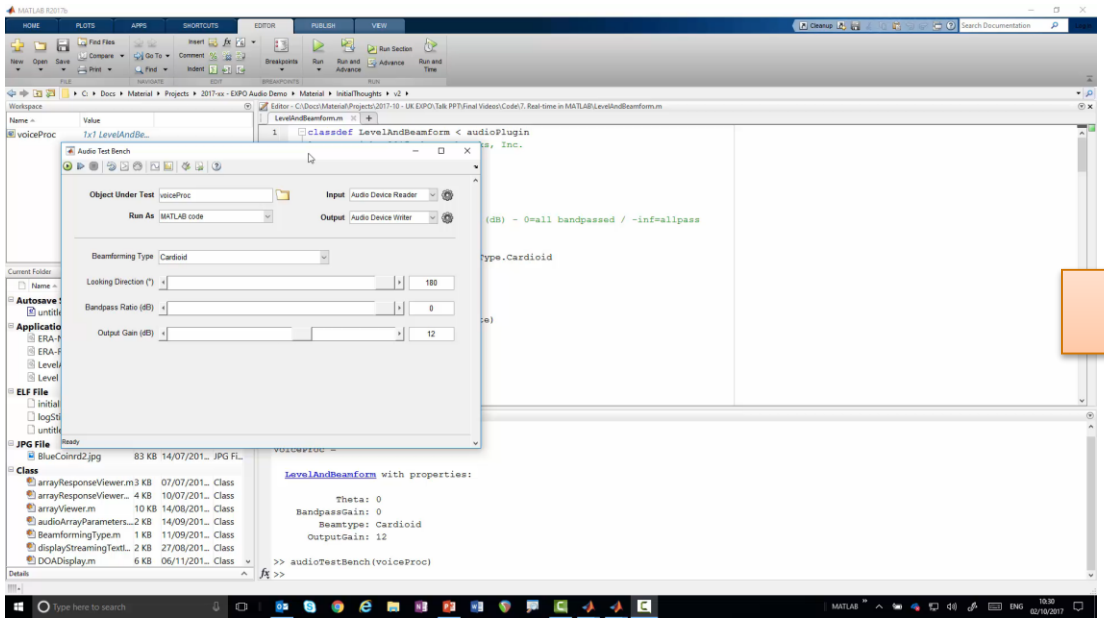
Creating Audio Testbench



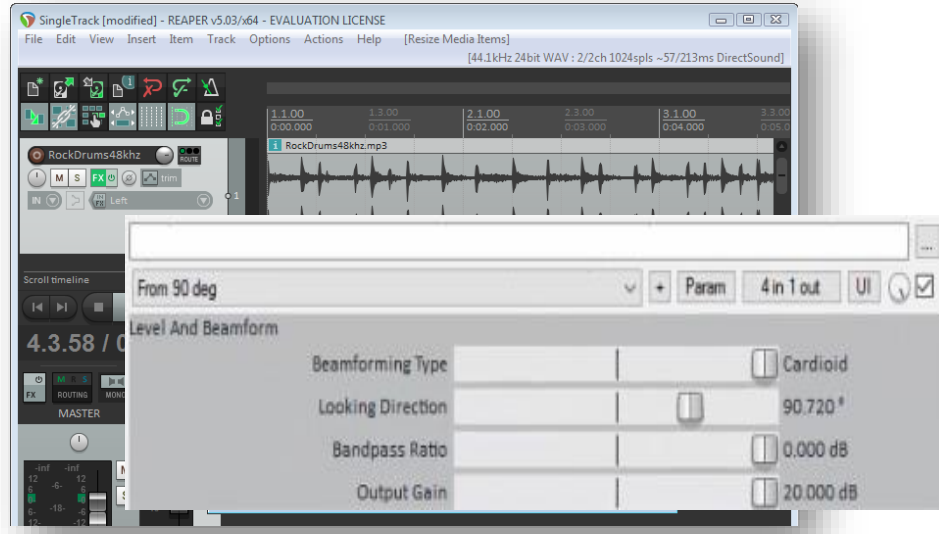
- Debug your audio plugin
- Time and frequency domain visualization of your processing
- Option to bypass algorithm under test for interactive A/B testing

Reuse your design in Digital Audio Workstations

Generating VST Plugins

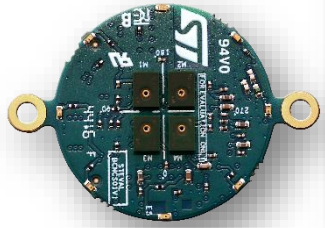


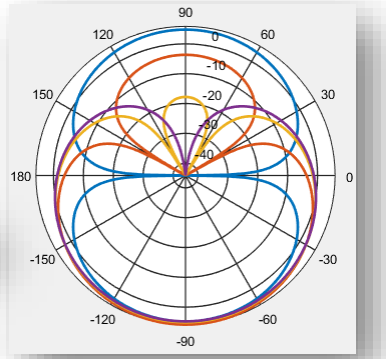
VST plugin



How Can I…

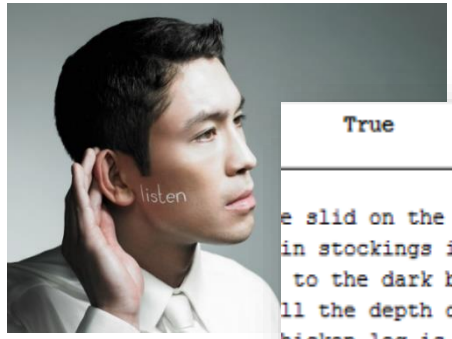
1. Design a voice interface?





2. Validate if my voice interface can work in real-life scenarios?

3. **Test the performance of my system?**



True	Guessed
e slid on the smooth planks"	"the birth canal we sl
in stockings is hard to sell"	"a large size in stock
to the dark blue background"	"blue the sheet to the
ll the depth of a well"	"it's easy to tell the
hicken leg is a rare dish"	"these days a chicken
served in round bowls"	"rice is often served

How To Measure Performance?

```
while ishandle(hui)

    % Audio acquisition (from file)
    rxAll = readaudio();
    inAll = levelChannels(rxAll);

    % Event segmentation and DOA
    [toTune.doa, segmentation, eventsDetected] = bufferAndSearch(inAll, fs, searchBufferLength);
    % Point array towards estimated DOA
    toTune.pointingTheta = toTune.doa;

    % Select correct subarray for pointing
```

Subjective Analysis

Output audio "sounds good"

Speech-to-text

True	Guessed
e slid on the smooth planks"	"the birth canal we sl
in stockings is hard to sell"	"a large size in stock
to the dark blue background"	"blue the sheet to the
ll the depth of a well"	"it's easy to tell the
hicken leg is a rare dish"	"these days a chicken
served in round bowls"	"rice is often served

"91.5% of spoken sentences correctly converted"

Test performance with speech-to-text services

```
>> [samples, fs] = audioread('helloaudioPD.wav');
>> soundsc(samples, fs)
>> speechObject = speechClient('Google','languageCode','en-US');
>> outInfo = speech2text(speechObject, samples, fs);

>> outInfo.TRANSCRIPT =

ans =
    'hello audio product Developers'

>> outInfo.CONFIDENCE =

ans =
    0.9385
```

Importing data into MATLAB

```
while ishandle(hui)

    % Audio acquisition (from file)
    rxAll = readaudio();
    inAll = levelChannels(rxAll);

    % Event segmentation and DOA
    [toTune.doa, segmentation, eventsDetected] = bufferAndSearch(inAll, fs, searchBufferLength);
    % Point array towards estimated DOA
    toTune.pointingTheta = toTune.doa;

    % Select correct subarray for pointing
```

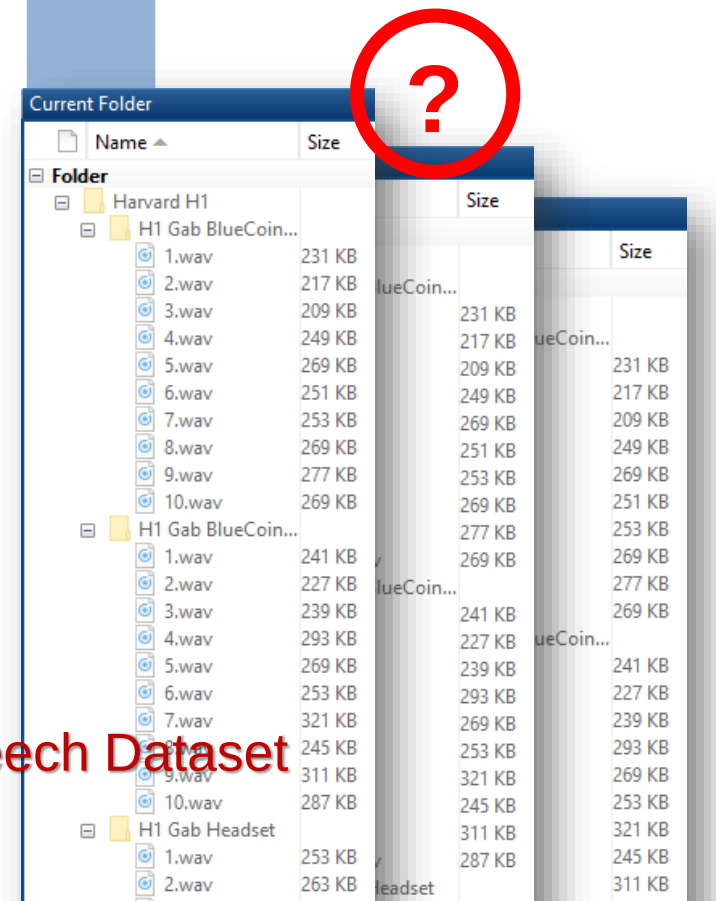
Voice Interface

Connecting MATLAB to Cloud

Cloud Services

Speech to Text

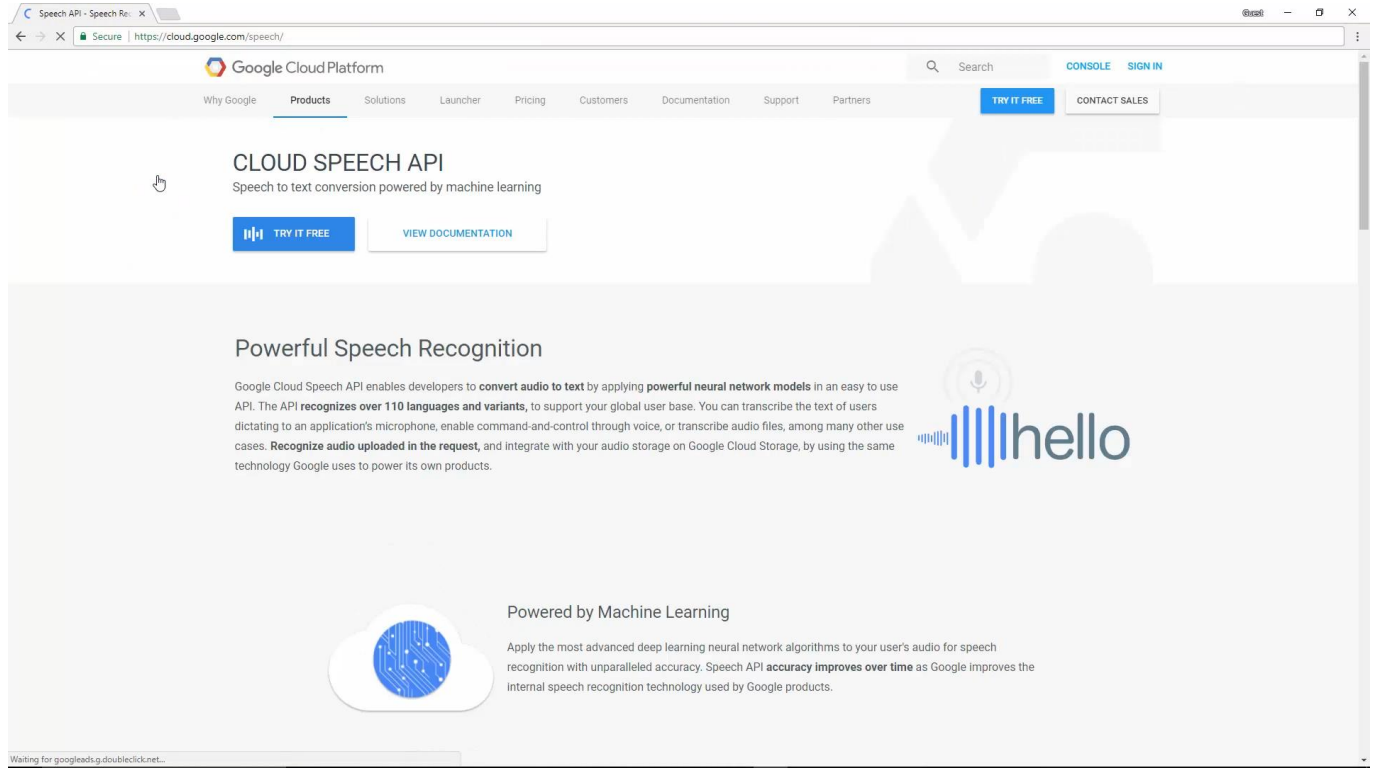
Speech Dataset



True	GuessedBaseline	ConfidenceBaselir
"the birch canoe slid on the smooth planks"	"the birth canal we slid on the smooth planks"	0.74933
" a large size in stockings is hard to sell"	"a large size in stockings is hard to sell"	0.9366
"glue the sheet to the dark blue background"	"blue the sheet to the dark blue background"	0.95009
"its easy to tell the depth of a well"	"it's easy to tell the depth of a well"	0.97403
"these days a chicken leg is a rare dish"	"these days a chicken leg is a redfish"	0.86686
"rice is often served in round bowls"	"rice is often served in rum balls"	0.95946
"the juice of lemons makes fine punch"	"the cheese of lemons makes flying punch"	0.79089
"the box was thrown beside the parked truck"	"the box was thrown beside the box truck"	0.82796
"the hogs were fed chopped corn and garbage"	"the Hawks were fed chops corn and garbage"	0.81863
"four hours of steady work faced us"	"4 hours of study world fastest"	0.82947

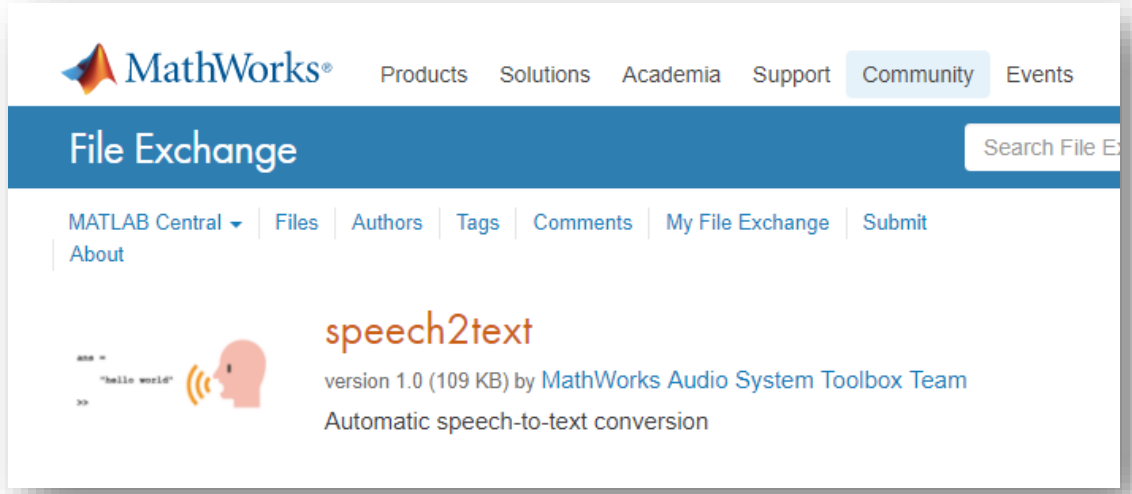
Connecting to Cloud Services for Speech Recognition

- Google® Speech API
- IBM® Watson Speech API
- Microsoft® Azure Speech API



Speech-To-Text – Access 3rd-party web services from MATLAB

- Automate content labelling of speech datasets
- Validate speech enhancement algorithms for transcription performance
- Run text analytics on auto-transcribed voice recordings
- Choice of
  - Google® Speech API
  - IBM® Watson Speech API
  - Microsoft® Azure Speech API
- Requires separate credentials for service provider of choice



https://www.mathworks.com/matlabcentral/fileexchange/65266-speech2text

Importing data into MATLAB

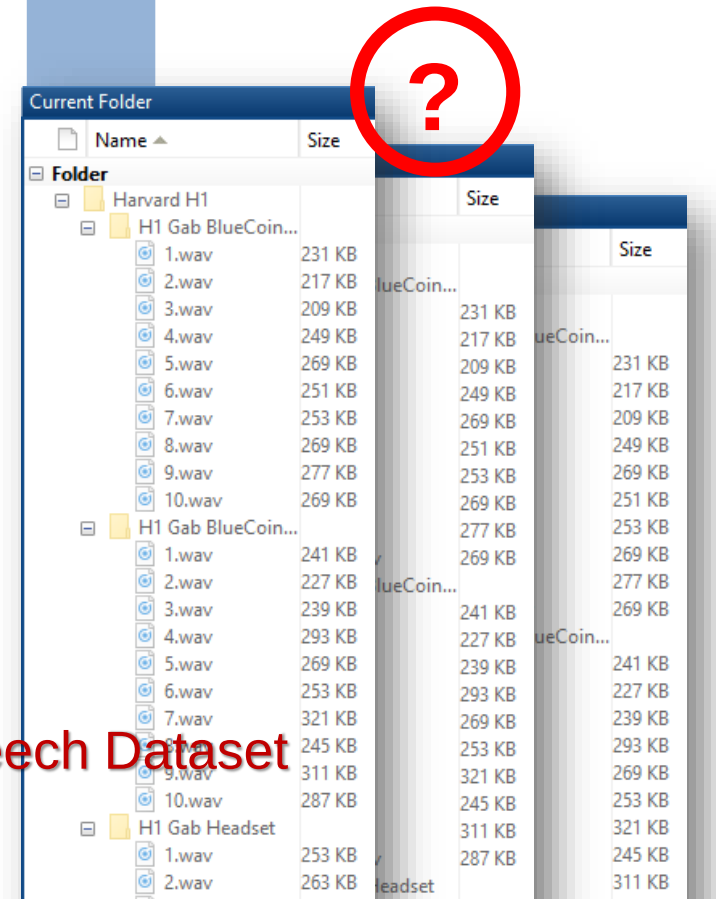
```
while ishandle(hui)

    % Audio acquisition (from file)
    rxAll = readaudio();
    inAll = levelChannels(rxAll);

    % Event segmentation and DOA
    [toTune.doa, segmentation, eventsDetected] = bufferAndSearch(inAll, fs, searchBufferLength);
    % Point array towards estimated DOA
    toTune.pointingTheta = toTune.doa;

    % Select correct subarray for pointing
```

Voice Interface

Connecting MATLAB to Cloud

Cloud Services

Speech to Text

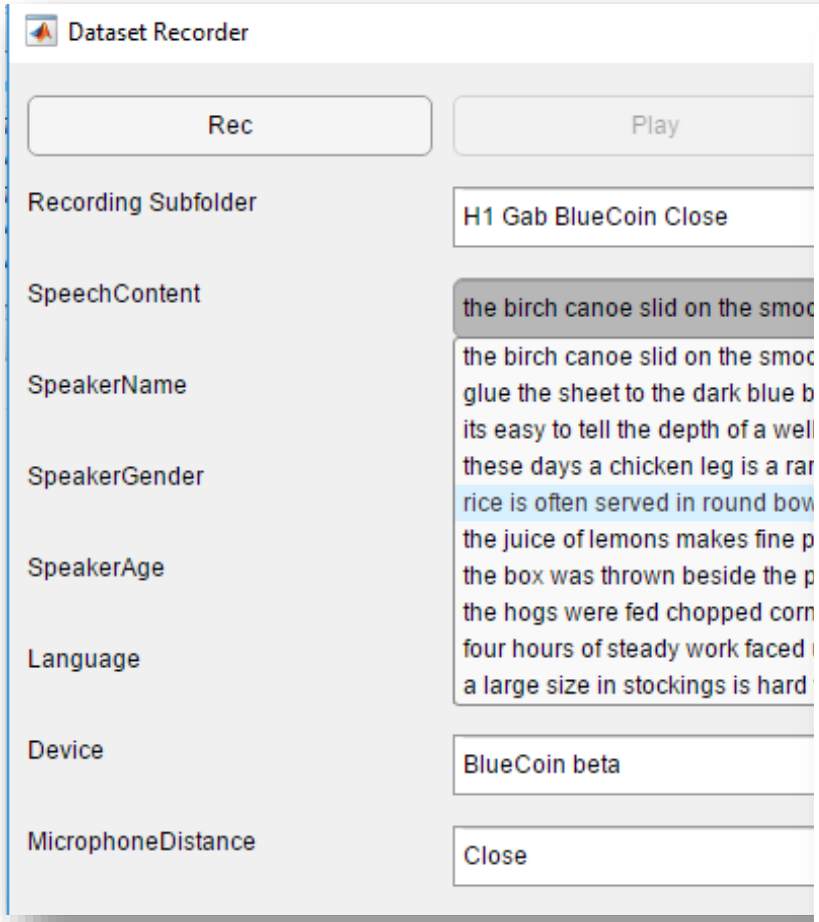


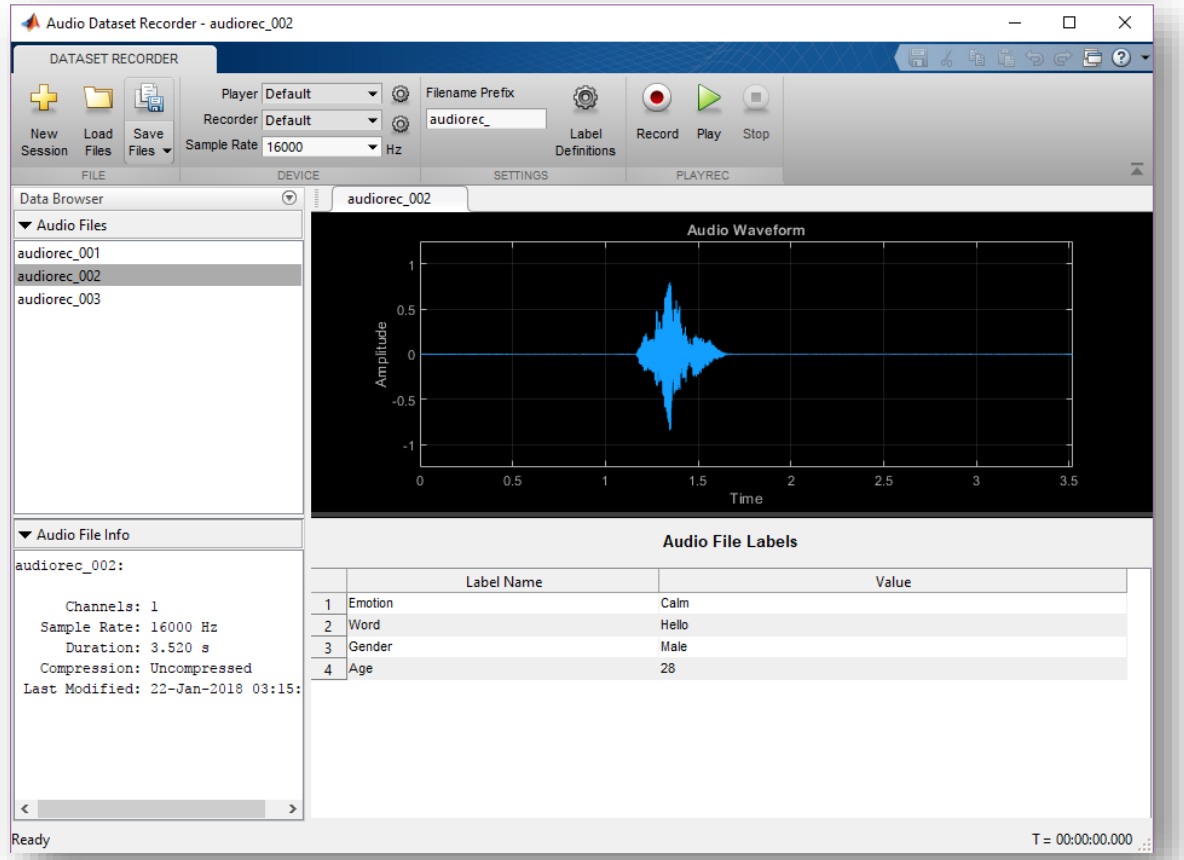
Speech Dataset

True	GuessedBaseline	ConfidenceBaselir
"the birch canoe slid on the smooth planks"	"the birth canal we slid on the smooth planks"	0.74933
" a large size in stockings is hard to sell"	"a large size in stockings is hard to sell"	0.9366
"glue the sheet to the dark blue background"	"blue the sheet to the dark blue background"	0.95009
"its easy to tell the depth of a well"	"it's easy to tell the depth of a well"	0.97403
"these days a chicken leg is a rare dish"	"these days a chicken leg is a redfish"	0.86686
"rice is often served in round bowls"	"rice is often served in rum balls"	0.95946
"the juice of lemons makes fine punch"	"the cheese of lemons makes flying punch"	0.79089
"the box was thrown beside the parked truck"	"the box was thrown beside the box truck"	0.82796
"the hogs were fed chopped corn and garbage"	"the Hawks were fed chops corn and garbage"	0.81863
"four hours of steady work faced us"	"4 hours of study world fastest"	0.82947

Building a small speech dataset quickly

Example: an App with automated content labelling*





See also Dataset Recorder App prototype in example "Record Audio Datasets" (From R2018a in Audio System Toolbox)

MATLAB R2017b

HOME PLOTS APPS SHORTCUTS

Cleanup | Search Documentation | Gabriele

New Script | New Live Script | New | Open | Find Files | Compare | Import Data | Save Workspace | New Variable | Open Variable | Clear Workspace | Analyze Code | Run and Time | Clear Commands | Simulink | Layout | Preferences | Set Path | Parallel | Add-Ons | Help | Community | Request Support | Learn MATLAB

FILE VARIABLE CODE SIMULINK ENVIRONMENT RESOURCES

C: ▸ Users ▸ gbunkhei ▸ Desktop ▸ AES Material ▸ Booth Demos ▸ Background ▸ Voice Interface with BlueCoin ▸ 10. Test With Google Speech ▸ Rec ▸

Workspace | Command History | Current Folder

Name	Size	Type	Date M...
Harvard H1		Folder	13/1...
H1 Gab Blue...		Folder	13/1...
1.wav	231 KB	WAV File	13/0...
2.wav	217 KB	WAV File	13/0...
3.wav	209 KB	WAV File	13/0...
4.wav	249 KB	WAV File	13/0...
5.wav	269 KB	WAV File	13/0...
6.wav	251 KB	WAV File	13/0...
7.wav	253 KB	WAV File	13/0...
8.wav	269 KB	WAV File	13/0...
9.wav	277 KB	WAV File	13/0...
10.wav	269 KB	WAV File	13/0...
H1 Gab Blue...		Folder	13/1...
1.wav	241 KB	WAV File	13/0...
2.wav	227 KB	WAV File	13/0...
3.wav	239 KB	WAV File	13/0...
4.wav	293 KB	WAV File	13/0...
5.wav	269 KB	WAV File	13/0...
6.wav	253 KB	WAV File	13/0...
7.wav	321 KB	WAV File	13/0...
8.wav	245 KB	WAV File	13/0...
9.wav	311 KB	WAV File	13/0...
10.wav	287 KB	WAV File	13/0...
H1 Gab Head...		Folder	13/1...
1.wav	253 KB	WAV File	13/0...
2.wav	263 KB	WAV File	13/0...
3.wav	271 KB	WAV File	13/0...

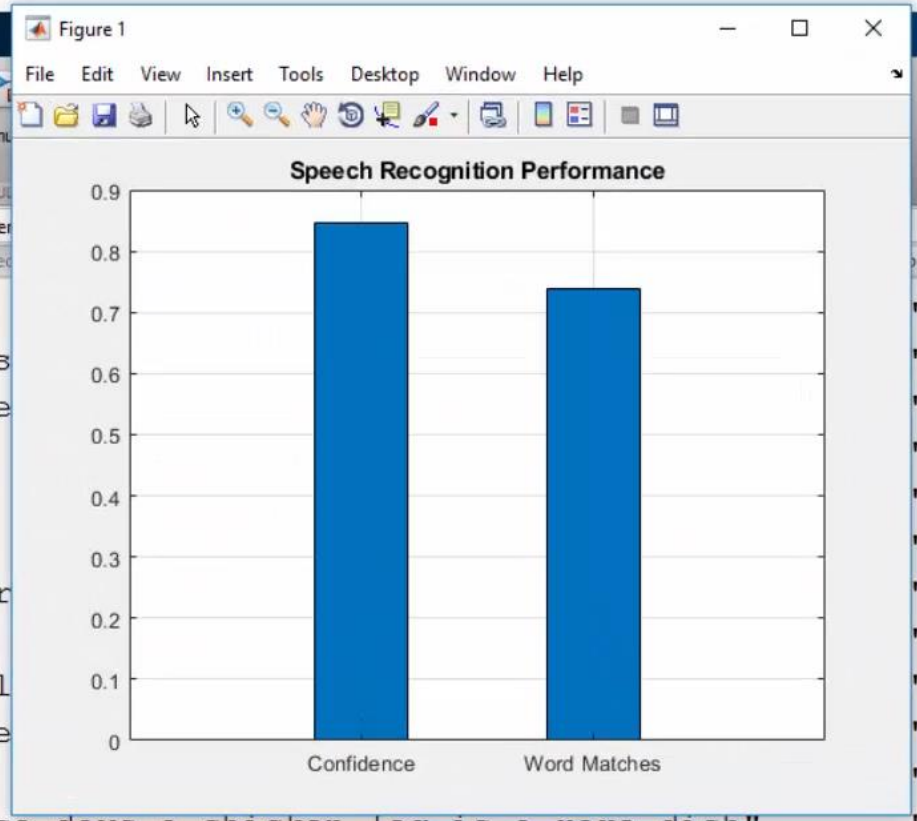
Details

Editor - assessSpeechRecognitionPerformance.m

Command Window

```
>> assessSpeechRecognitionPerformance
Sending "the birch canoe slid on the smooth planks"...
Sending " a large size in stockings is hard to sell"...
Sending "glue the sheet to the dark blue background"...
Sending "its easy to tell the depth of a well"...
Sending "these days a chicken leg is a rare dish"...
Sending "rice is often served in round bowls"...
Sending "the juice of lemons makes fine punch"...
Sending "the box was thrown beside the parked truck"...
Sending "the hogs were fed chopped corn and garbage"...
Sending "four hours of steady work faced us"...
Sending "the birch canoe slid on the smooth planks"...
```

Click and drag to move Editor...



APPS SHORTCUTS

Cleanup

Search Documen

Find Files
Compare
Open
Import Data
Save Workspace
New Variable
Open Variable
Clear Workspace
Analyze Code
Run and Time
Clear Commands
VARIABLE
CODE

Users ▸ gbunkhei ▸ Desktop ▸ AES Material ▸ Booth Demos ▸ Background ▸ Voice Inter

Command History | Current Folder

	Size	Type	Date M...
1		Folder	13/1...
Blue...		Folder	13/1...
	231 KB	WAV File	13/0...
	217 KB	WAV File	13/0...
	209 KB	WAV File	13/0...
	249 KB	WAV File	13/0...
	269 KB	WAV File	13/0...
	251 KB	WAV File	13/0...
	253 KB	WAV File	13/0...
	269 KB	WAV File	13/0...
	277 KB	WAV File	13/0...
	269 KB	WAV File	13/0...
Blue...		Folder	13/1...
	241 KB	WAV File	13/0...
	227 KB	WAV File	13/0...
	239 KB	WAV File	13/0...
	293 KB	WAV File	13/0...
	269 KB	WAV File	13/0...
	253 KB	WAV File	13/0...
	321 KB	WAV File	13/0...
	245 KB	WAV File	13/0...
	311 KB	WAV File	13/0...
	287 KB	WAV File	13/0...

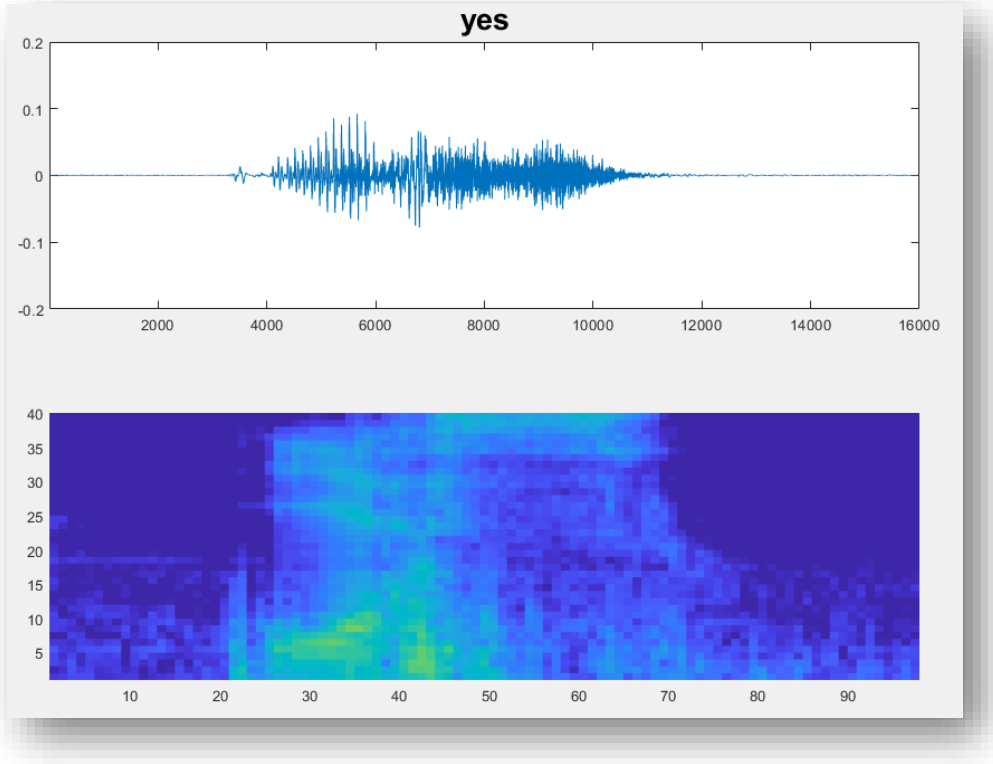
Editor - assessSpeechRec

```
"its
"thes
"rice
"the
"the
"the
"four
"the
" a l
"glue
"its
"these days a chicken leg is a rare dish"
"rice is often served in round bowls"
"the juice of lemons makes fine punch"
"the box was thrown beside the parked truck"
"the hogs were fed chopped corn and garbage"
"four hours of steady work faced us"
"the birch canoe slid on the smooth planks"
" a large size in stockings is hard to sell"
"glue the sheet to the dark blue background"
"its easy to tell the depth of a well"
"these days a chicken leg is a rare dish"
"rice is often served in round bowls"
"the juice of lemons makes fine punch"
```

ommand Window

```
"it's easy to tell the depth of a well"
"please days a chicken leg is a red dis
"rice is often served in round balls"
"the juice of lemons makes wine punch"
"the box was thrown beside the parked t
"the Hogs were fat chops corn and garba
"4 hours of steady work faces"
"the first canoe slid on the smooth pla
"a large size in stockings is hard to s
"do the sheets to the dark blue backgro
"excuse me to tell the depth of a well"
"these things of chicken leg is a redfi
"price is often served in ramen bowls"
"but use of lemons makes fine punch"
"the box was thrown beside the parts tr
"Bogs Boots at Shopko in garbage"
"4 hours of study worth-based office"
"the birth canal we slid on the smooth
"a large size in stockings is hard to s
"glue the sheet to the dark blue backgr
"it's easy to tell the depth of a well"
"these days a chicken leg is a red dish
"Ross is often served in round balls"
"the Jews of lemons make spine punch"
```

Speech Command Recognition with Deep Learning
(New Example)



- Train a Convolutional Neural Network (CNN) to recognize speech commands
- Work with Google's speech command dataset
- Leverage helper code for:
  - Reading and managing large datasets (New **audio datastore** prototype)
  - Transforming 1D signals into 2D images via perceptually-aware spectrograms (New **auditory spectrogram** prototype)
- Prototype trained network in real-time on live audio

To Learn More about Deep Learning with MATLAB

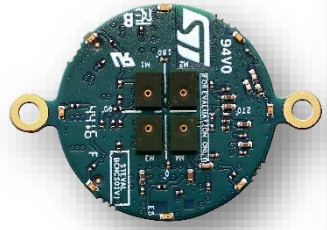
Demystifying Deep Learning

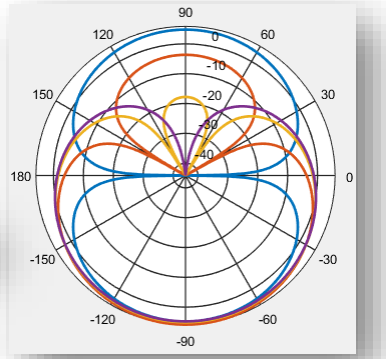
14:30–15:15

Deep learning can achieve state-of-the-art accuracy for many tasks considered algorithmically unsolvable using traditional machine learning, including classifying objects in a scene or recognizing optimal paths in an environment. Gain practical knowledge of the domain of deep learning and discover new MATLAB® features that simplify these tasks and eliminate the low-level programming. From prototype to production, you'll see demonstrations on building and training neural networks and hear a discussion on automatically converting a model to CUDA® to run natively on GPUs.

Dr. Amod Anandkumar,
Senior Team Lead,
MathWorks India

How Can I…





✓ Design a voice interface?

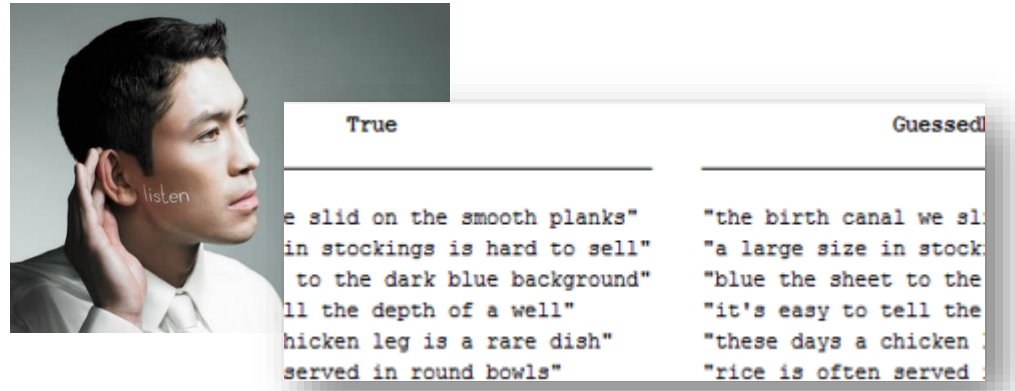
Phased Array System Toolbox

- Microphone Array Design
- Beamforming and Direction of Arrival Estimation

✓ Validate if my voice interface can work in real-life scenarios?

Audio System Toolbox

- Live Acquisition from hardware
- Leveraging Audio Plugins for performance improvement



✓ Test the performance of my system?

MATLAB

- Using Cloud Services for Speech Recognition
- Creating your own speech dataset

Signal Processing with MATLAB

This two-day course shows how to analyze signals and design signal processing systems using MATLAB®, Signal Processing Toolbox™, and DSP System Toolbox™.

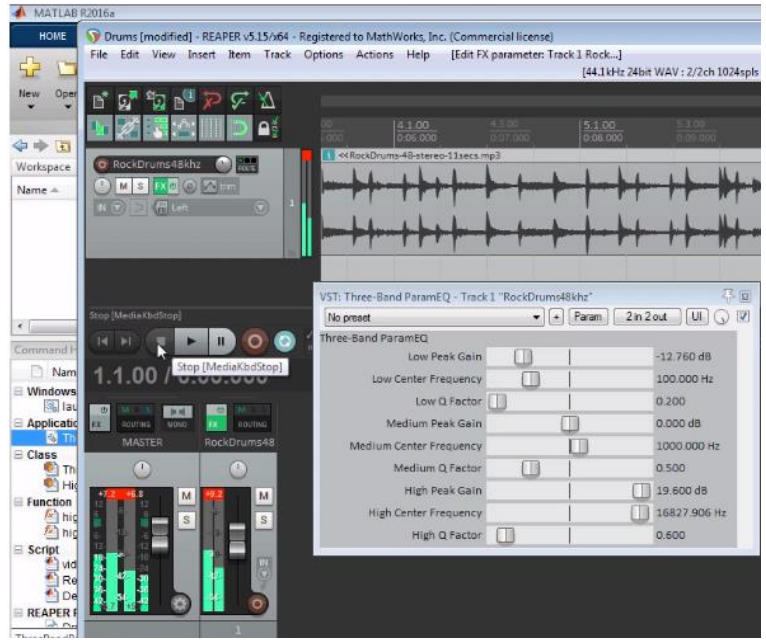
Topics include:

- Creating and analyzing signals
- Performing spectral analysis
- Designing and analyzing filters
- Designing multirate filters
- Designing adaptive filters

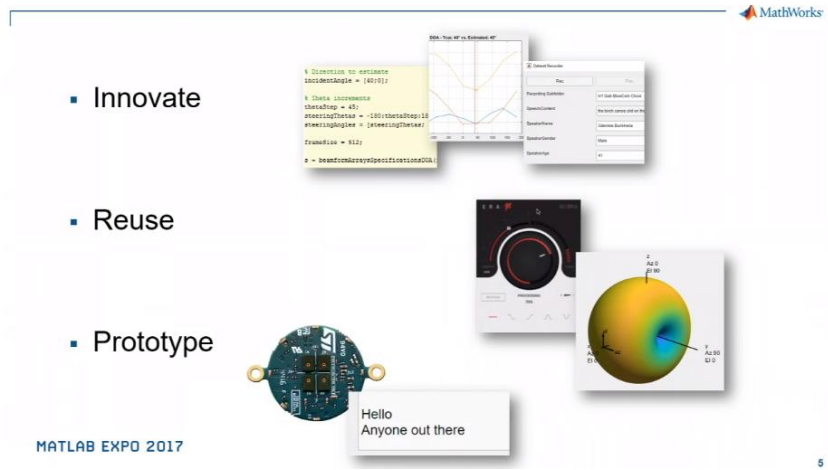
Call to Action

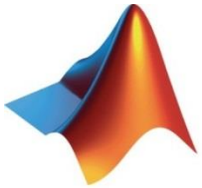
Videos and Webinars

- Real-time Audio Processing for Algorithm Prototyping and Custom Measurements



- Reusing and Prototyping Code to Accelerate Innovation: Smart Voice Interfaces





MathWorks®

Accelerating the pace of engineering and science

Speaker Details

Email: Vidya.Viswanathan@mathworks.in

LinkedIn:

www.linkedin.com/in/vidyaviswanathan

Contact MathWorks India

Products/Training Enquiry Booth

Call: 080-6632-6000

Email: info@mathworks.in

- **Share your experience with MATLAB & Simulink on Social Media**
  - Use #MATLABEXPO
  - I use #MATLAB because…… Attending #MATLABEXPO
  - Examples
    - I use #MATLAB because it helps me be a data scientist! Attending #MATLABEXPO
    - Learning new capabilities in #MATLAB and #Simulink at #MATLABEXPO.

- **Share your session feedback:**

  Please fill in your feedback for this session in the feedback form