ADDING VOICE RECOGNITION TO YOU ANDROID APP

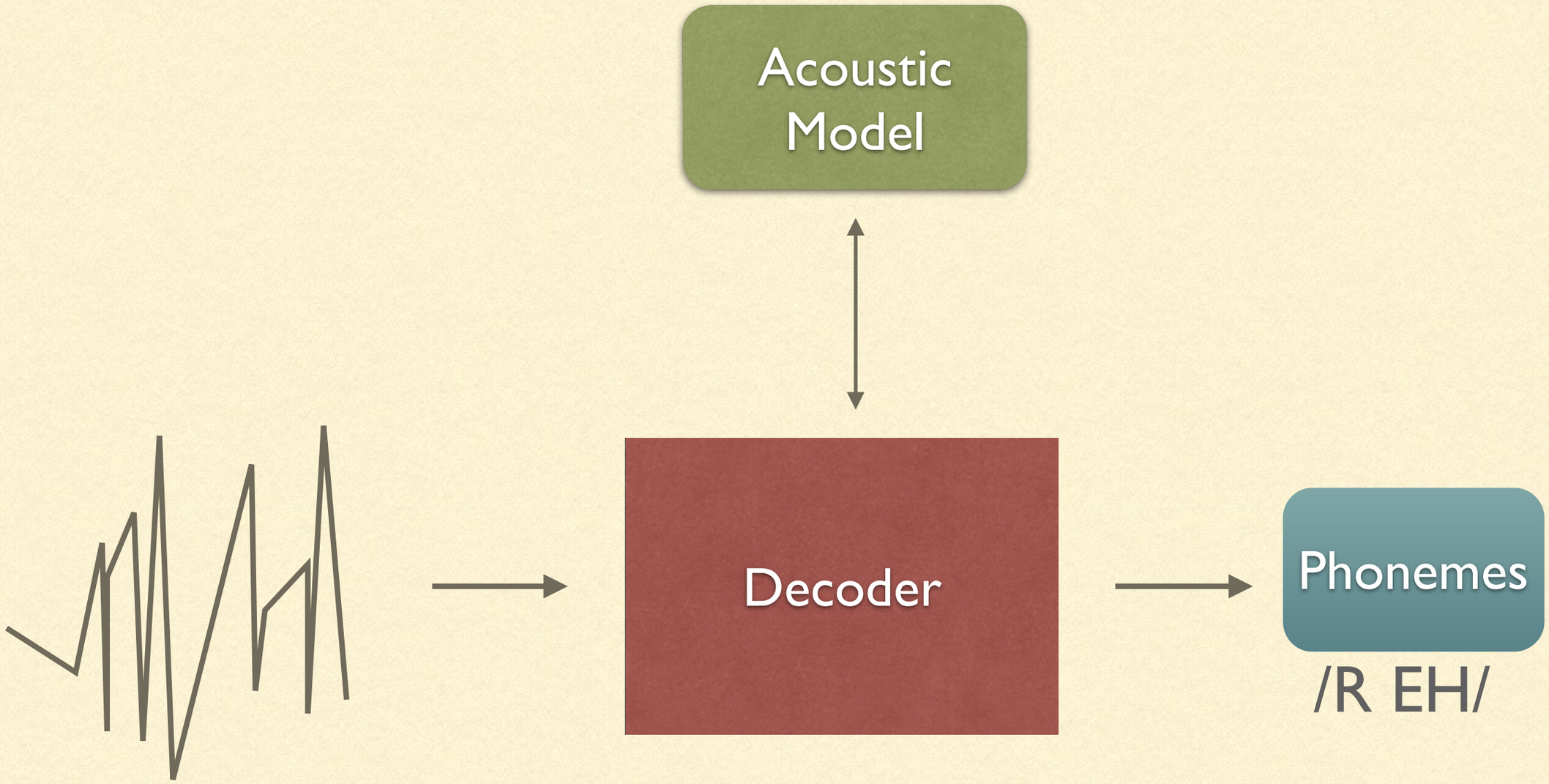
MANUEL NAKAMURAKARE
Mobile Software Engineer
Handsfree Learning

DEMO

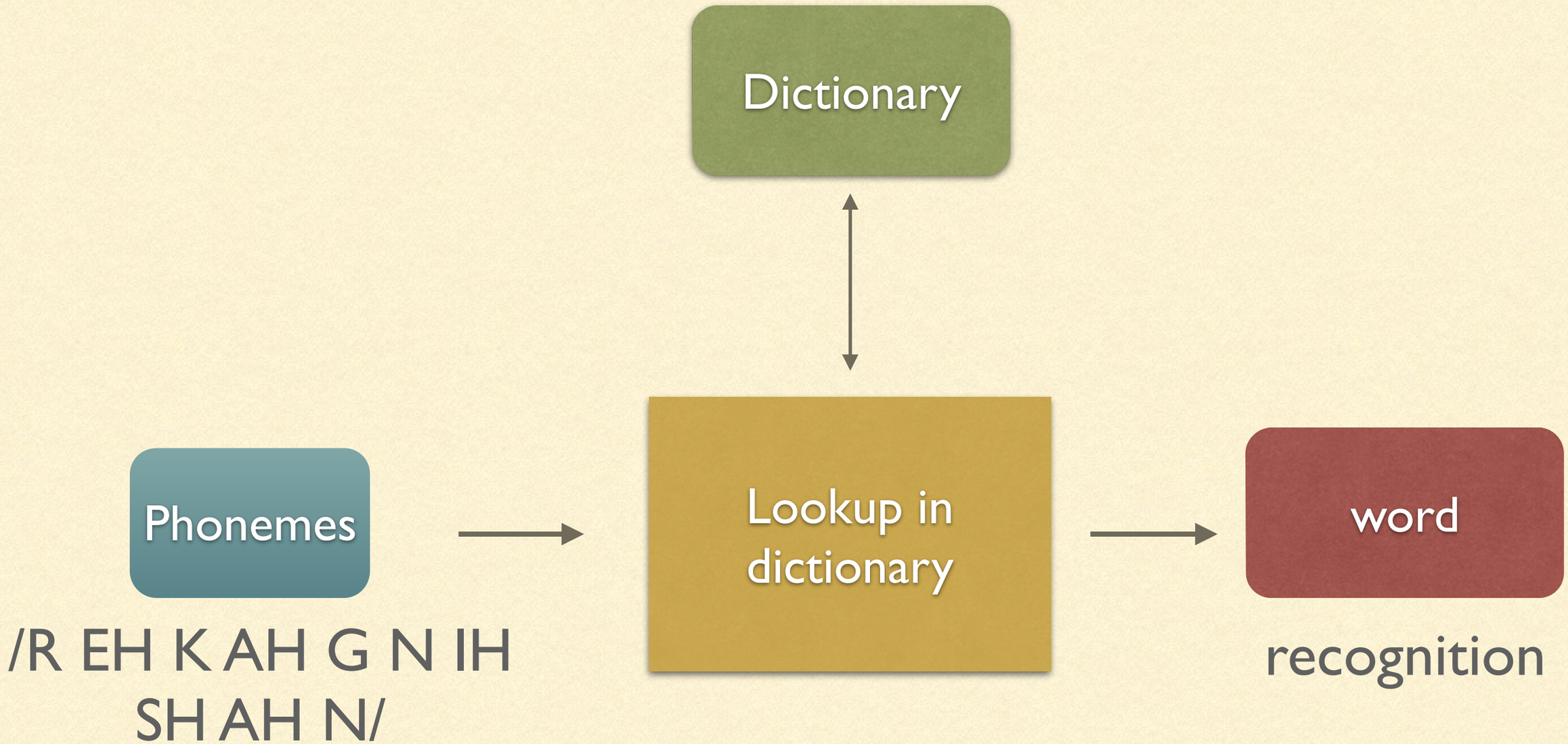
WHY DO I CARE?

- Because it's cool!
- There are situations where using your hands can be impossible or dangerous
- Improve user experience
- Disabled people

THE BASICS



THE BASICS


Dictionary
Phonemes
/R EH K AH G N IH SH AH N/
Lookup in dictionary
word
recognition

BUT… IT LOOKS VERY HARD!

- I will need to be a speech recognition expert, and I already have enough with Android!
- I will need to learn about signal processing and stuff !!

MOST POPULAR API'S

- CMU's Pocket Sphinx
- Google Speech Api
- Nuance
- Others: Wit AI, AT&T, etc

CMU'S POCKET SPHINX

- On device speech recognizer
- Limited dictionary
- Small acoustic model

GOOGLE SPEECH API

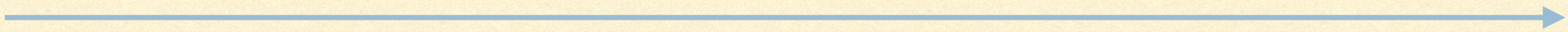
- Cloud based speech recognizer
- Free
- Bigger acoustic model
- Bigger dictionary
- Constantly being trained

NUANCE SPEECH API

- Cloud based speech recognizer
- Free version has limited number of transactions
- Slower compared to Google Speech API
- Bigger acoustic model
- Bigger dictionary
- Constantly being trained

CONTINUOUS LISTENING

take picture take picture take picture



CHALLENGES

CHALLENGES

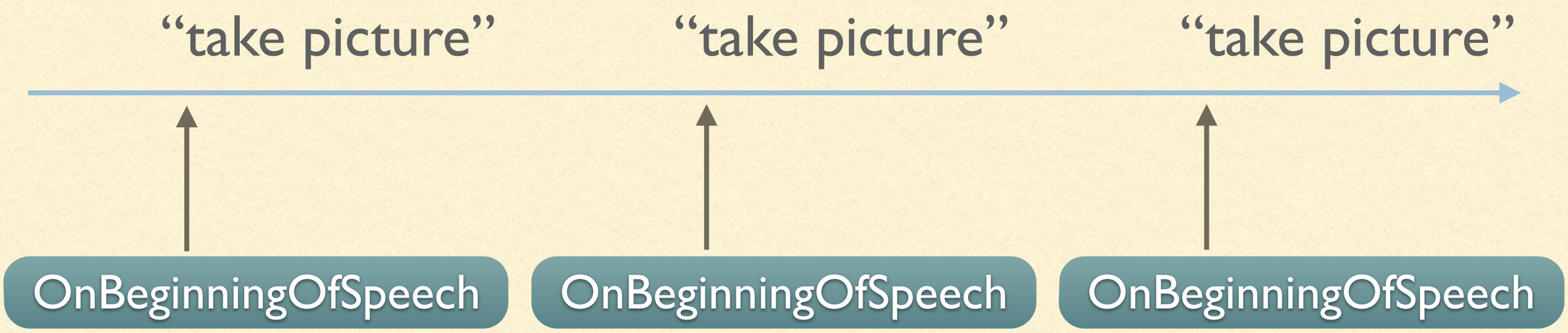
Pocket Sphinx:

- not very accurate

CHALLENGES

Google Speech API:

- beep at start (good and bad)

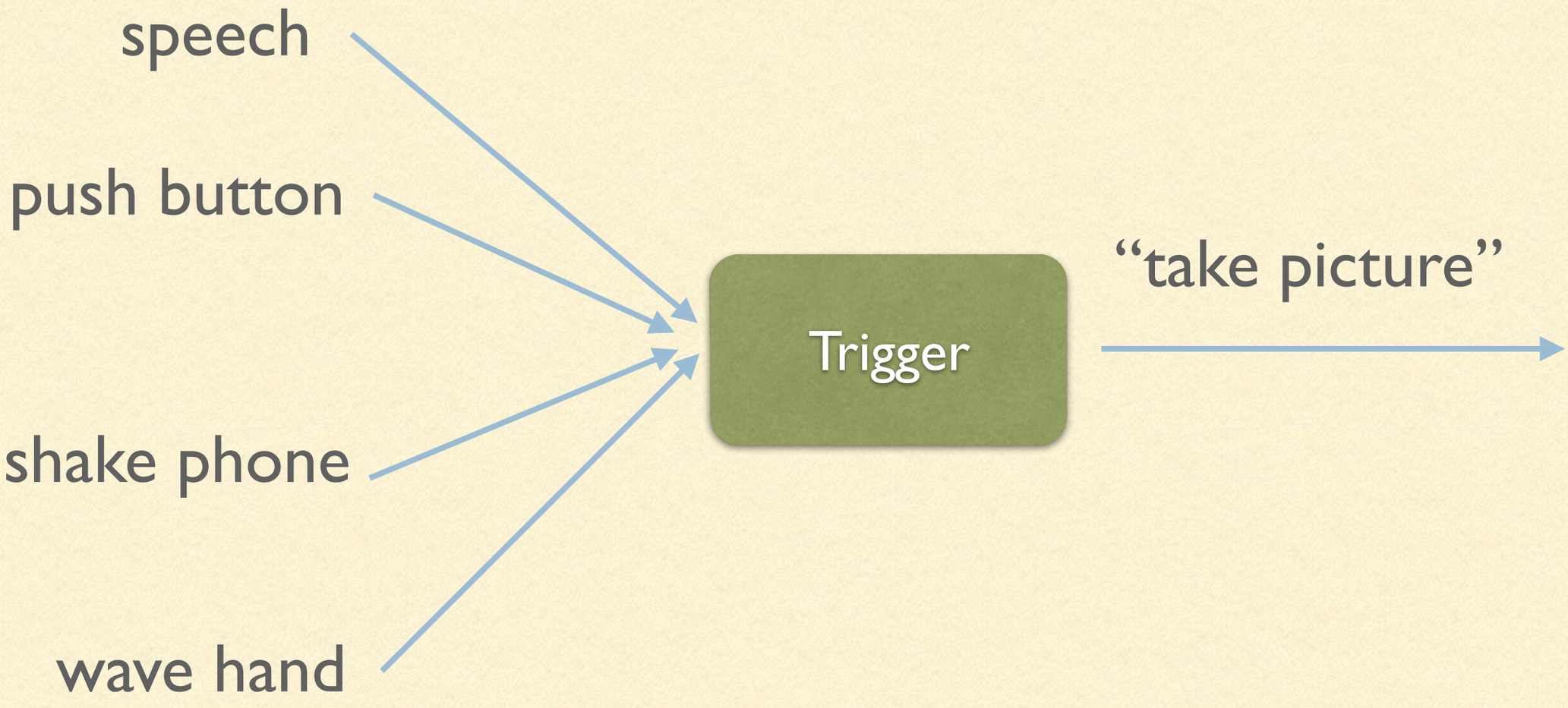


CHALLENGES

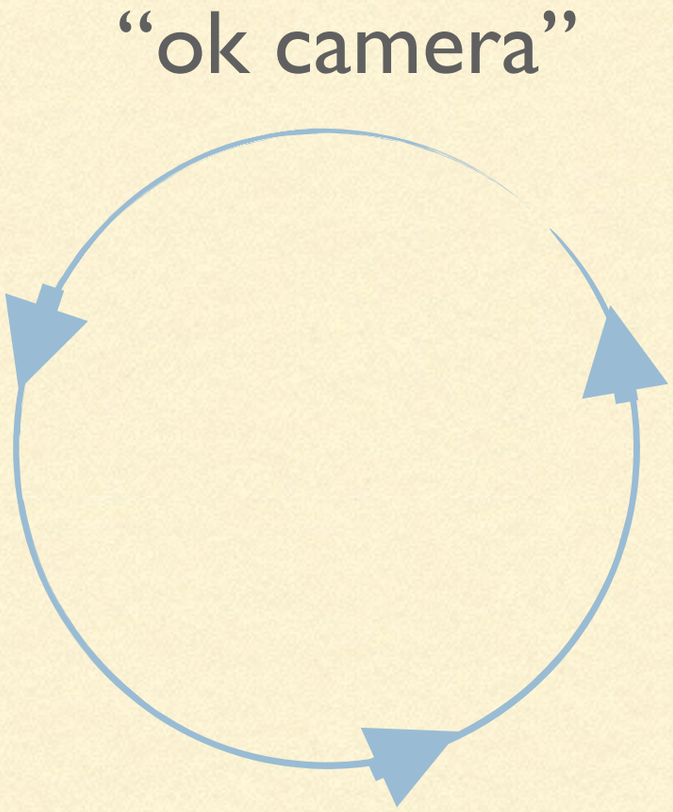
Nuance Speech API:

- start and stop of speech not necessarily in sync

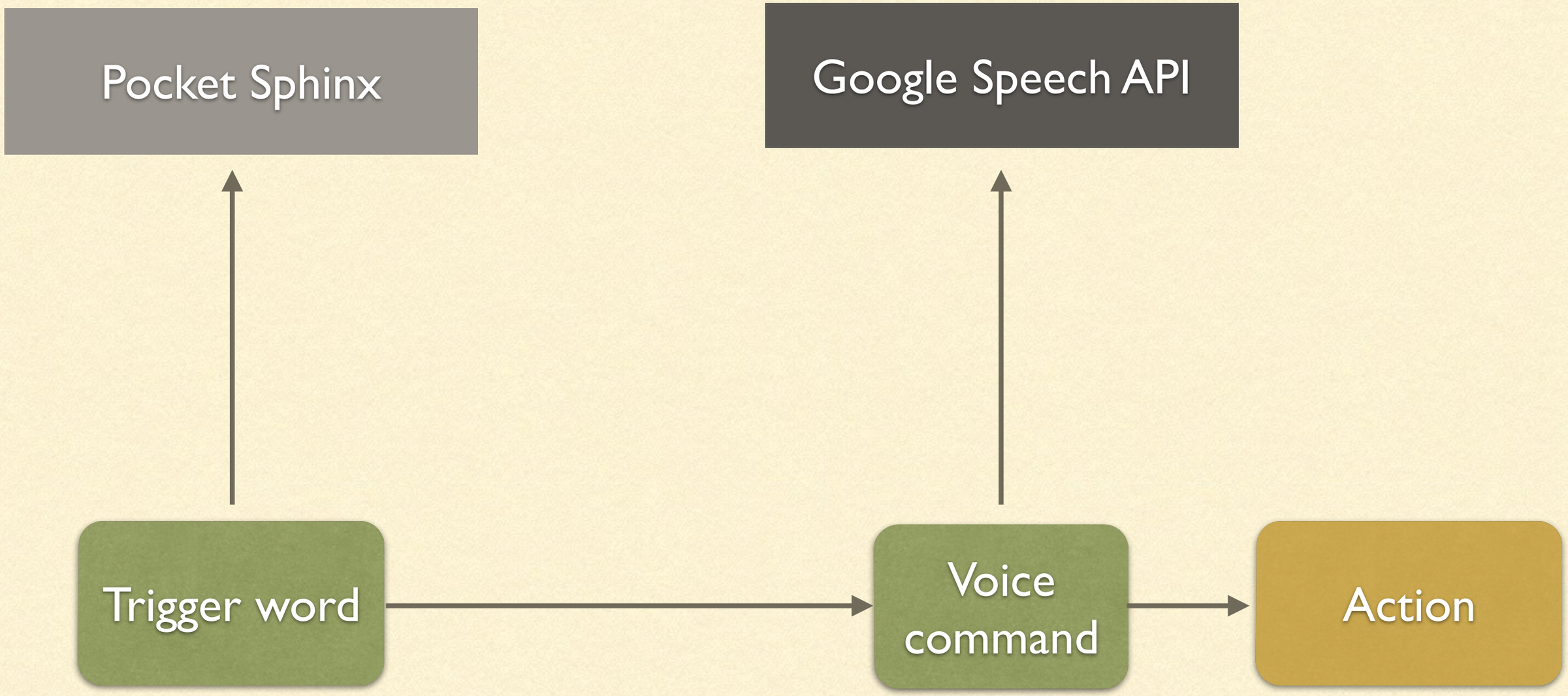
NON-CONTINUOUS LISTENING



MIXED APPROACH



ARCHITECTURE OVERVIEW



ANY QUESTIONS AT THIS POINT?

SHOW ME THE CODE

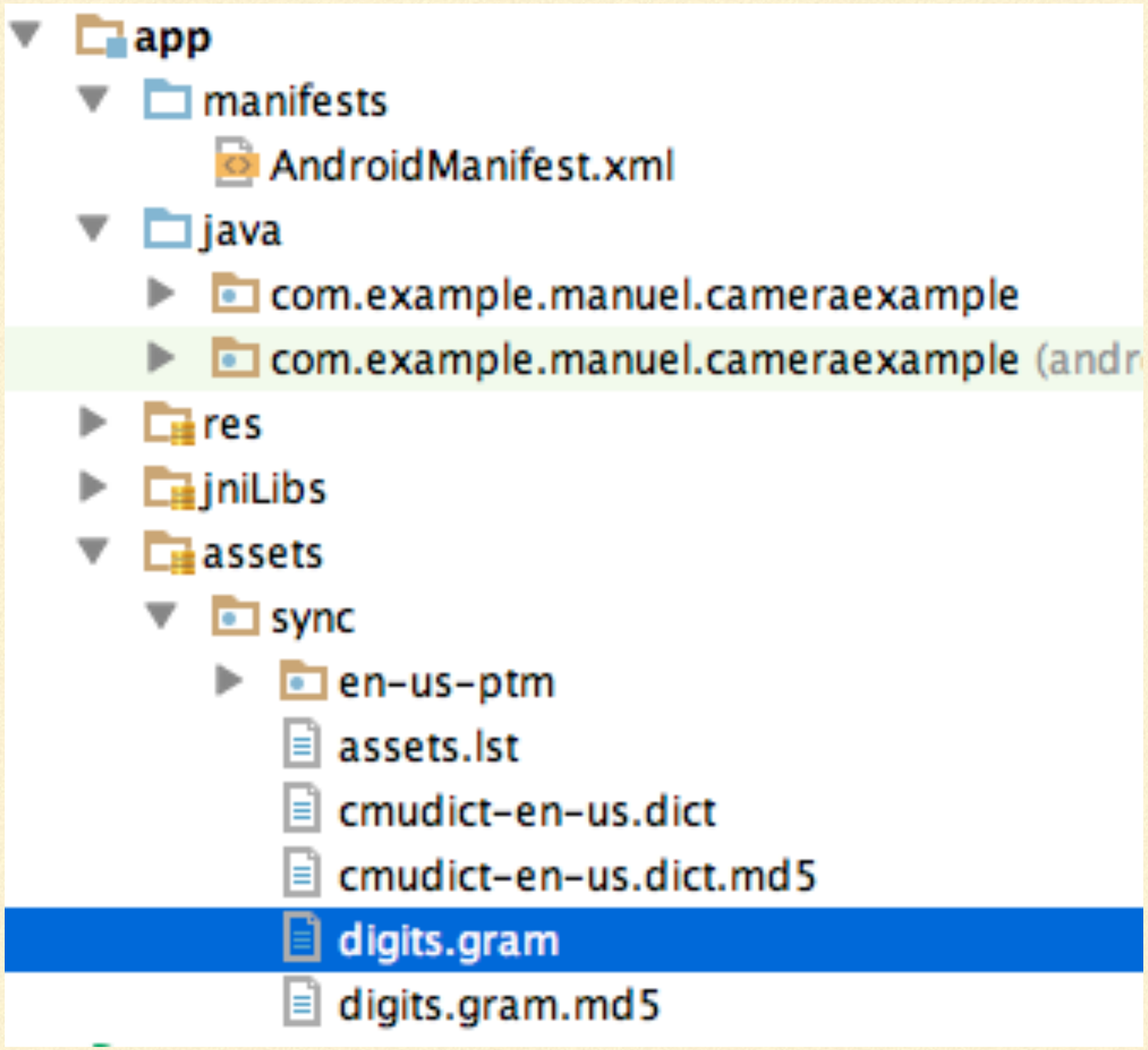
- All the code is in:

https://github.com/manask88/speechrecognitiondemoapp

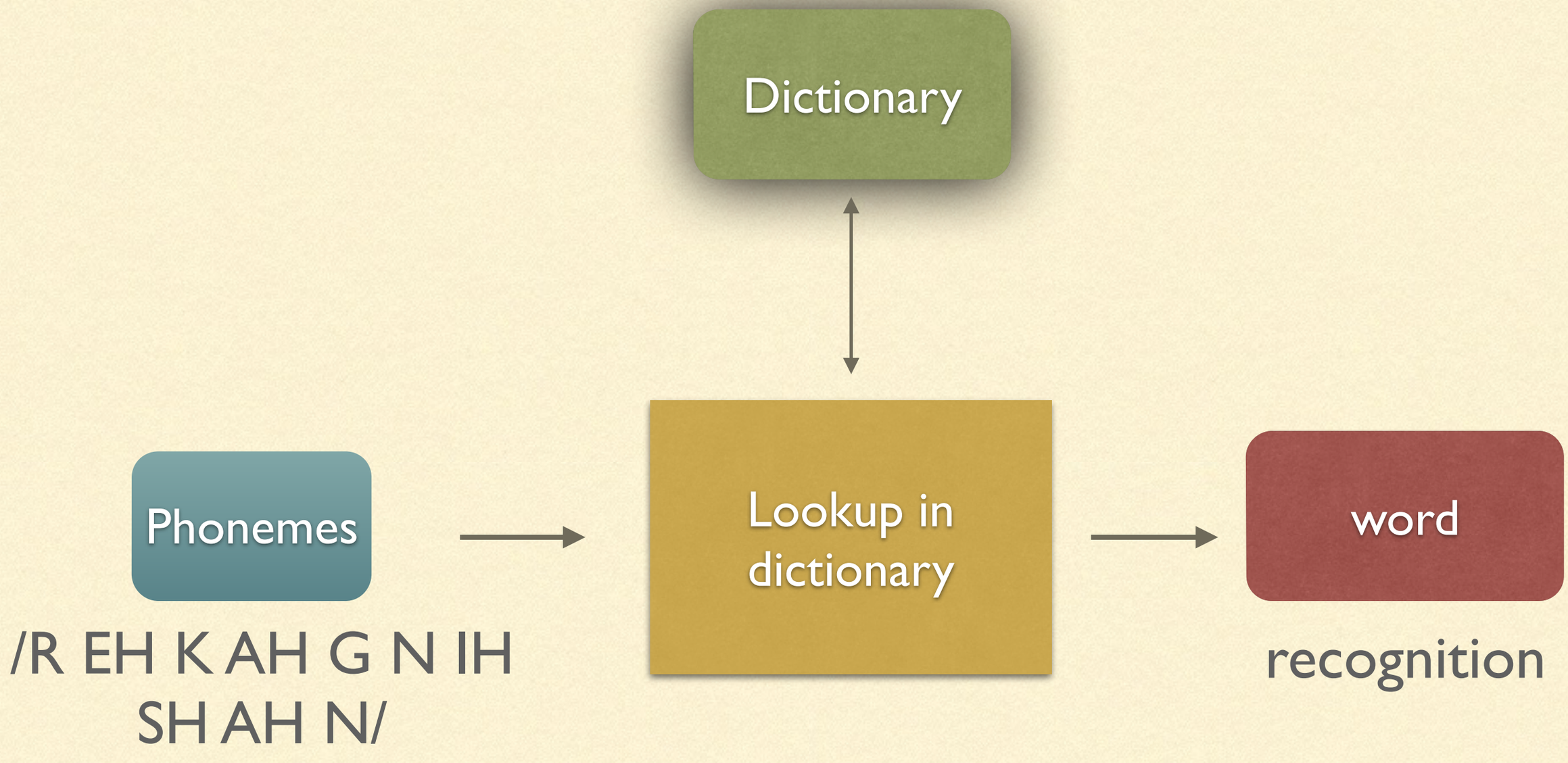
POCKET SPHINX INIT (1/2)

```
 Assets assets = new Assets(mContext);

//Performs the synchronization of assets in the application and
external storage
 File assetDir = assets.syncAssets();

//Creates a new speech recognizer builder with default
configuration
SpeechRecognizerSetup speechRecognizerSetup = defaultSetup();

speechRecognizerSetup.setAcousticModel(new File(assetDir, "en-
us-ptm"));

speechRecognizerSetup.setDictionary(new File(assetDir, "cmudict-
en-us.dict"));
```

ASSETS


app
manifests
AndroidManifest.xml
java
com.example.manuel.cameraexample
com.example.manuel.cameraexample (andr
res
jniLibs
assets
sync
en-us-ptm
assets.lst
cmudict-en-us.dict
cmudict-en-us.dict.md5
digits.gram
digits.gram.md5

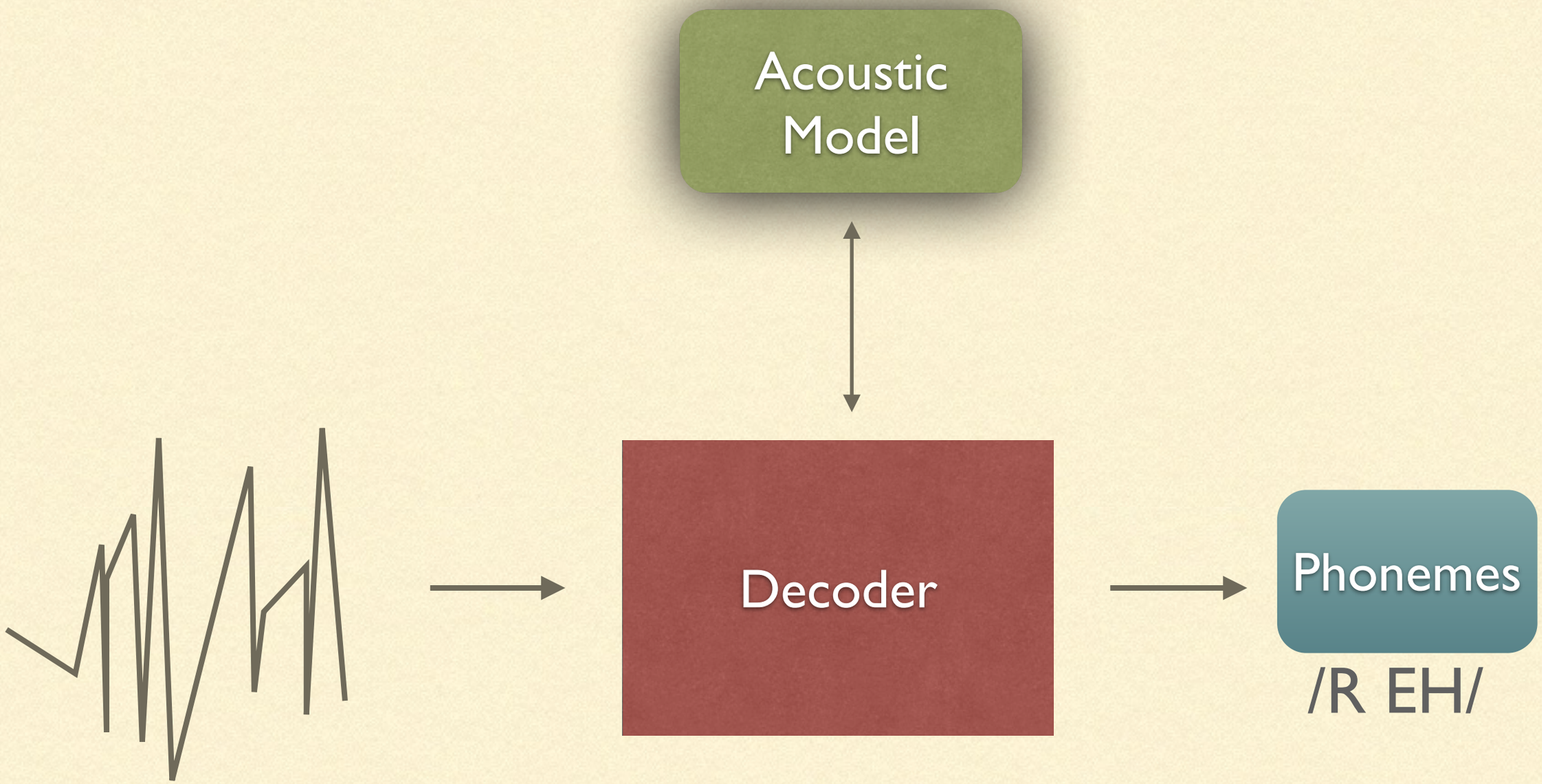
DICTIONARY


Dictionary
Phonemes
/R EH K AH G N IH SH AH N/
Lookup in dictionary
word
recognition

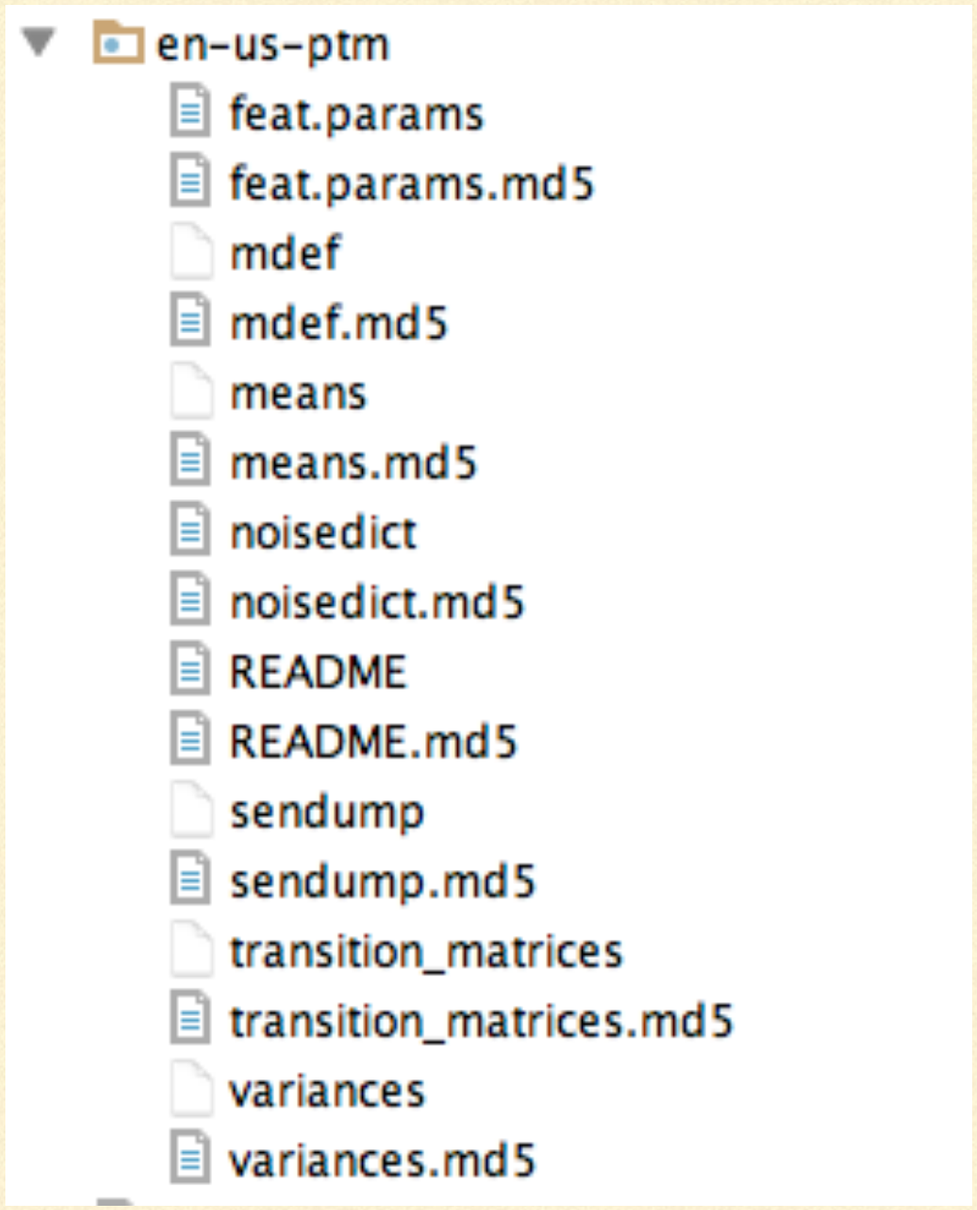
DICTIONARY

- aborts AH B AO R T S
- abound AH B AW N D
- abounded AH B AW N D IH D
- abounding AH B AW N D IH NG
- abounds AH B AW N D Z
- about AH B AW T

ACOUSTIC MODEL



ACOUSTIC MODEL



POCKET SPHINX INIT (2/2)

```
// Threshold to tune for keyphrase to balance between false
alarms and misses
speechRecognizerSetup.setKeywordThreshold(1e-45f);

//Creates a new SpeechRecognizer object based on previous set
up.
mPocketSphinxRecognizer =
speechRecognizerSetup.getRecognizer();

mPocketSphinxRecognizer.addListener(new
PocketSphinxRecognitionListener());

// Create keyword-activation search.
mPocketSphinxRecognizer.addKeyphraseSearch(KWS_SEARCH,
KEYPHRASE);
```

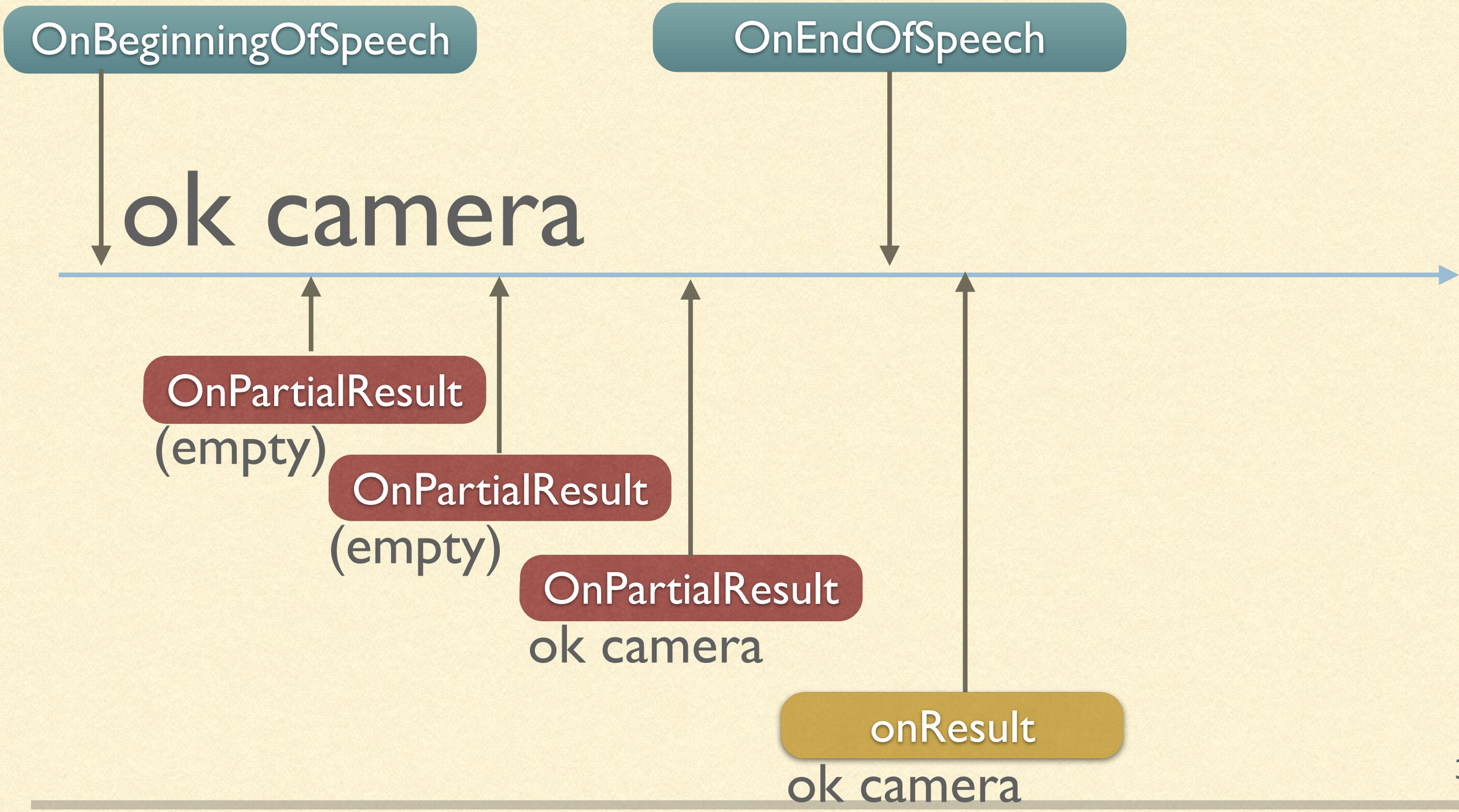
POCKET SPHINX LISTENER (1/2)

```
@Override
public void onBeginningOfSpeech() {
        }
@Override
public void onPartialResult(Hypothesis hypothesis) {
        if (hypothesis == null)
            return;

        String text = hypothesis.getHypstr();
        if (text.equals(KEYPHRASE)) {
            mGoogleSpeechRecognizer.
                        startListening(mSpeechRecognizerIntent);
            mPocketSphinxRecognizer.cancel();
        }
    }
```

POCKET SPHINX LISTENER (2/2)

```
@Override
public void onResult(Hypothesis hypothesis) {
        }

@Override
public void onEndOfSpeech() {
        }
@Override
public void onError(Exception error) {
        }

@Override
public void onTimeout() {
        }
```

SPEECH EVENTS IN DETAIL (KEYWORD SEARCH)



GOOGLE SPEECH INIT (1/2)

```
mAudioManager = (AudioManager) mContext
                .getSystemService(Context.AUDIO_SERVICE);
mGoogleSpeechRecognizer = android.speech.SpeechRecognizer
                          .createSpeechRecognizer(mContext);

mGoogleSpeechRecognizer.setRecognitionListener(new
                        GoogleRecognitionListener());

mSpeechRecognizerIntent = new Intent(
     RecognizerIntent.ACTION_RECOGNIZE_SPEECH);
```

GOOGLE SPEECH INIT (2/2)

```
mSpeechRecognizerIntent.putExtra(
                RecognizerIntent.EXTRA_LANGUAGE_MODEL,
                RecognizerIntent.LANGUAGE_MODEL_FREE_FORM);

mSpeechRecognizerIntent.putExtra(
                RecognizerIntent.EXTRA_CALLING_PACKAGE,
                mContext.getPackageName());
```

GOOGLE SPEECH LISTENER (1/2)

```
@Override
public void onResults(Bundle results) {
    if ((results != null) &&
         results(
         android.speech.SpeechRecognizer.
         RESULTS_RECOGNITION))        {
        //…(next slide)
    }

    mPocketSphinxRecognizer
                          .startListening(KWS_SEARCH);
}
```

GOOGLE SPEECH LISTENER (2/2)

```
ArrayList<String> heard = results(android.speech.

SpeechRecognizer.RESULTS_RECOGNITION);
float[] scores =
results.getFloatArray(android.speech.

SpeechRecognizer.CONFIDENCE_SCORES);

if (mOnResultListener!=null){
        mOnResultListener.OnResult(heard);
}
```

RESULT CONTENTS

```
onResultsheard:take picture confidence:0.861010
onResultsheard:take a picture confidence:0.0
onResultsheard:fake picture confidence:0.0
onResultsheard:they take picture confidence:0.0
onResultsheard:take it your confidence:0.0
```

ANY QUESTIONS AT THIS POINT?

LET'S GET OUT HANDS A LITTLE BIT DIRTY!

- Get the project from:

https://github.com/manask88/workshopdemoapp

- It’s an camera demo app.
- PocketSphinx and Google Speech API are already integrated.
- Write code necessary to take pictures using your voice

ANALYZE RECOGNIZED WORDS

```
@Override
        public void onResults(Bundle partialResults) {
            if ((partialResults != null)
                    && partialResults
                            .containsKey(android.speech.SpeechR
ecognizer.RESULTS_RECOGNITION)) {
                    //…

                    //analyze list of words here
                }

mPocketSphinxRecognizer.startListening(KWS_SEARCH);
            }
```

TRIGGER ACTION

- Use callbacks, handlers, etc to communicate action to Activity
- To take a picture, use the following line of code in the Activity:

```
mCamera.takePicture(null,null,mPictureCallBack);
```

NEXT STEPS

INTEGRATE SPEECH RECOGNITION APIS TO YOUR APP

1. Install the Pocket Sphinx library in your project
2. Initialize Pocket Sphinx and create a Listener for its events
3. Initialize Google Speech API and create a Listener for its events
4. Communicate Google Speech API results to your app

How to add Pocket Sphinx and Google Speech API wiki:

https://github.com/manask88/wiki/wiki/Speech-Recognition-Tutorial

If you have any questions or feedback, please contact me at:

me@manuelnakamurakare.com

BACKUP SLIDES

SPEECH EVENTS IN DETAIL (NO KEYWORD SEARCH)