

# Understanding Impaired Speech

Kobi Calev , Morris Alper  
January 2016

**Voiceitt**



# Our Problem Domain

- We deal with **phonological disorders**
- They may be either
  - resonance or phonation
  - physiological or neural
  - developmental or adult onset
- Must be consistent (e.g. not stuttering)
- Examples:
  - aphasia (e.g. caused by stroke)
  - dysarthria (e.g. caused by CP)
  - apraxia (e.g. childhood apraxia of speech)

# Example Recordings

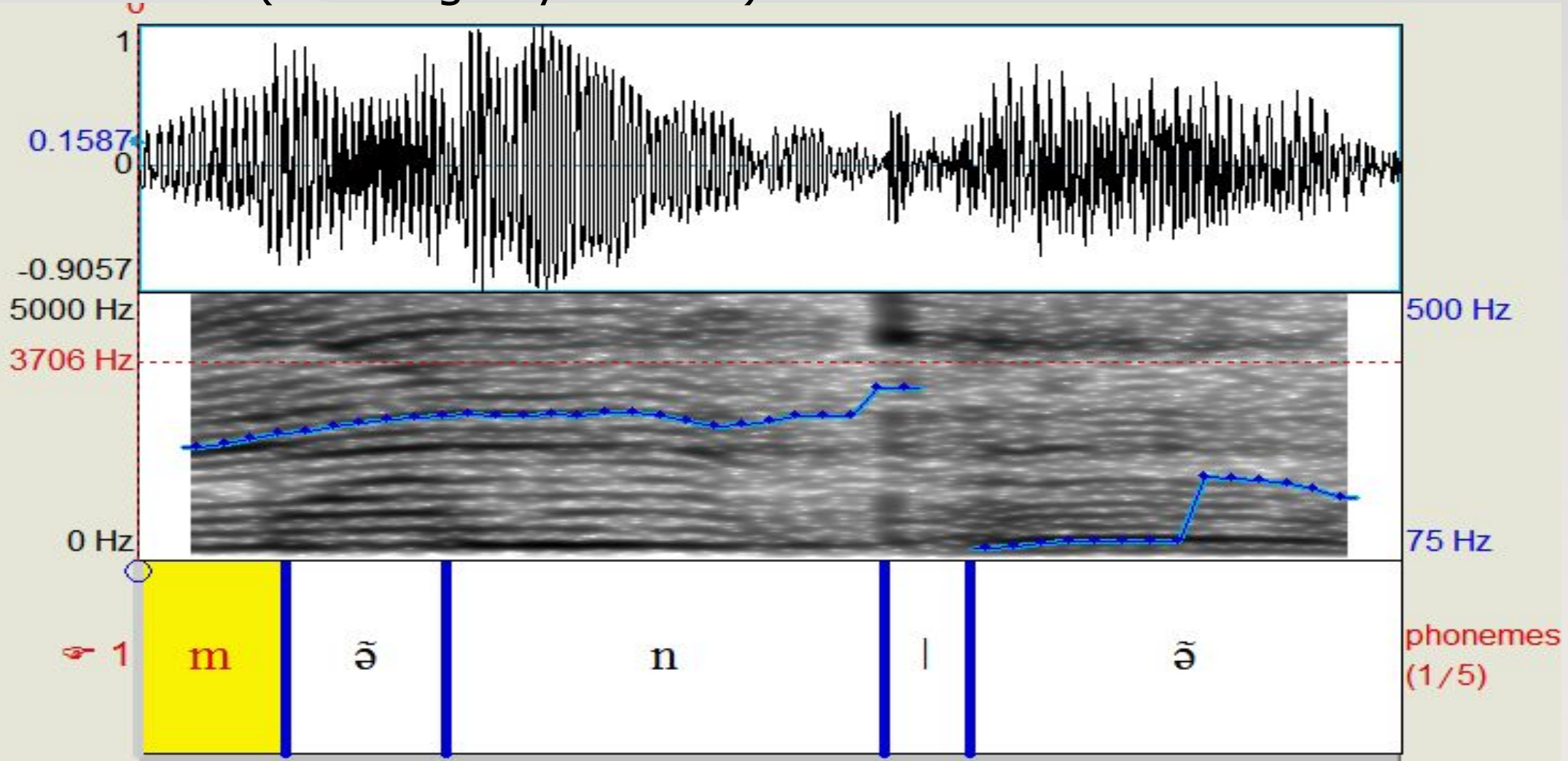
## Cerebral Palsy

- Stiff muscles
- Poor respiratory control
- “Risk of learned helplessness and becoming passive communicators”
- Not indicative of intellectual ability

## DiGeorge Syndrome

- Palatal abnormalities
- Causes hyper-nasalization

# “Window” (DiGeorge syndrome) with our extIPA annotation



Hyper-nasalized, “d” realized as a click

# Speech production

**Respiration** – control of breath during speech.

**Phonation** – vibration of vocal cords.

**Resonation** – shaping cavities of speech mechanism.

**Articulation** – coordinating the movement of speech organs.

If any of these are impaired, it leads to a phonological speech disorder.

# Modes of phonation

- Speaking
- Whispering
- Singing
- Shouting
- Whistling

Standard Speech-Recognition breaks in all modes except for “Speaking”.

# Disordered Phonetics

Extended IPA (extIPA): phonemes that only occur in disordered speech

- velopharyngeal fricative /fŋ/, bilabial percussive /w̥/, etc.
- various diacritics
  - e.g. whistled articulation, sliding articulation, ingressive airflow

Unusual substitutions (e.g. plosives → clicks)

These are informative phonemes in the Speaker's idiolect, even if not contained in the "target" host-language (eg, standard English, Hebrew, etc)

# Existing Solutions for population

Existing solutions *avoid speech*

Speech boards , head tracking , eye tracking , sip and puff.

Limitations:

- Slow
- Mobility limitations
- Do not encourage use of speech; communicative feedback loop

*Can speech recognition be used for this target population?*



# Existing Speech Recognition systems

- Requires masses of speech data, preferably transcribed and time-aligned
- Improvement focuses on data-mining scripts (ie, how to feed the system with more data)
- Brute-force - no linguistic model assumed
  - Simple and regular
  - Double edged sword [less heuristic]
- What makes these masses of data necessary?
  - Arbitrary time segmentation
    - Informative transient information get smeared later to be “recovered” statistically

# Our Solution Domain

**Insight:** People can adapt to understand disordered speech

## Speech recognition:

- A speech-to-speech system.
- Where *phonology* is speaker dependent.
- No pre-given training data
- “Standard” speech recognition doesn’t work (ie, HMMs etc requires data bootstrapping)
- Similar to working with **Under-Resourced Languages** / “Zero Speech” settings

# Additional advantages of our solution

- Unsupervised extraction of specific user phonological units
  - Domain-Adaptation: Into pre-trained language
  - Speech-Therapist
- Data acquisition: No existing databases for impaired speech

Possible future directions with same technology:

- Accent adaptation / analysis
- Bootstrapping speech recognition for under-resourced languages

# Our Solution

## Adaptive speech recognition.

### Advantages:

- Encourages use of speech, leading to improvement
- Data collection: no extant corpus of impaired speech
- Useful diagnostic tool for speech pathologists
- Potential uses in other areas of zero-resource speech recognition

# Zero-Resource Speech Processing

“Unsupervised linguistic unit discovery”.

- Keyword spotting
- Summarization of audio
- Voice dialing
- “Bootstrapping” speech recognition (e.g. accent adaptation)

New application: impaired speech.

Other possible uses: whispered speech, singing, or any other mode

*Come in!* WE ARE  
**HIRING**

## Voiceitt

“The mission of Voiceitt is to create innovative solutions that helps aid the lives of people with disabilities by fostering independence and social inclusion and thus dramatically improving their quality of life.”

Positions available in algorithmics and signal processing

<http://www.voiceitt.com/>

<http://www.voiceitt.com>



# Links

- <http://www.zerospeech.com> - The Zero Resource Speech Challenge
- <http://www.mica.edu.vn/sltu/> - International Workshop on Spoken Language Technologies for Under-resourced Languages
- [https://en.wikipedia.org/wiki/Extensions\\_to\\_the\\_International\\_Phonetic\\_Alphabet](https://en.wikipedia.org/wiki/Extensions_to_the_International_Phonetic_Alphabet)

# Bibliography

- The Handbook of Language and Speech Disorders [ Wiley-Blackwell ] / Müller ,J. Ball
- Whistled Languages A Worldwide Inquiry on Human Whistled Speech [ springer] / J. Meyer

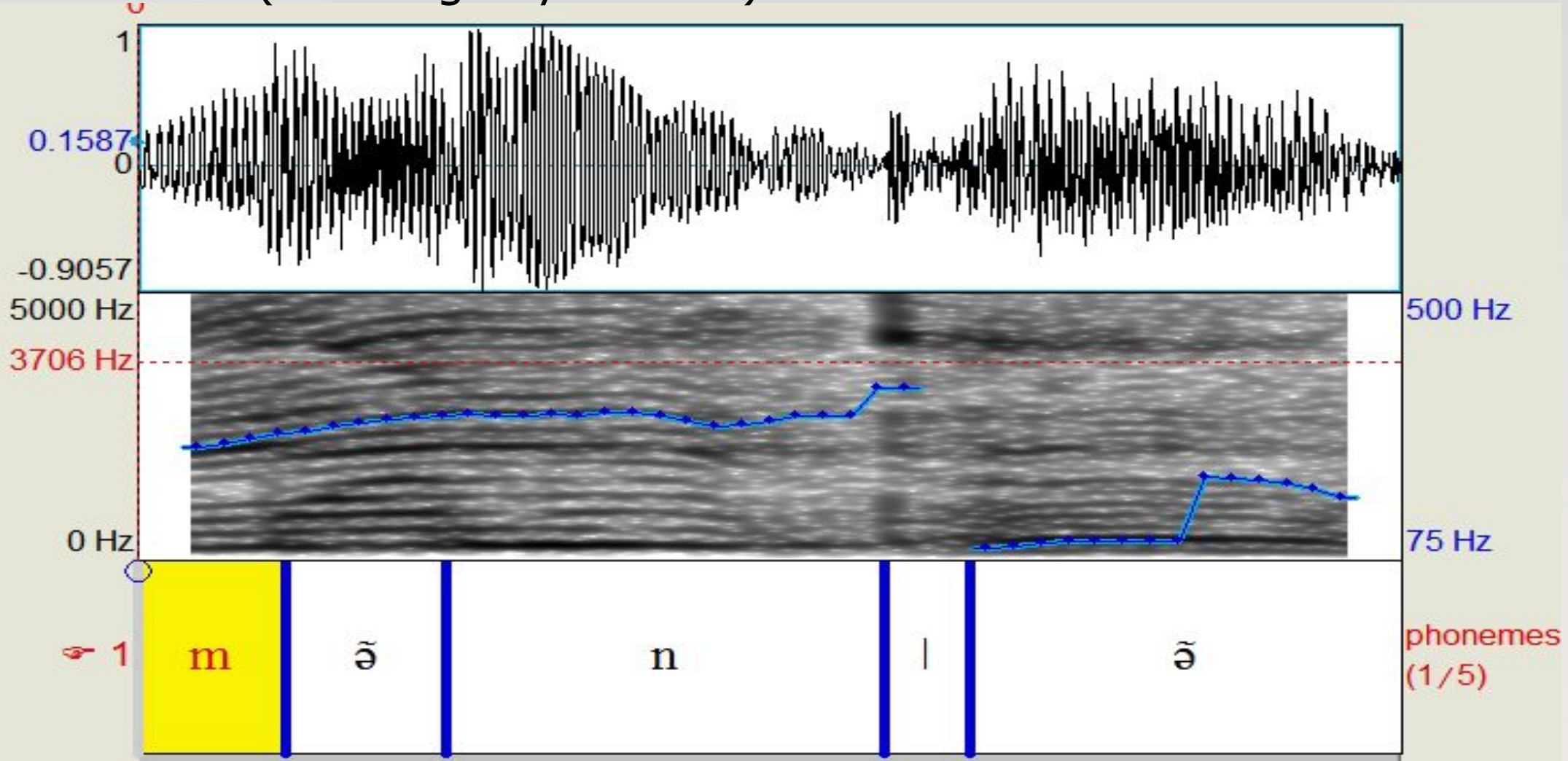


# Thank You!

# Questions?

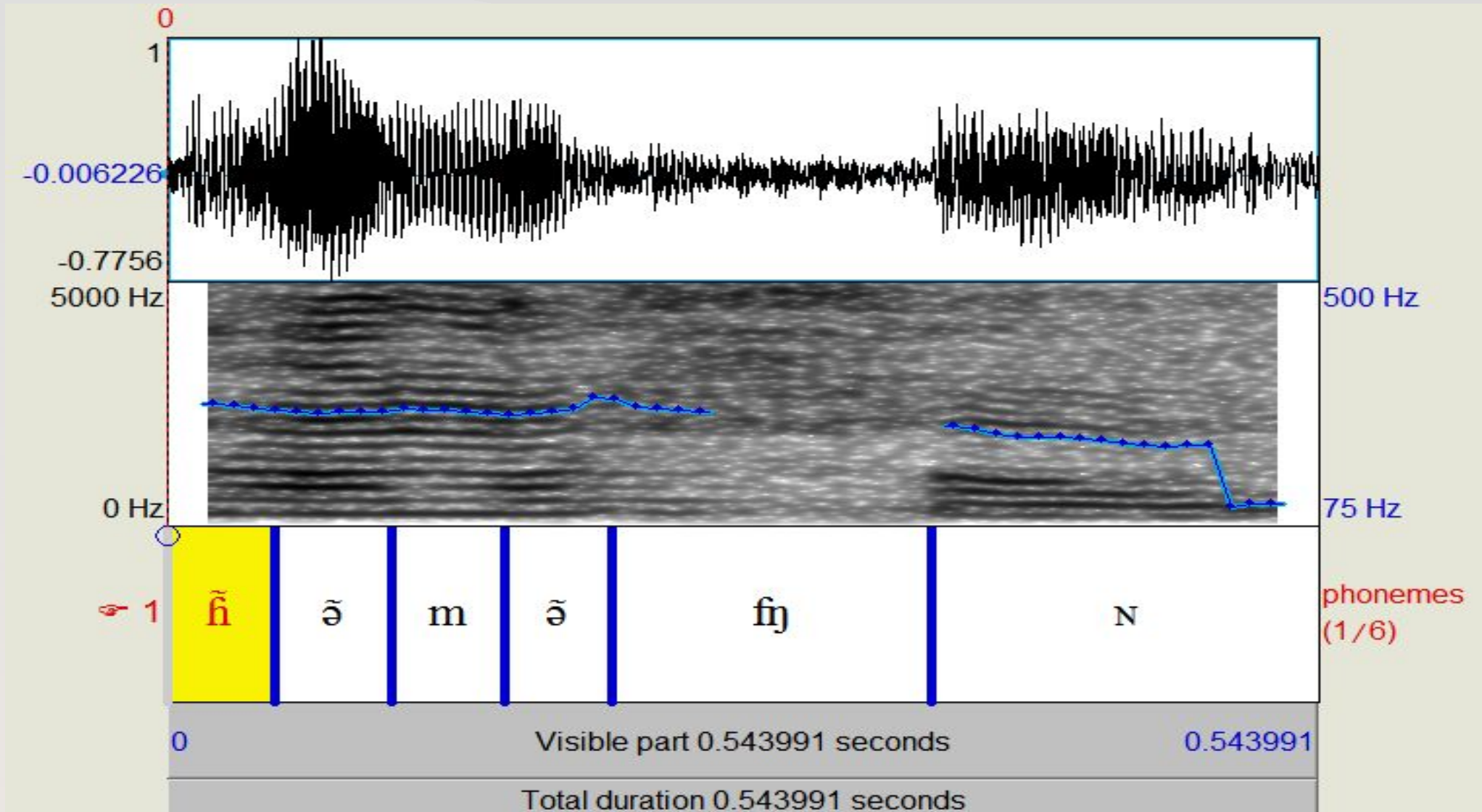
Feel free to contact us at  
[kobic@voiceitt.com](mailto:kobic@voiceitt.com)  
[malper@voiceitt.com](mailto:malper@voiceitt.com)

# “Window” (DiGeorge syndrome) with our extIPA annotation



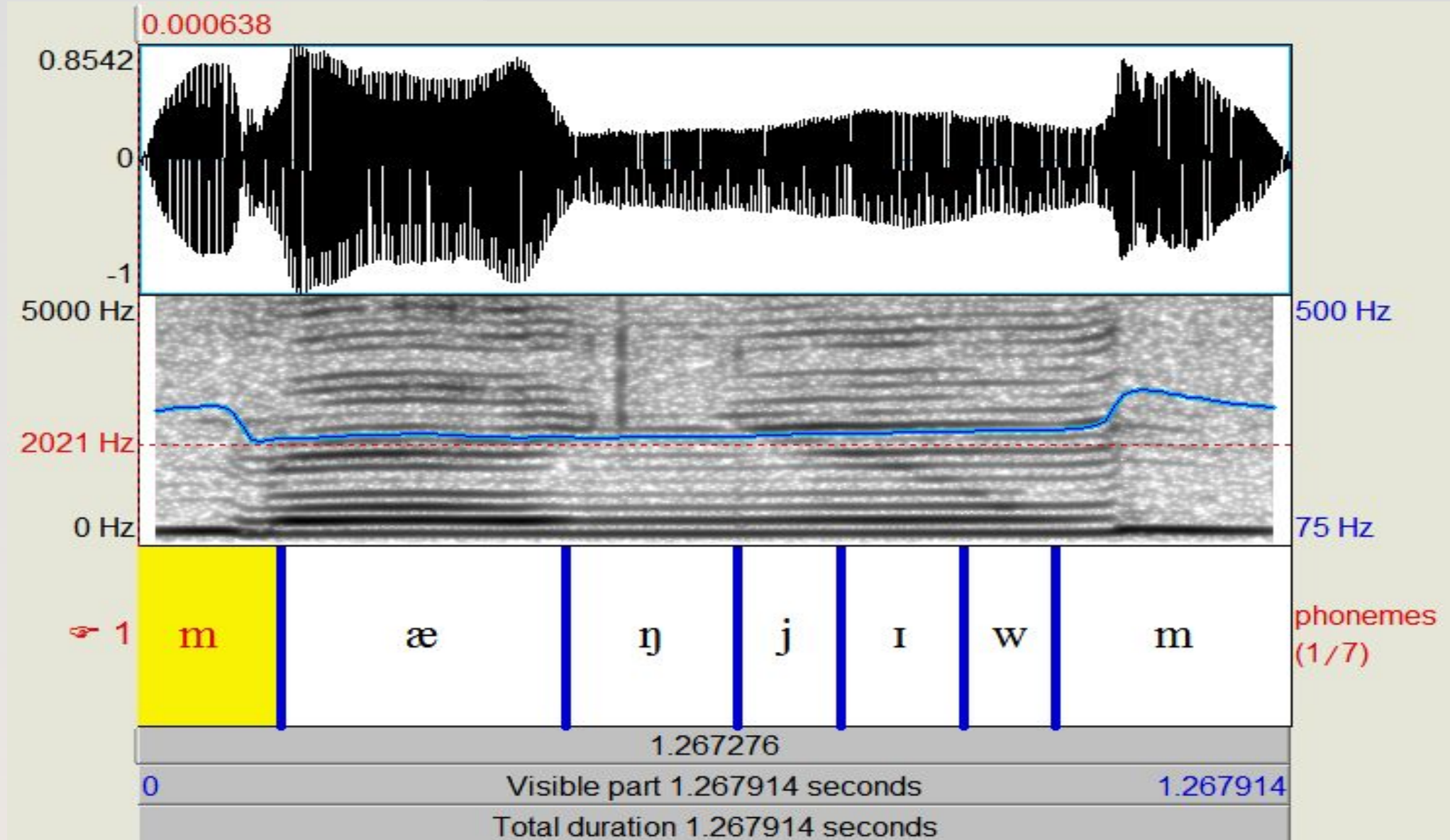
Hyper-nasalized, “d” realized as a click

# “Telephone” (DiGeorge syndrome)



“t” realized as a nasal-fricative

# "vacuum" (Cerebral Palsy) with our extIPA annotation





# "Plane" (Cerebral Palsy) with our extIPA annotation

