Speech-envelope enhancement to improve cocktail-party listening

Daniel Cardosi, Lucas Baltzell, Virginia Best

1. Background

How do we recognize what one person is saying when others are speaking at the same time . . . ?

Cherry, C. (1953). “Some experiments on the recognition of speech, with one and with two ears”

A few of the factors which give mental facility might be the following:

a) **The voices come from different directions.**
b) Lip-reading, gestures, and the like.
c) Different speaking voices, mean pitches, mean speeds, male and female, and so forth.
d) Accents differing
e) Transition-probabilities (subject matter, voice dynamics, syntax)

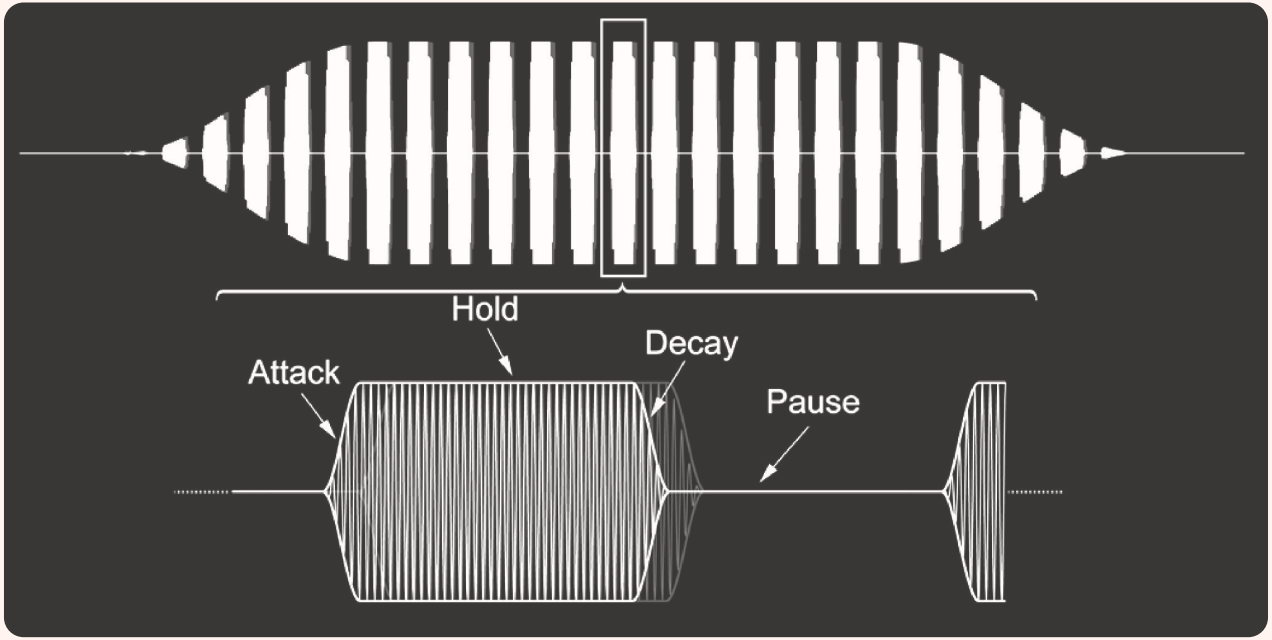
Colin Cherry (1953), “Some experiments on the recognition of speech, with one and with two ears”

Speech "Enhancement"

- Aim: to boost **speech intelligibility**
- Method: temporal-envelope manipulation
- Can also affect **binaural perception** through the alteration of **onsets**

** See our poster 4pPPb3 this afternoon! **

- Could help to solve the cocktail-party problem for the hearing-impaired



Klein-Hennig, M., Dietz, M., Hohmann, V., and Ewert, S. D. (2011). “The influence of different segments of the ongoing envelope on sensitivity to interaural time delays,” J. Acoust. Soc. Am. 129, 3856-3872.

Speech "Enhancement"

- Aim: to boost **speech intelligibility**
- Method: temporal-envelope manipulation
- Can also affect **binaural perception** through the alteration of **onsets**

** See our poster 4pPPb3 this afternoon! **

- Could help to solve the cocktail-party problem for the hearing-impaired

Expansion

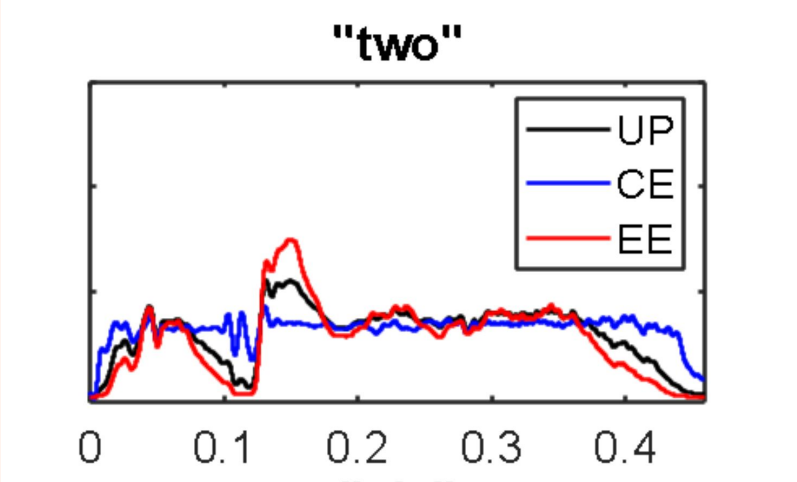
Increases amplitude variability

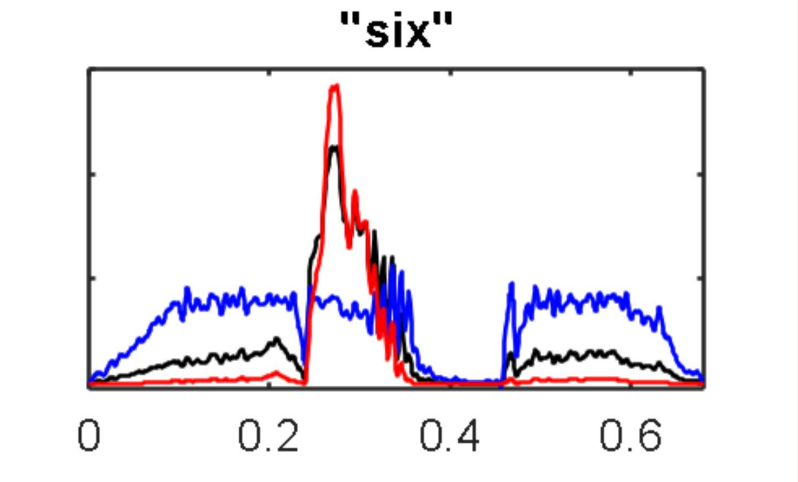
(Clarkson & Bahgat 1991; Fu & Shannon 1998; Lorenzi *et al*. 1999; *Apoux et al*., 2001, 2004).

Compression

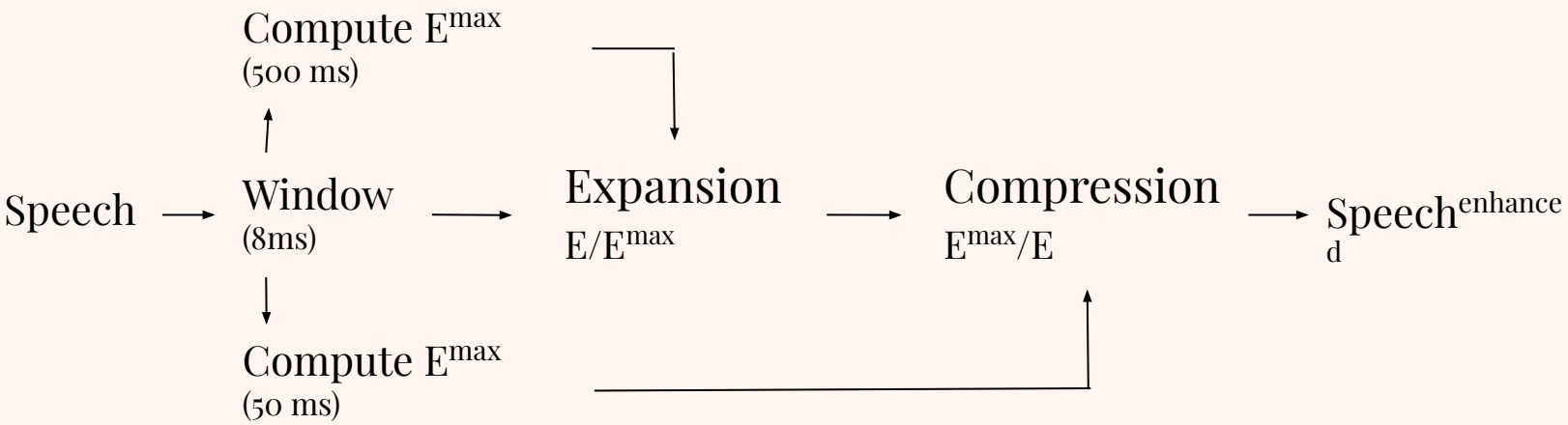
Decreases amplitude variability

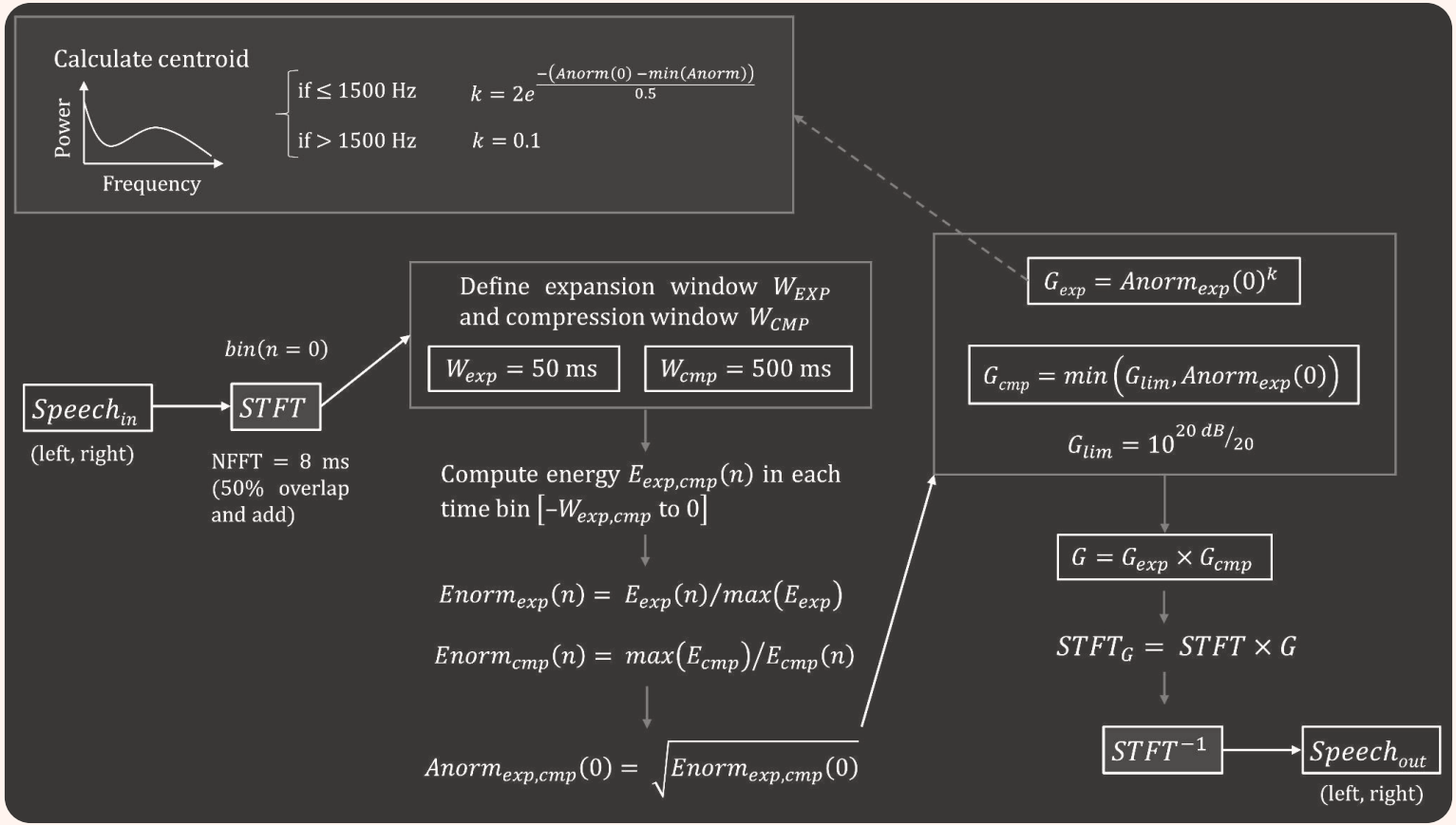
(Vandali 2001; Desloge *et al*. 2017; Goldsworthy *et al*. in press).


"two"
UP
CE
EE
0
0.1
0.2
0.3
0.4

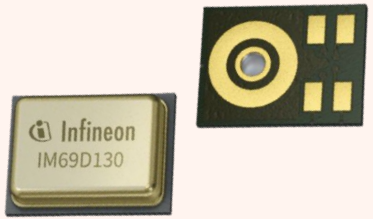

"six"
0
0.2
0.4
0.6

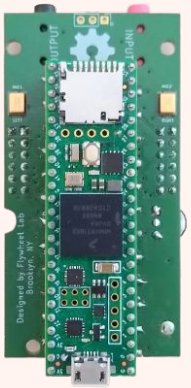
2. Algorithm


Compute E^{max}
(500 ms)
Speech
Window
(8ms)
Expansion
E/E^{max}
Compression
E^{max}/E
Speech^{enhanced}
Compute E^{max}
(50 ms)

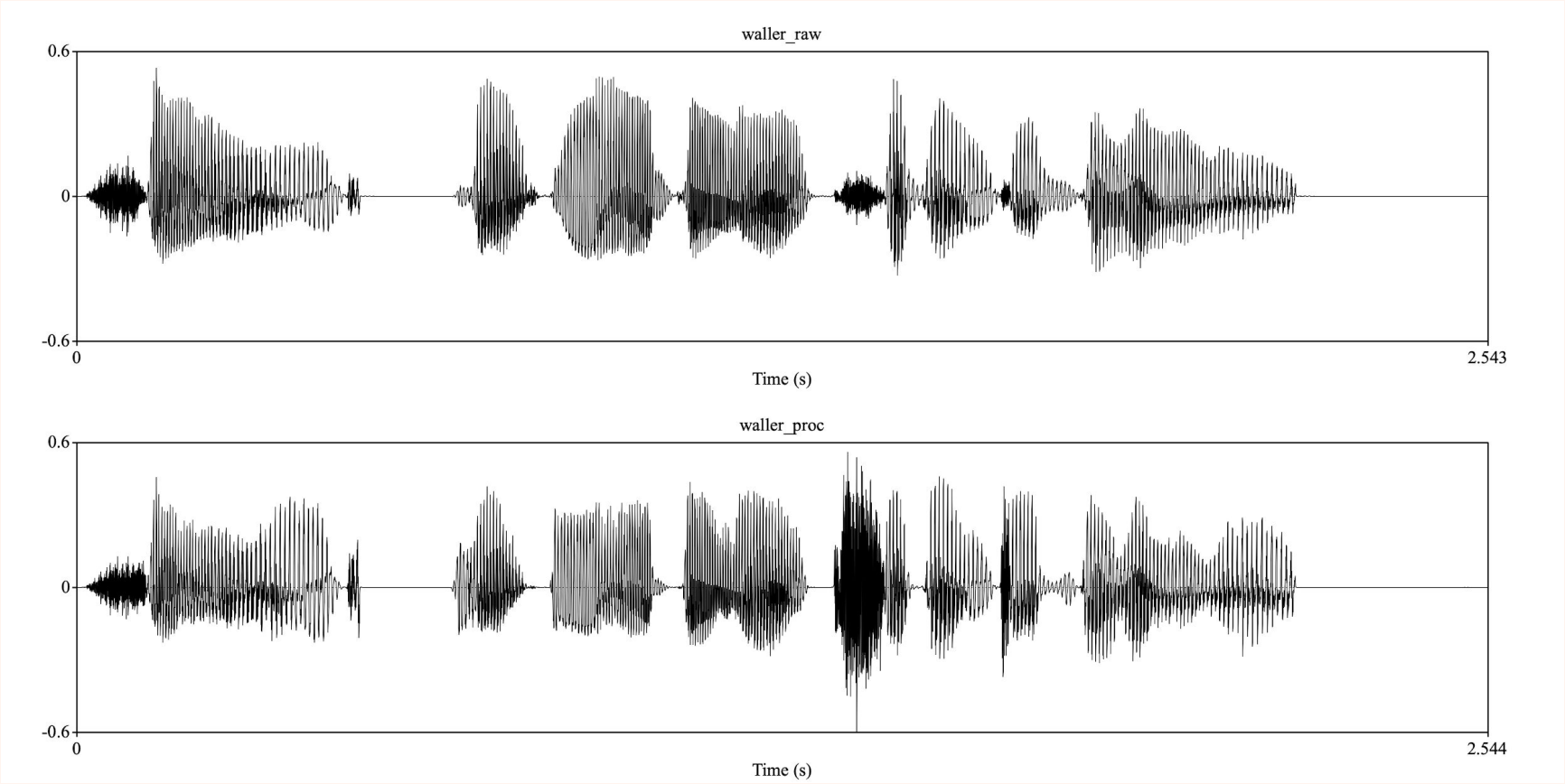

Calculate centroid
Power
Frequency
if ≤ 1500 Hz
$k = 2e^{\frac{-(Anorm(0) - min(Anorm))}{0.5}}$
if > 1500 Hz
$k = 0.1$
$bin(n = 0)$
$Speech_{in}$
(left, right)
$STFT$
NFFT = 8 ms
(50% overlap
and add)
Define expansion window W_{EXP}
and compression window W_{CMP}
$W_{exp} = 50$ ms
$W_{cmp} = 500$ ms
Compute energy $E_{exp,cmp}(n)$ in each
time bin $[-W_{exp,cmp}$ to $0]$
$Enorm_{exp}(n) = E_{exp}(n)/max(E_{exp})$
$Enorm_{cmp}(n) = max(E_{cmp})/E_{cmp}(n)$
$Anorm_{exp,cmp}(0) = \sqrt{Enorm_{exp,cmp}(0)}$
$G_{exp} = Anorm_{exp}(0)^k$
$G_{cmp} = min\left(G_{lim}, Anorm_{exp}(0)\right)$
$G_{lim} = 10^{20\ dB/20}$
$G = G_{exp} \times G_{cmp}$
$STFT_G = STFT \times G$
$STFT^{-1}$
$Speech_{out}$
(left, right)

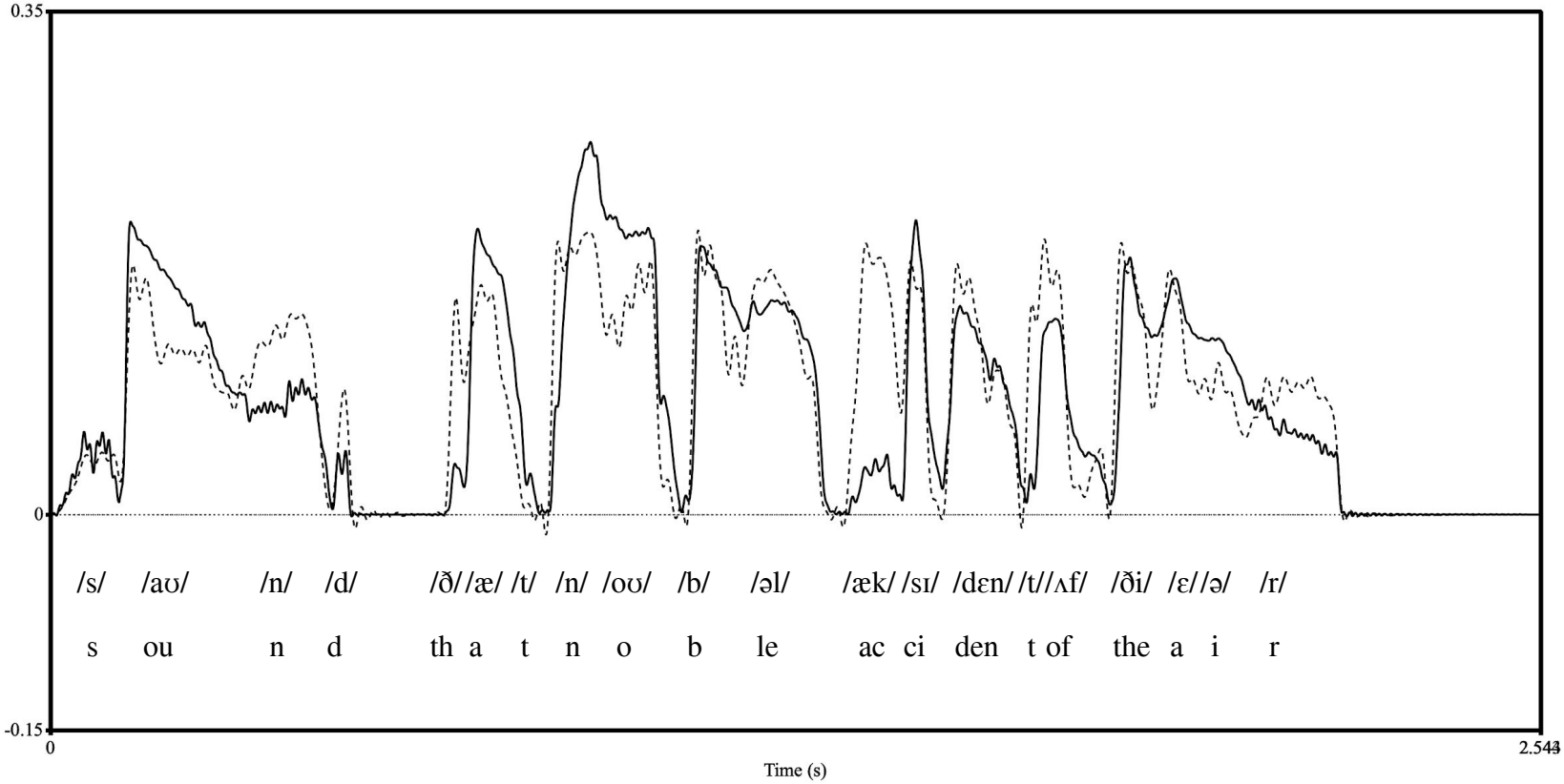
3. Project

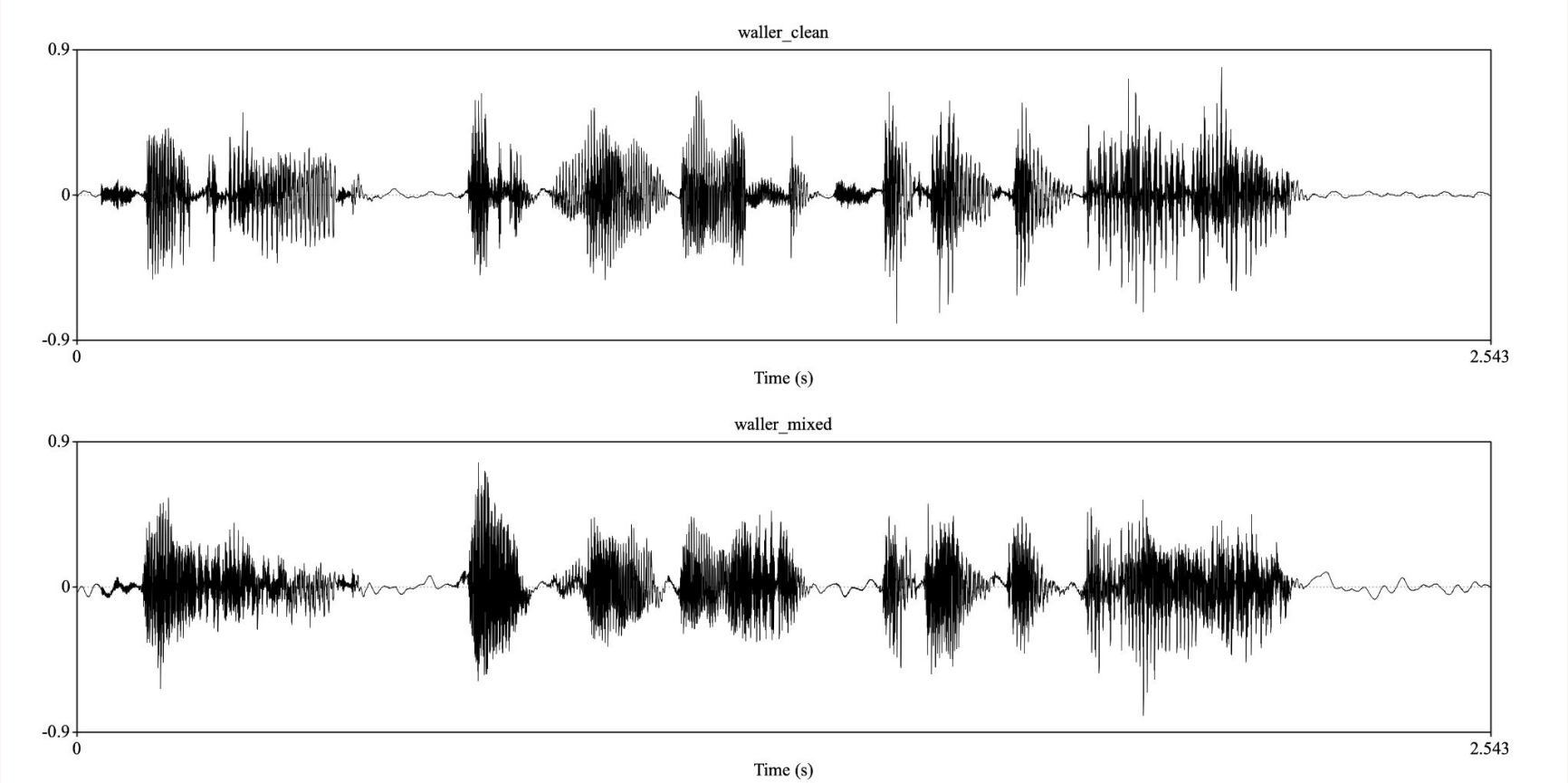


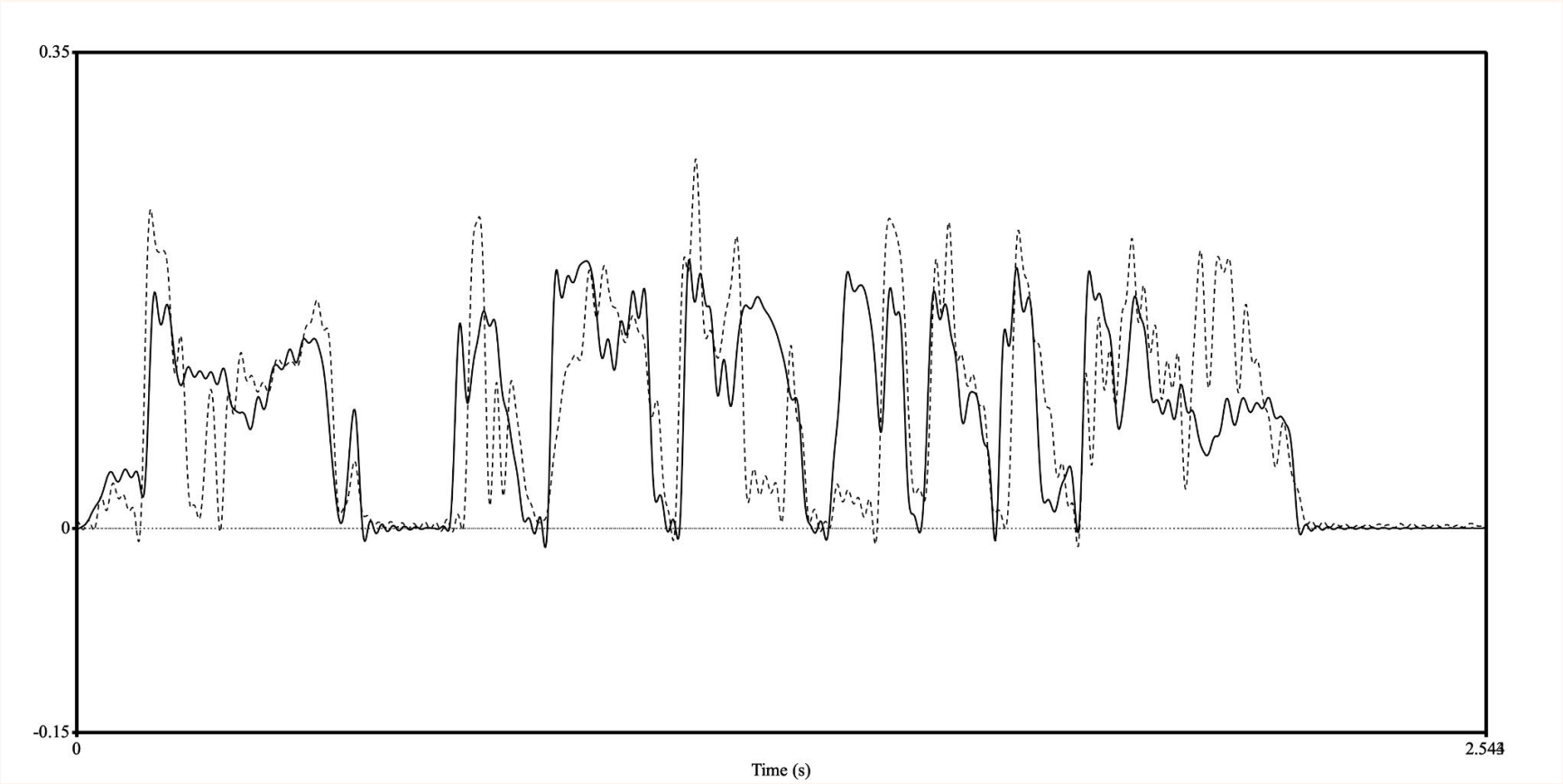


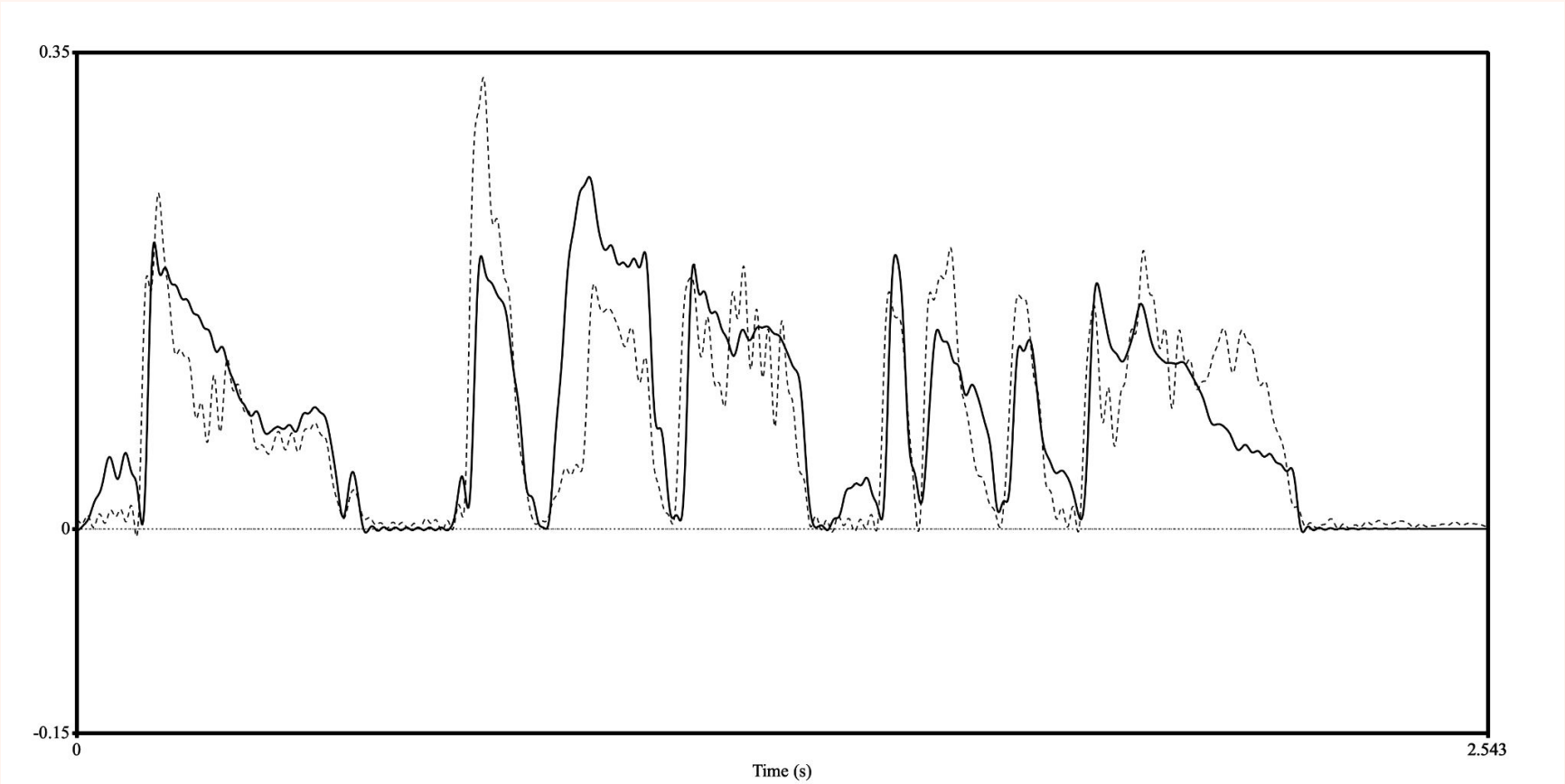
Infineon IM69D130 → *Teensy 4.1* → *Tympan BTE Earpieces*


waller_raw
0.6
0
-0.6
0
2.543
Time (s)
waller_proc
0.6
0
-0.6
0
2.544
Time (s)


0.35
0
-0.15
0
2.543
Time (s)
/s/ /aʊ/ /n/ /d/ /ð//æ/ /t/ /n/ /oʊ/ /b/ /əl/ /æk/ /sɪ/ /dɛn/ /t//ʌf/ /ðɪ/ /ɛ//ə/ /r/
s ou n d th a t n o b le ac ci den t of the a i r


waller_clean
0.9
0
-0.9
0
2.543
Time (s)
waller_mixed
0.9
0
-0.9
0
2.543
Time (s)


0.35
0
-0.15
0
2.543
Time (s)


0.35
0
-0.15
0
2.543
Time (s)

Applications

- Laboratory experiments: different methods of envelope enhancement
- Field trials
- Improve performance of hearing devices

Thank You!