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
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


Compression

Decreases amplitude variability

(Vandali 2001; Desloge et al. 2017; Goldsworthy et al. in press).
2. Algorithm
Speech \rightarrow \text{Window (8ms)} \rightarrow \text{Expansion} \rightarrow \text{Compression} \rightarrow \text{Speech}^{\text{enhance}}

\text{Speech} \\
\text{Compute } E^{\text{max}} (500 \text{ ms}) \\
\uparrow \quad \uparrow \quad \downarrow \quad \downarrow \\
\text{Window (8ms)} \quad \text{Expansion} \quad \text{Compression} \quad \text{Speech}^{\text{enhance}} \\
\downarrow \quad \downarrow \quad \downarrow \\
\text{Compute } E^{\text{max}} (50 \text{ ms})
Calculate centroid

\[
\begin{align*}
\text{Power} & \quad \text{Frequency} \\
\begin{cases}
\text{if } \leq 1500 \text{ Hz} & k = 2e^{-\frac{(\text{Anorm}(0) - \text{min(Anorm))}}{0.5}} \\
\text{if } > 1500 \text{ Hz} & k = 0.1
\end{cases}
\end{align*}
\]

Define expansion window \( W_{\text{EXP}} \) and compression window \( W_{\text{CMP}} \)

\[
\begin{align*}
W_{\text{exp}} &= 50 \text{ ms} \\
W_{\text{cmp}} &= 500 \text{ ms}
\end{align*}
\]

Compute energy \( E_{\text{exp,cmp}}(n) \) in each time bin \([-W_{\text{exp,cmp}} \text{ to } 0]\)

\[
\begin{align*}
\text{Enorm}_{\text{exp}}(n) &= E_{\text{exp}}(n)/\max(E_{\text{exp}}) \\
\text{Enorm}_{\text{cmp}}(n) &= \max(E_{\text{cmp}})/E_{\text{cmp}}(n)
\end{align*}
\]

\[
\text{Anorm}_{\text{exp,cmp}}(0) = \sqrt{\text{Enorm}_{\text{exp,cmp}}(0)}
\]

\[
\begin{align*}
G_{\text{exp}} &= \text{Anorm}_{\text{exp}}(0)^k \\
G_{\text{cmp}} &= \min\left(G_{\text{lim}}, \text{Anorm}_{\text{exp}}(0)\right) \\
G_{\text{lim}} &= 10^{\frac{20 \text{ dB}}{20}} \\
G &= G_{\text{exp}} \times G_{\text{cmp}} \\
\text{STFT}_G &= \text{STFT} \times G
\end{align*}
\]

\[
\begin{align*}
\text{Speech}_{\text{out}} & \quad \text{(left, right)}
\end{align*}
\]
3. Project
Infineon IM69D130 → Teensy 4.1 → Tympan BTE Earpieces
sound that noble accident of the air
Applications

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