1. Summary

- Stop consonants are an important component of the speech signal
- Their dynamic and unpredictable nature makes them difficult to encode
- A system to detect, segment, and modify stop consonants in a speech signal is proposed
- Results from subjective evaluation open new perspectives for speech coding

2. Stop consonants

- Are produced by forming a closure in the vocal tract, building up pressure in the mouth behind this closure, and releasing the closure
- Are typically decomposed into 3 or 4 main phases:
  - Closure
  - Burst
  - Aspiration
  - Coarticulation

3. Overview of the system

4. Detection and Segmentation

- Detection of the Closure-Burst Transition (CBT) using the Plosion Index [Ananthapadmanabha 2014]
  \[ PI(n_0,m_1,m_2) = |s(n_0)| / \text{savg}(m_1,m_2) \]
- Uses a “constrained” PI to eliminate false positives
- Excellent detection performance and temporal resolution
- However: tendency towards late detection (“Original CBT instant”)
- An improvement is presented that uses the raw plosion index and instants of glottal closure (GCI) to avoid late detections (“Modified CBT instant”)

5. The alteration: muting the burst

6. Subjective evaluation (TIMIT)

- Comparison Category Rating (CCR) listening test
- Test (altered vs orig.) and control (orig. vs orig.) sequences
- 12 non-native English expert listeners
- Comparison Mean Opinion Score (CMOS) for Test: -0.12
- T-test marginally fails at the 5% significant level: [-0.15; -0.08]

7. Conclusions and Perspectives

- Stop consonants can be drastically altered at a very limited cost in terms of subjective quality
- On the TIMIT corpus, up to 1.88% of the total coding rate might be saved by simply muting the segmented burst phases (assuming fixed rate and ideal adaptive framing)