**1 Introduction**

**Problem:** Estimate phase \( \phi \) from given magnitude spectrum \( M \) such that a consistent time signal is achieved via inverse short-time Fourier transform (ISTFT).

**Applications:**
- Speech enhancement and speech separation
- Speech synthesis and voice conversion

---

**2 System Overview**

Two-stage phase reconstruction system (similar to [1]):
1. Use deep neural networks (DNNs) to estimate phase derivatives
2. Reconstruct phase from its estimated derivatives

**Proposed improvements:**
- A novel regularized cosine loss function
- Shift correction (SC) as a pre-processing step
- A novel phase reconstruction method

---

**3 Phase Derivatives Estimation**

- Train two equally structured DNNs using combined loss:
  \[ L_{\text{total}} = L[\Delta\hat{\psi}] + \lambda \cdot L[\Delta\hat{\psi}] \]
  \( L \) should consider \( 2\pi \) ambiguity and have a limited solution space
- **Novelty I** - regularized cosine loss function:
  \[ L_{\text{reg}}[\Delta\hat{\psi}] := \sum_{k,m} - \cos \{ \Delta\hat{\psi}(k, m) \} + \lambda \cdot |\Delta\hat{\psi}(k, m)|^4 \]
  Here: \( \lambda = \frac{\pi}{2} \)
- Systematic offsets occur in the calculation of \( \hat{\psi} \) and \( \hat{\psi}_d \)
- Offset in \( \hat{\psi}_d \) can be described by the shift theorem of the DFT:
  \[ x[n-S] \leftrightarrow X(k) \cdot e^{-i\omega S} \]
- Systematic shift in \( \hat{\psi}_d \) can be observed empirically
- **Novelty II** - shift correction:
  \[ \psi_i(k, m) = W[\psi_i(k, m) - \pi/2] \]
  \[ \psi_d(k, m) = W[\psi_d(k, m) + \pi] \]

**4 Phase Reconstruction Method**

- Combine \( \hat{\psi}_i \) and \( \hat{\psi}_d \) such that a consistent \( \hat{\phi} \) is achieved
- **Novelty III** - averaging of weighted estimates \( \hat{\varphi}_P \) from \( P \) paths:
  \[ \hat{\phi}(k, m) = \frac{1}{P} \sum_{p=1}^{P} \alpha_p(k, m) \cdot e^{i\hat{\phi}_p(k, m)} \]
  with estimation quality indicators \( \alpha_p \):
  \[ \alpha_1(k, m) = M(k - 1, m) \]
  \[ \alpha_2(k, m) = M(k, m - 1) \]
  \[ \alpha_3(k, m) = \min_{l \in \{-1, 0, 1\}} M(k + l, m + l) \]
- Polar histograms of path error \( \hat{\varphi}_P(k, m) - \hat{\phi}(k, m) \) demonstrate suitability of chosen weights

---

**5 Evaluation**

Results after phase reconstruction using different methods:
- Two-stage reconstruction method performs reference algorithms
- Similar performance to Griffin-Lim (100 it.) although no iterations are required

**6 Conclusion**

- Proposed novelties significantly improve phase reconstruction system
- Novelty I - regularized cosine loss function stabilizes training
- Novelty II - shift correction further stabilizes and accelerates training
- Novelty III - phase reconstruction method outperforms reference algorithms

**References**

---

Lars Thieling, Daniel Wilhelm, Peter Jax

International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2021, Toronto, Canada