The effect of retiming speech on masked intelligibility

Martin Cooke, Ikerbasque & Language and Speech Lab, Universidad del País Vasco, Spain
Vincent Aubanel, The MARCS Institute, University of Western Sydney, Australia

Introduction

Many ‘near end’ speech modification techniques (e.g., [1]) operate in the spectral domain, achieving gains of more than 5 dB in international evaluations [2]. Temporal adjustments are also possible and have shown substantial benefits for fluctuating maskers [3, 4]. The current study asks whether spectral and temporal modifications are synergistic and investigates the basis for gains produced by speech retiming.

Previous work

The GCRetime algorithm [3]

• Fine-scale expansion [9] of speech
• Optimisation of Glimpse proportion [5] (to improve masked audibility) and Cochlear scaled entropy [6] (to weight important speech information)
• Largest gains (≈ 4 dB) for competing speech masker in Hurricane Challenge [4]

SSDRC [1]

• Spectral Shaping and Dynamic Range Compression
• Largest gains (≈ 5 dB) for speech-shaped noise masker in Hurricane Challenge [4]

Issues

1. Are spectral and temporal modifications synergistic?
2. Do temporal modifications result in speech which is intrinsically more intelligible when taken out of the inducing-masker context?
3. To what extent is any benefit of retiming due to mere elongation?

The new study also tests earlier findings using speech material in a different language (Spanish) [7].

Cost functions

\[ c(t, j) = G(t, j) = G(t, j)W_{CSE}(t) \]

where

\[ W_{CSE} = (w - 1) \frac{CSE - \mu}{\Delta} + 1 \]

Here, expansion-only path constraints

Retiming paths for GP only (dotted) and GP+CSE (solid). Red line is CSE index, with high-CSE regions shaded.

Experiment 1: spectro-temporal synergy?

Native Spanish listeners identified keywords in sentences masked by a competing speaker in 2 SNRs and 5 conditions:

- plain
- unmodified G
- temporal GC
- spectral S
- SSDRC alone SSDRC alone
- SGC SGC
- SSDRC followed by GC

Experiment 2: role of mere elongation?

Conditions:

- plain
- retimed using GC condition of exp 1
- elongated: linear stretch to duration of retimed sentence via WSOLA

Masked by:

- competing speech (CS)
- speech-modulated noise (SMN)
- speech-shaped noise (SSN)

Discussion

• Spectral and temporal intelligibility-enhancing modifications can be complementary, at least for fluctuating maskers
• Elongation alone is beneficial for fluctuating maskers (cf. lack of benefits in earlier studies; e.g. [8, 9, 10, 11]), - increased likelihood of speech information appearing in masker dips?
• Why is CSE not beneficial?
  - At these SNRs perhaps audibility takes precedence
  - Transitional information promoted by CSE may be easily masked
• Further gains for retiming may be realisable if speech-rate changes – presumably responsible for reduced intelligibility in SSN – are mitigated. Informally, retiming is evident only in SSN.
• Low-delay implementations with compression and expansion show promise but more work required to retain important speech information.

References


Acknowledgments: We thank Yannis Stylianou for providing code implementing the SSDRC algorithms. This work was supported in part by the EU Project LUSC (The Listening Tailor).