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 [& compression] of speech
- Optimisation of **G**limpse proportion [5] (to improve masked audibility) and **C**ochlear scaled entropy [6] (to weight important speech information)
- Largest gains (≈ 4 dB) for competing speech masker in Hurricane Challenge [4]

SSDRC [1]

- **S**pectral **S**haping and **D**ynamic **R**ange **C**ompression
- 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].

Retiming

Glimpse proportion

$$GP = \frac{1}{TF}\sum_{f=1}^{F}\sum_{t=1}^{T}\mathcal{H}[\,S(t,f) - M(t,f) - \alpha\,]$$

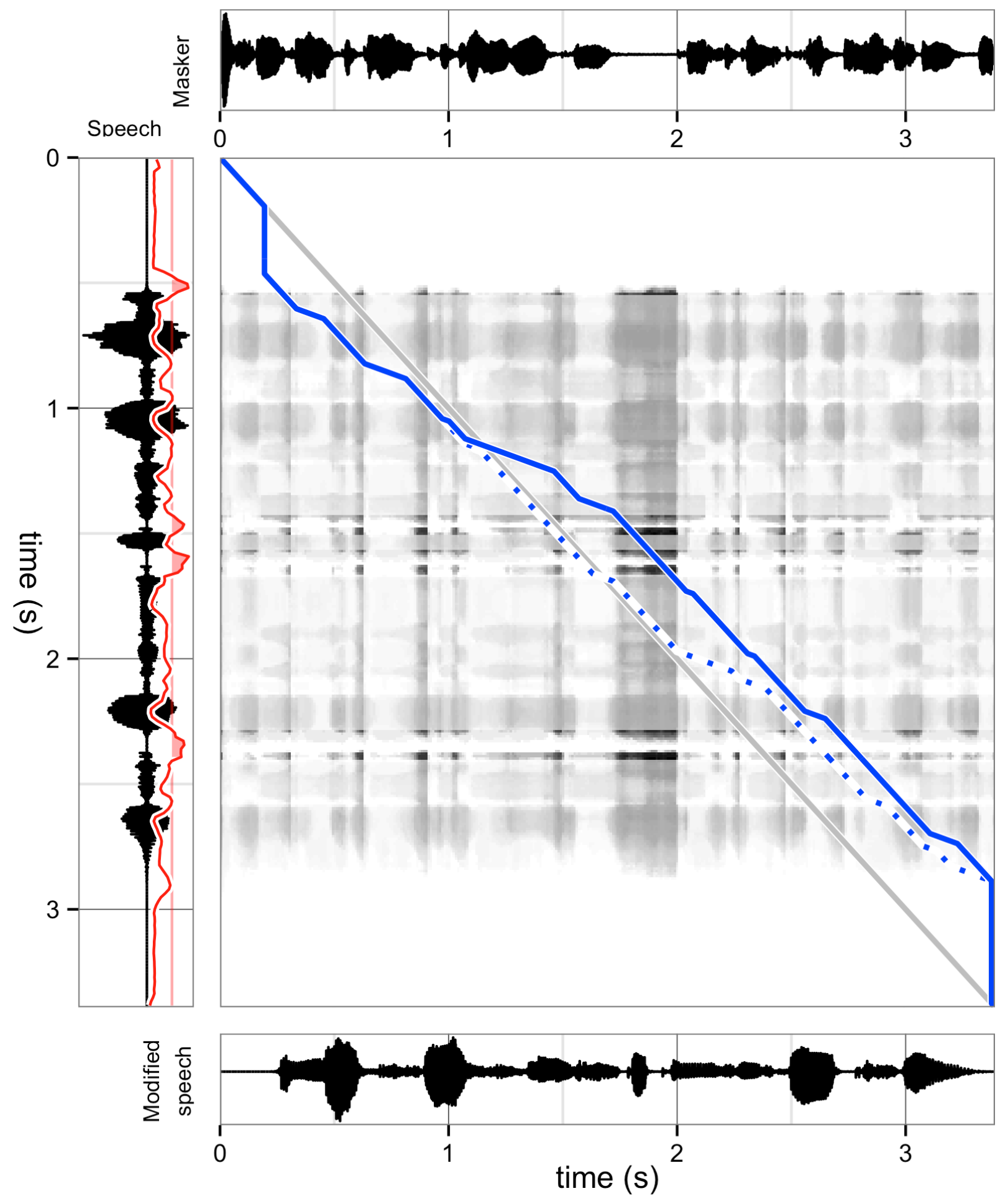
where S and M are auditory spectro-temporal excitation patterns in dB for speech and masker; $\alpha = 0$ dB

Cochlear-scaled entropy

$$CSE(t) = \sum_{k=-b/2}^{b/2} d(t+k), \text{with } b = 5 \text{ frames } (80\,ms)$$

where

$$d^2(t) = \sum_{f=1}^{F}[S(t+1,f) - S(t,f)]^2$$



Cost functions

$$\begin{aligned} c(i,j) &= GP_f(i,j) \\ &= GP_f(i,j)\,W_{\mathsf{CSE}}(i) \end{aligned}$$

where

$$W_{\mathsf{CSE}} = (w-1)\,\mathcal{H}[CSE - \beta] + 1$$

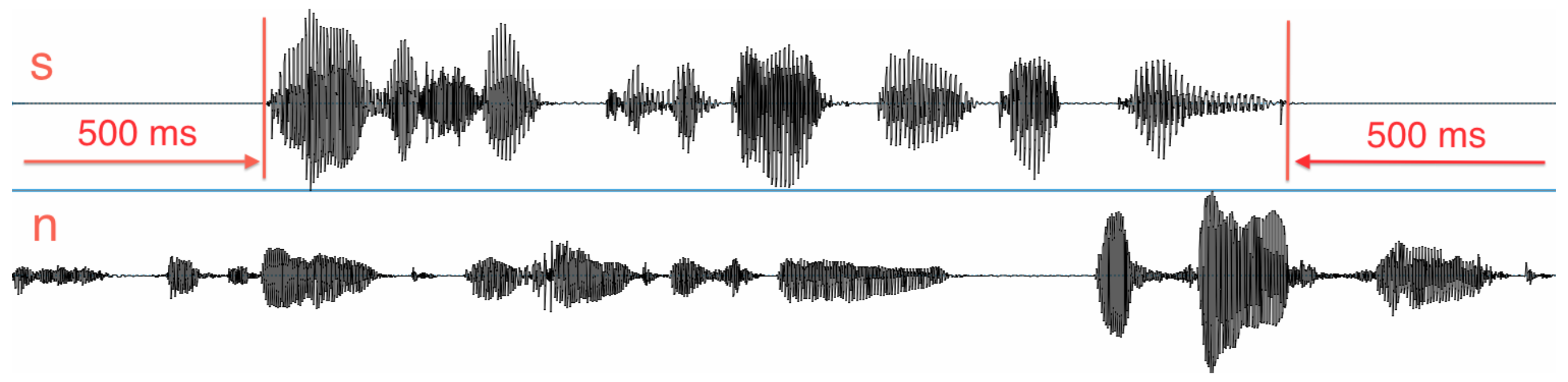
Here, expansion-only path constraints

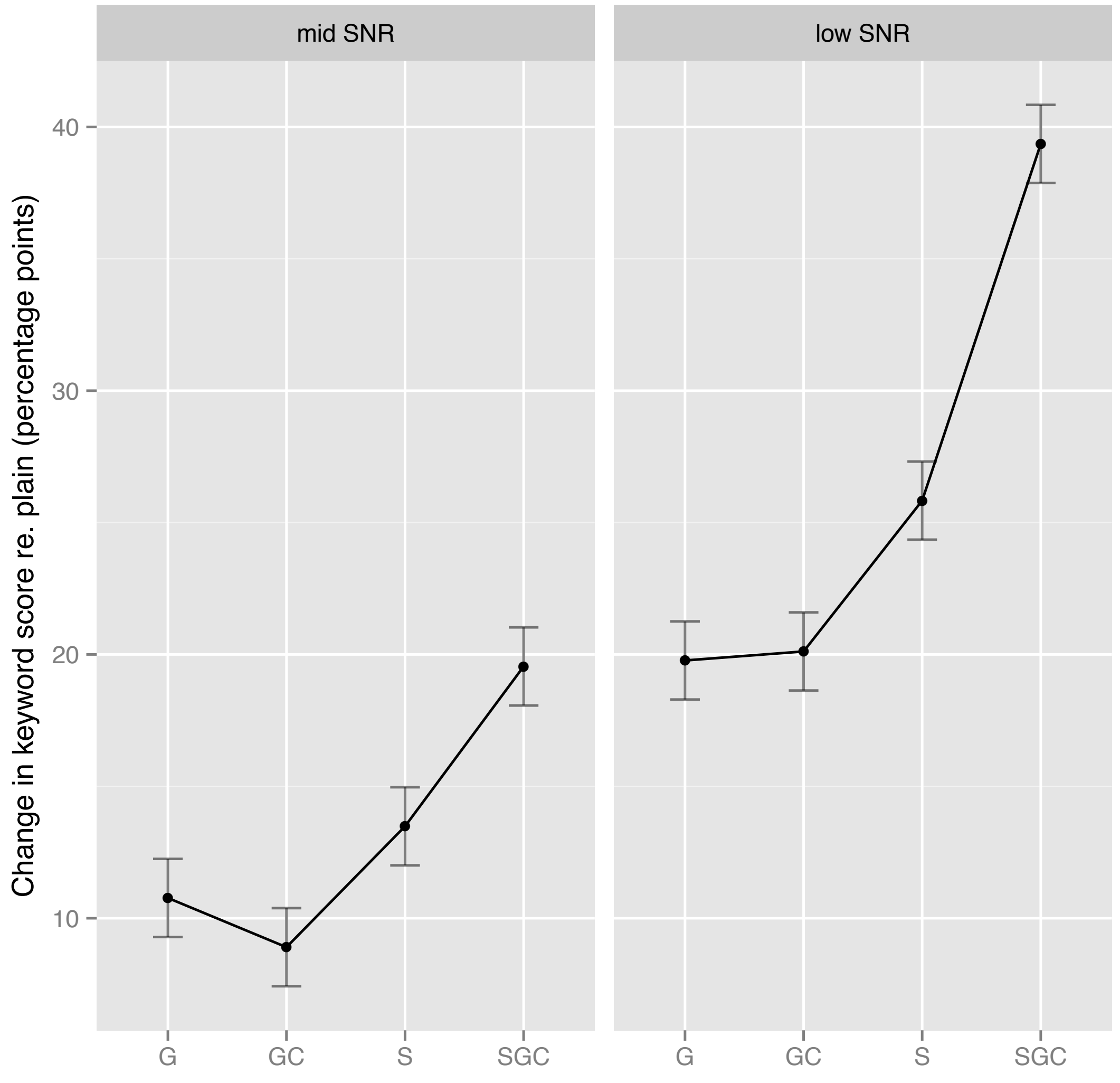
Retiming paths for GP only (dotted) and GP+CSE (solid). Red line is CSE index, with high-CSE regions shaded.

Experiment 1: spectro-temporal synergy?

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

plain	unmodified	baseline
G	temporal	retiming based on GP
GC	temporal	retiming based on GP and CSE
S	spectral	SSDRC alone
SGC	spectral+temporal	SSDRC followed by GC





- Significant gains over baseline in all conditions
- No benefit of CSE
- Complementary but sub-additive effect of spectral and temporal modifications

$N = 22$. Error bars indicate ± 0.5 Fisher LSD

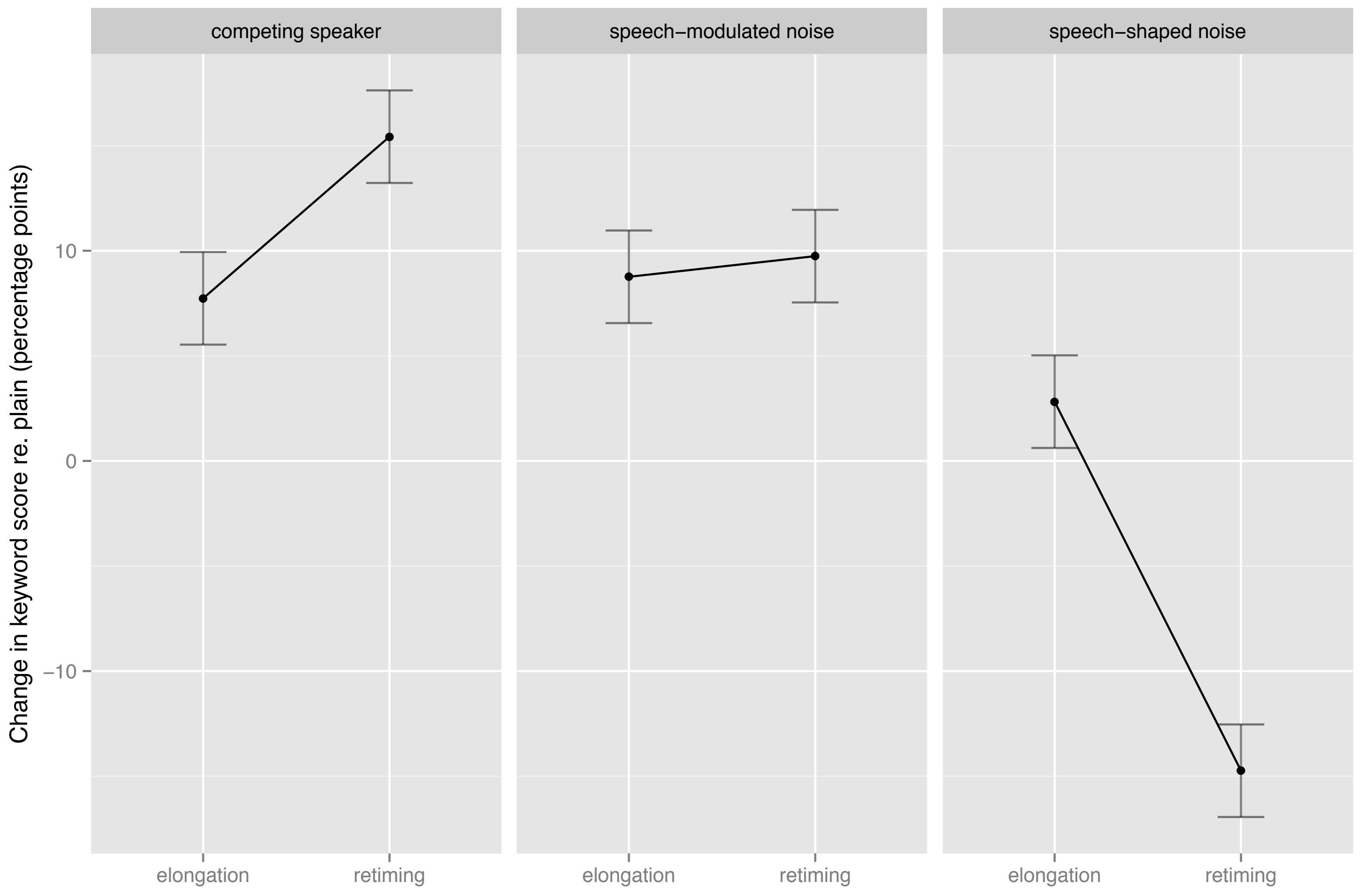
Experiment 2: role of mere elongation?

Conditions:

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

Masked by

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



- For CS: about half benefit of retiming comes from elongation
- But for SMN no additional retiming benefit
- In SSN, retiming is harmful ⇒ reduced intrinsic intelligibility

First finding of masker-dependent benefits of elongated speech?

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

[1] Zorila, T.-C., Kandia, V., & Stylianou, Y. (2012). Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression. *Proc. Interspeech*, Portland, USA, 635-638.

[2] Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Comm.*, 55, 572–585.

[3] Aubanel, V., & Cooke, M. (2013). Information-preserving temporal reallocation of speech in the presence of fluctuating maskers. *Proc. Interspeech*, Lyon, France, 3592–3596.

[4] Cooke, M., Mayo, C., & Valentini-Botinhao, C. (2013). Intelligibility-enhancing speech modifications: the Hurricane Challenge. *Proc. Interspeech*, Lyon, France, 3552–3556.

[5] Cooke, M. (2006). A glimpsing model of speech perception in noise. *J. Acoust. Soc. Am.*, 119:1562–1573.

[6] Stilp, C. & Kluender, K. (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *PNAS*, 107:12387–12392.

[7] Aubanel, V., Garcia Lecumberri, M. L., & Cooke, M. (in revision). The Sharvard corpus: A phonemically-balanced Spanish sentence resource for audiology. *Int. J. Audiology*.

[8] Schmitt, J. F. (1983). The effects of time compression and time expansion on passage comprehension by elderly listeners, *J. Speech Lang. Hear. R.*, 26:373.

[9] Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. J. Speech Hear. Res., 39, 494–509.

[10] Nejime,Y., & Moore, B.C.J. (1998). Evaluation of the effect of speech-rate slowing on speech intelligibility in noise using a simulation of cochlear hearing loss. J. Acoust. Soc. Am., 103, 572–576.

[11] Cooke, M., Mayo, C., & Villegas, J. (2014). The role of durational increases in the Lombard speech intelligibility benefit. *J. Acoust. Soc. Am.*, in press.

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