EFFECTS OF SPECTRAL TILT ON LISTENERS' PREFERENCES AND INTELLIGIBILITY

Olympia Simantiraki[1], Martin Cooke[2], Yannis Pantazis[3]

[1] University of the Basque Country, Spain
[2] Ikerbasque (Basque Science Foundation), Spain
[3] Institute of Applied and Computational Mathematics, FORTH, Greece

ENRICH EUROPEAN TRAINING NETWORK

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

ICASSP 2020

BACKGROUND

Near-end listening enhancement (NELE) algorithms improve intelligibility in adverse conditions *(Schepker et al, 2015)*

Supra-intelligibility factors also important:

Effortful listening : increased frustration
fatigue
decreased concentration

impact on work performance (*Hétu et al, 1988; Al-Hanbali et al, 2016)*

Lack of naturalness : synthetic speech less acceptable

listeners prefer to listen to natural speech (*Chang, 2011*)

Schepker et al (2015) J. Acoust. Soc. Am. 138, 2692–2706
Hétu et al (1988) Br. J. Audiol. 22, 251–264
Al-Hanbali et al (2016) Ear Hear. 38, e39–e48
Chang (2011) ACLCLP, 64–78

BACKGROUND

Supra-intelligibility factors can be explored through **listener preferences**.

Listener preferences (LP) in different conditions have been measured:

- via subjective scales on a five-point scale (*too slow, slow but ok, preferred, fast but ok, too fast)*
  - speech rate (*Moore et al, 2007; Adams & Moore, 2009)*
- by allowing listeners to modify speech properties in real-time using virtual adjustment tools
  - speech rate (*Wingfield & Ducharme, 1999; Simantiraki & Cooke, 2019)*
  - local criterion for retaining or removing time-frequency regions (*Zhang & Shen, 2019)*

Moore et al (2007) Int. J. Audiol. 46, 154–160
Adams and Moore (2009) J. Am. Acad. Audiol. 20, 28–39
Wingfield and Ducharme (1999) J. Gerontol. B Psychol. Sci. Soc. Sci. 54B, P199–P202
Simantiraki and Cooke (2019) ICA, 5736-5738
Zhang and Shen (2019) Interspeech, 1383–1387

BACKGROUND

In noise talkers produce Lombard speech
(*Zollinger & Brumm, 2011)*

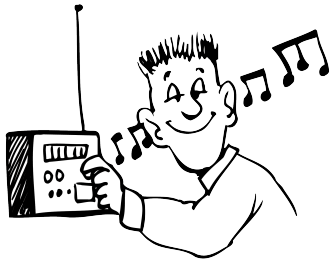
feature	**compared to neutral**	**intelligibility gains**
duration	elongated speech	only in babble noise (*Adams & Moore, 2009)*
F0	higher pitch	no (*Lu & Cooke, 2009)*
spectral tilt	flatter tilt	yes (*Lu & Cooke, 2009)*

Main research questions:

1. Do listener preferences show a pattern different from intelligibility?
2. Do spectral tilt preferences change in challenging conditions?
3. Can listeners' preferences be modelled?

Zollinger and Brumm (2011) Behaviour 148, 1173–1198
Adams and Moore (2009) J. Am. Acad. Audiol. 20, 28–39
Lu and Cooke (2009) Speech Commun. 51, 1253–1262

LISTENER PREFERENCES EXPERIMENT



SpeechTuner: A tool for acquiring listener preferences!

Tool uses pre-computed stimuli to ensure that changes can apply to **any** parameter and are smooth.

CODE RELEASE SOON…

LISTENER PREFERENCES EXPERIMENT

Participants: 35 normal-hearing, native Spanish listeners (30 female, mean age 20)

Stimuli:

- male talker, Spanish Harvard sentences corpus (5 keywords)
- presented through headphones

Maskers: quiet and SSN at 0, -3, -6 dB SNR

Instructions: tune the speech in real-time until you can recognise as many words as possible (23 available levels)

LISTENER PREFERENCES EXPERIMENT

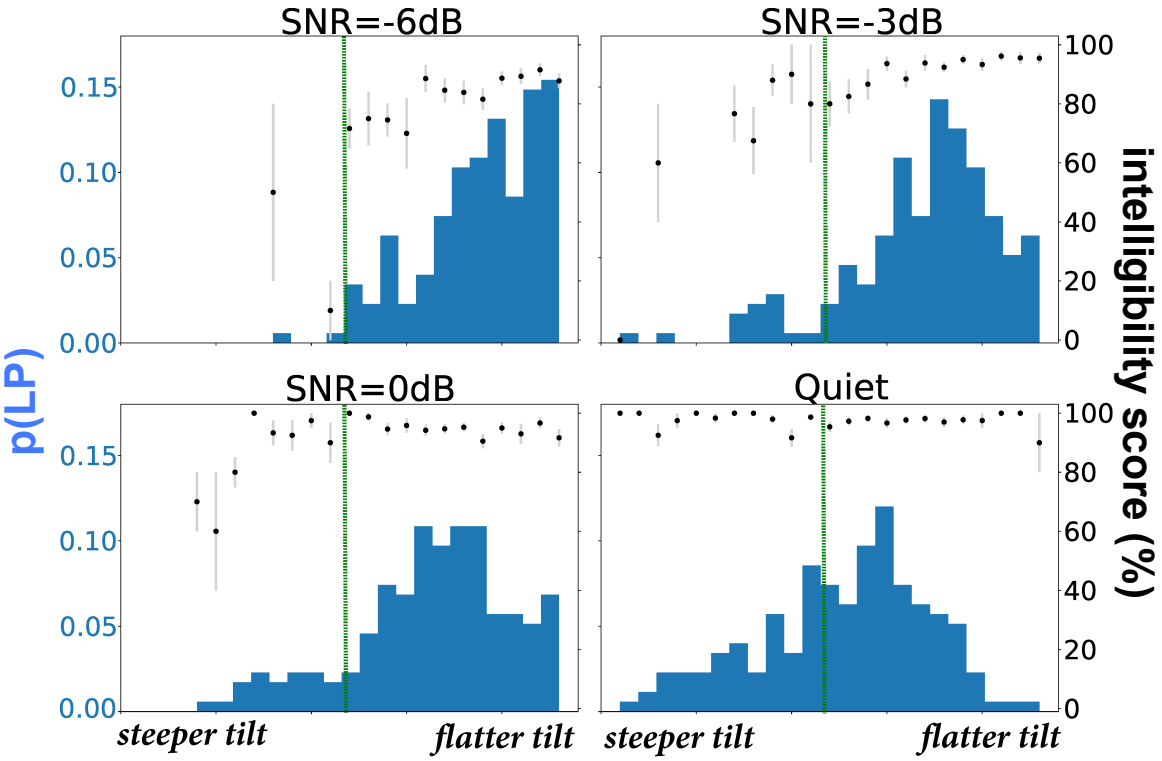
Procedure:

- 4 blocks divided by condition
  - 5 trials for modifying spectral tilt
- each trial:

  adjustment phase:
  - started at a random spectral tilt value
  - must listen to at least 5s of speech before proceeding to the test phase

  test phase:
  - type out what was heard (sequence of 2 sentences)
- prior to the experiment
  - familiarization phase of 5 trials
- block ordering across participants using balanced Latin square design

RESULTS



Effects of spectral tilt on LP and intelligibility

- with increasing SNR, more tilt values with intelligibility scores at ceiling and LPs occupy a wider range of tilts
- distinct preferences even when intelligibility is at or near ceiling
- poorer intelligibility score when steeper tilt than the original was chosen

Spectral tilt preferences across different noise levels

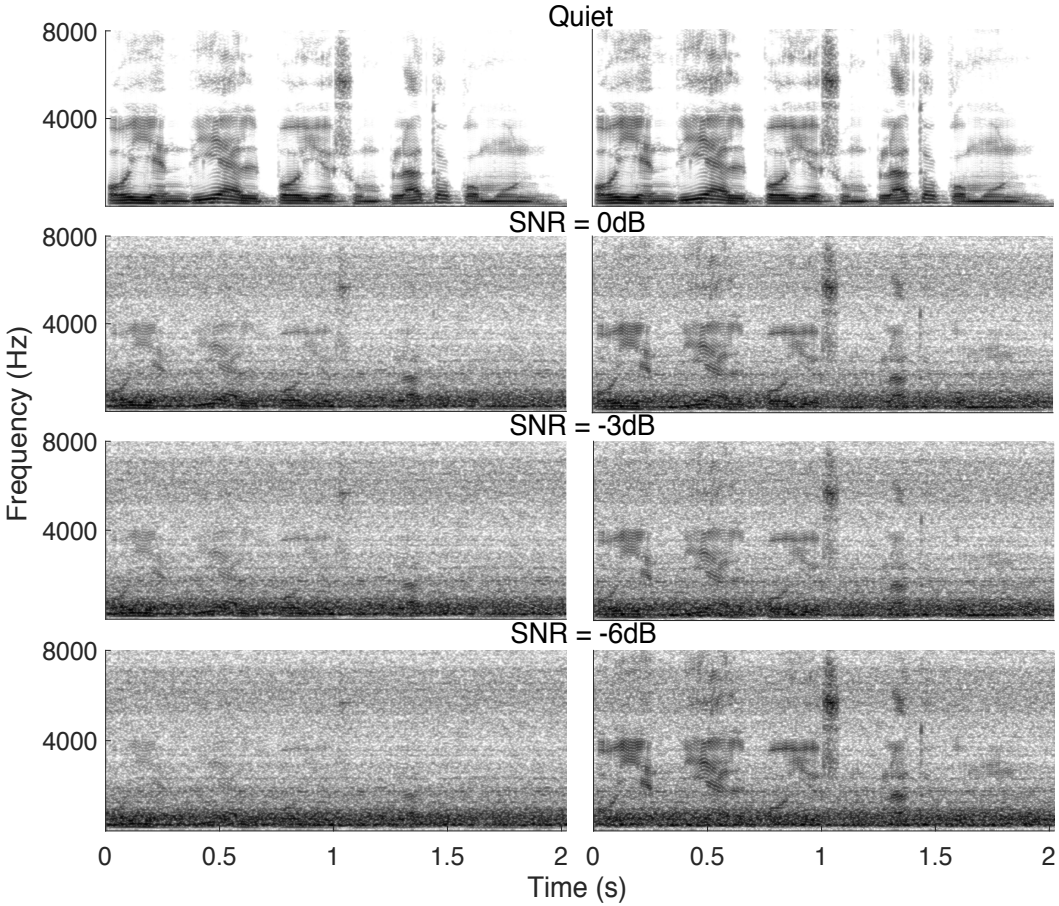
- distributions differ significantly (K-S test), apart from 0 and -3 dB SNR
- clear effect of SNR on LPs (one-way ANOVA)
- with increasing noise level, flatter tilt preferred (post-hoc comparisons)

MODELLING LISTENER PREFERENCES

INTELLIGIBILITY FACTOR

LPs are, in part, related to the intelligibility scores

- most probable preferences led to high intelligibility scores



First component in our model, the *extended glimpse proportion metric (GP_{ext}):* an objective measure of energetic masking and good predictor of intelligibility (*Tang & Cooke, 2016*)

Tang and Cooke (2016) Interspeech, 2488–2492

MODELLING LISTENER PREFERENCES

SUPRA-INTELLIGIBILITY FACTORS
Under less severe conditions, LPs are influenced by factors beyond intelligibility

Captured with a *Gaussian-type* component in our model

The proposed model defined by:

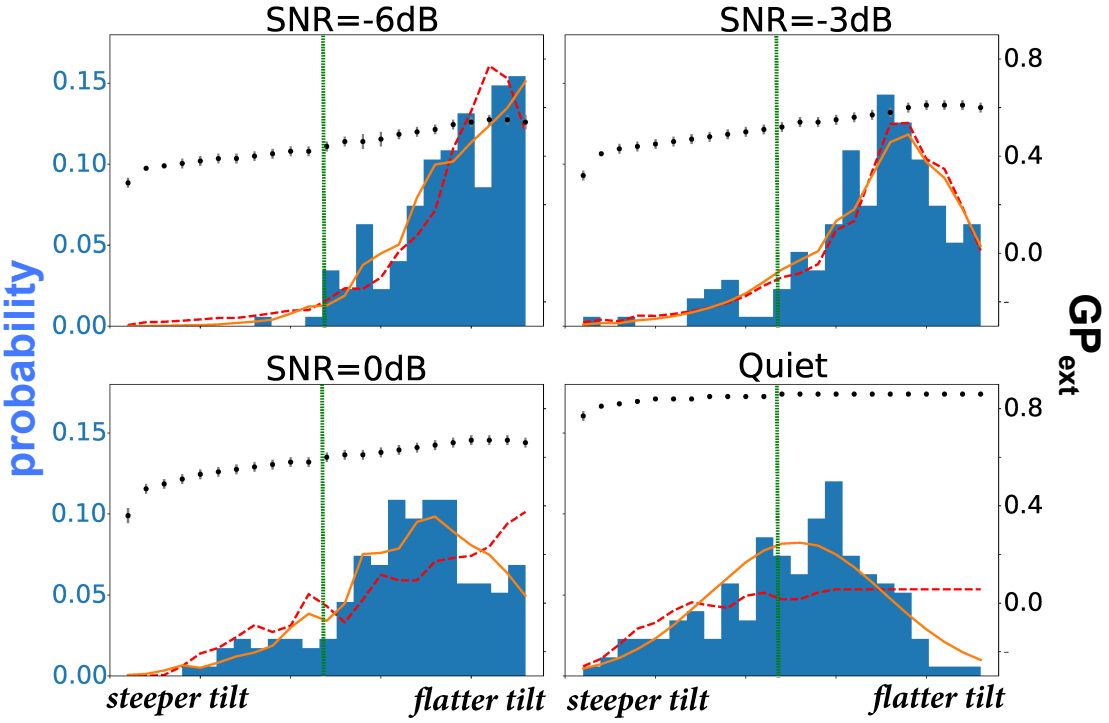
$$p_{LP}(x) \propto e^{\alpha_0 g(x-\alpha_1)-\beta_0(x-\beta_1)^2}$$

α_0 and β_0: scale parameters that widen or contract the PDF's shape

α_1 and β_1: translation parameters which move the PDF to the left or right

$g(x)$: glimpsing model for the spectral tilt levels, x in [1, 23]

MODELLING LISTENER PREFERENCES



Model parameters estimated from experimental data:

- mean GP computed using the test phrases
- KLD minimization of LPs used with respect to the proposed distribution
- optimal values estimated using gradient descent algorithm

Ablation test: reduced model (red-dashed line) $p_{LP}(x) \propto e^{\alpha_0 g(x-\alpha_1)-\beta_0(x-\beta_1)^2}$

- Quadratic term helped the model to fit LPs especially in high SNRs (KLD)

DISCUSSION

Listeners' spectral tilt preferences in additive noise were investigated.

<u>In noise, listeners preferred flatter tilts</u>

- facilitates intelligibility due to masking release (*Lu & Cooke, 2009*)

<u>Even with intelligibility at ceiling, listeners showed distinct tilt preference</u>

- in quiet, neutral speech characterised by less flat spectral tilt (*Summers et al, 1988*)
- Lombard speech is affected by the level of background noise. Higher noise level leads to flatter spectral tilt (*Varadarajan & Hansen, 2006*)

Lu and Cooke (2009) Speech Commun. 51, 1253–1262
Summers et al (1988) J. Acoust. Soc. Am. 84, 917–928
Varadarajan and Hansen (2006) Interspeech 937–940

DISCUSSION

In adverse conditions, main concern was optimising intelligibility while in less noisy conditions other factors e.g., speech quality (*Tang et al, 2018*)

- information complementary to objective intelligibility metrics that can be used in NELE algorithms
- SpeechTuner can be used for computing speech feature weighting patterns
- optimal weightings of unseen situations might be acquired by generalising the probabilistic model

Preferences might be linked to listening effort

- listening effort and speech intelligibility can assess speech perception and algorithm performance at high SNR and very low SNR respectively (*Rennies et al, 2018*)

Tang et al (2018) J. Otorhinolaryngol. Hear. Balance Med. 1
Rennies et al (2018) J. Acoust. Soc. Am 144, EL315–EL321

CONCLUSIONS

Main research questions:

1. Do listener preferences show a pattern different from intelligibility?
   yes, in adverse conditions a model based on glimpses could fit LP
2. Do spectral tilt preferences change in challenging conditions?
   increasingly flatter tilts as noise level increases
3. Can listeners' preferences be modelled?
   proposed probabilistic model

SpeechTuner can be used for testing different speech features, stimuli and masking conditions for acquiring:

- listener preferences
- intelligibility scores

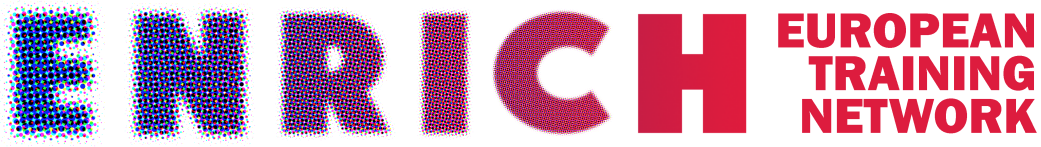
FUTURE WORK

Relation between **listening effort** and **listener preferences** will be investigated.
Measuring:

- listening effort using pupillometry
- listener preferences using SpeechTuner

We believe that listeners choose the speech feature value with which they exert the least listening effort.

Thank you!



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 675324