Investigating the role of harmonic cancellation in masked speech intelligibility

Luna Prud'homme[1], Mathieu Lavandier[1], Virginia Best[2]

1. Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment

2. Department of Speech, Language and Hearing Sciences, Boston University

luna.prudhomme@entpe.fr

In cocktail party scenarios, there are several factors that can improve speech intelligibility: spatial separation, masker amplitude modulation, and differences in harmonicity.

Previous studies (e.g., de Cheveigné et al., 1995; Deroche and Culling, 2011) suggested differences in harmonicity may be exploited via harmonic cancellation, in which listeners suppress the masker based on its F0. However, Leclère et al. (2017) suggested that this effect may be reduced when the masker F0 varies over time (as is the case for speech).

The aim of this study was to investigate the role of harmonic cancellation in speech intelligibility for various kinds of maskers ranging from noise to natural speech.

The experiment was focused on "energetic" aspects of masking and thus "informational" masking was deliberately minimized.

Methods

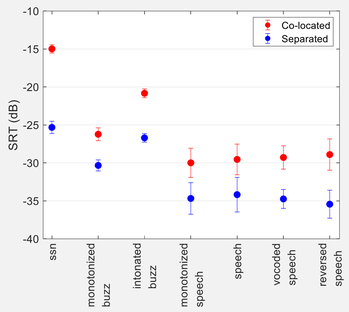
Target: female talker (F0=180 Hz), 0° azimuth , closed set matrix sentences

All maskers derived from a monologue spoken by a male talker (mean F0 = 112 Hz):

- **speech-shaped noise** (average spectrum matching the monologue)
- **monotonized buzz** (harmonic complex, mean F0 112 Hz)
- **intonated buzz** (harmonic complex with F0 contour extracted from the monologue using PRAAT)
- **monotonized speech** (F0 fixed at 112 Hz using PRAAT)
- **speech** (random segment of monologue)
- **vocoded speech** (8 channel vocoder, envelope extracted from monologue)
- **reversed speech** (time-reversed segment of monologue)

Spatial conditions: masker co-located (0°) or separated (60°)
SRTs extracted from psychometric functions (TMRs -40 to -5 dB)

Results

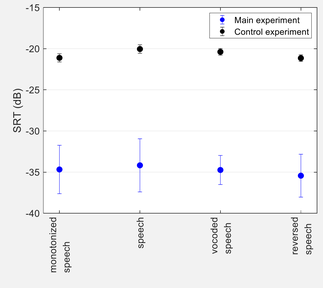


RMANOVA (all maskers)
Spatial condition F(1,9)=1159.5,p<0.001
Masker type F(6,54) = 42.9, p<0.001
Interaction F(6,54) = 14.1, p<0.001

RMANOVA (envelope modulated maskers)
Spatial condition F(1,9)=590.1,p<0.001
Masker type F(3,27) = 0.2, p=0.866
Interaction F(3,27) = 1.6, p=0.224

Main experiment (n = 10)

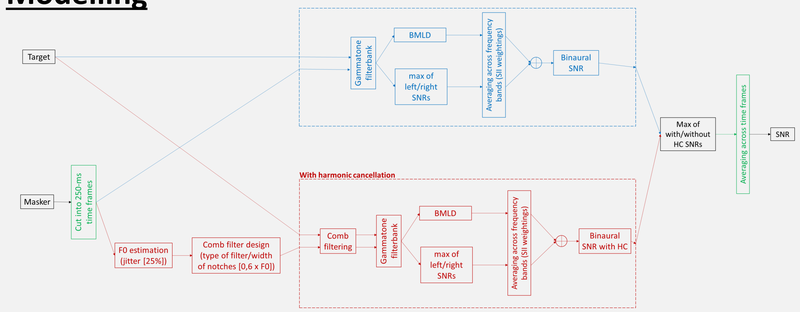
- Lower SRTs for separated vs. co-located : **spatial release from masking**
- Highest SRTs for speech-shaped noise : **no envelope modulation, no harmonicity**
- SRTs lower for monotonized buzz than intonated buzz : **possible role of harmonic cancellation impaired by intonation** (consistent with Leclère et al., 2017)
- SRTs lower for envelope modulated maskers : **listening in the gaps/dips**
- No significant difference between SRTs for all envelope modulated maskers : **minimal influence of harmonicity**
- No difference between forward and reversed speech : **informational masking successfully minimized**



RMANOVA (envelope modulated maskers)
Experiment F(1,4)=258.4,p<0.001
Masker type F(3,12) = 0.4, p=0.757
Interaction F(3,127) = 1.8, p=0.194

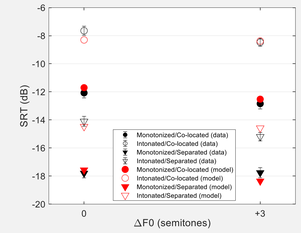
Control experiment (n = 5)
Separated condition re-tested with more challenging open-set target sentences to ensure that the similar SRTs for speech-based maskers were not due to a floor effect. SRTs were higher but still equivalent across maskers.

Modelling



A speech intelligibility model that includes a harmonic cancellation component was developed.

The model can predict differences between SRTs for monotonized and intonated buzzes using data from Leclère et al. (2017).



It is now being modified to handle envelope modulated stimuli and partially harmonic stimuli and will be applied to the current data.

Conclusions

SRTs for buzz maskers provide evidence for harmonic cancellation (Deroche et al., 2014) that would be impaired by intonation (Leclère et al., 2017).

SRTs for envelope modulated maskers suggest that the contribution of harmonic cancellation for speech might be minor. Potential reasons for this include the presence of intonation in speech, the presence of unvoiced parts, and possibly a dominant role of amplitude modulation.

The modelling results will provide further insights regarding these possiblities.

References

Deroche, M., Culling, J., Chatterjee, M., and Limb, C. (2014). "Speech recognition against harmonic and inharmonic complexes: Spectral dips and periodicity," J. Acoust. Soc. Am. 135, 2873-2884.

de Cheveigne, A., McAdams, S., Laroche, J., and Rosenberg, M. (1995). "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement," J. Acoust. Soc. Am. 97, 3736–3748.

Leclère, T., Lavandier, M., and Deroche, M. L. (**2017**). "The intelligibility of speech in a harmonic masker varying in fundamental frequency contour, broadband temporal envelope, and spatial location," Hear. Res., **350**, 1–10. doi:10.1016/j.heares.2017.03.012