



HAL
open science

The Role of Harmonicity in the Masking of Speech: Behavioural and Modelling results

Kurt Steinmetzger, Stuart Rosen

► **To cite this version:**

Kurt Steinmetzger, Stuart Rosen. The Role of Harmonicity in the Masking of Speech: Behavioural and Modelling results. e-Forum Acusticum 2020, Dec 2020, Lyon, France. pp.3487-3488, 10.48465/fa.2020.0080 . hal-03234174

HAL Id: hal-03234174

<https://hal.science/hal-03234174>

Submitted on 26 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE ROLE OF HARMONICITY IN THE MASKING OF SPEECH: BEHAVIOURAL AND MODELLING RESULTS

Kurt Steinmetzger^{1,2}

Stuart Rosen²

¹ Section of Biomagnetism, Department of Neurology, Heidelberg University Hospital, Germany

² Department of Speech, Hearing & Phonetic Sciences, University College London, United Kingdom

kurt.steinmetzger@med.uni-heidelberg.de

ABSTRACT

Harmonic tone complexes are much less effective maskers of speech than a noise with the same overall spectral envelope. Three main factors are thought to contribute to this effect: Firstly, harmonic complexes allow ‘glimpsing’ into frequency regions between harmonics where there is little masker energy. Secondly, the periodicity of the complexes allows them to be more effectively segregated from the target speech. Thirdly, tone complexes contain far fewer envelope modulations than noise, resulting in reduced modulation masking. To estimate their respective contributions, complexes which varied regarding all three factors were devised. Data obtained from normal-hearing listeners revealed that speech reception thresholds (SRTs) were generally worse for complexes with dynamic as opposed to static F_0 contours and improved with increasing masker F_0 , providing evidence for spectral glimpsing. In contrast, inharmonic complexes eliciting a weaker pitch percept led to similar SRTs than harmonic complexes, suggesting that periodicity as such does not aid the segregation of speech and masker. Modulation masking, on the other hand, was found to be an important factor, as SRTs were consistently worse for complexes that contained a broader range of envelope modulations. The predictions of a modulation-based speech intelligibility model, the sEPSM^{corr2}, were able to capture the effects of spectral glimpsing and modulation masking. However, the model generally overestimated the masking effectiveness of the tone complexes, as it does not take pitch-related effects that contribute to increased stream segregation relative to noise maskers into account.

1. INTRODUCTION

Periodic sounds are less effective maskers of speech than aperiodic noises [1,2]. Possible reasons for this effect include inter-harmonic spectral ‘glimpsing’, differences in the patterns of modulation power, and especially periodicity-related voice pitch cues that aid stream segregation.

Regarding the latter factor, the harmonic relation of the component tones in periodic sounds, i.e. the property that the component frequencies are integer multiples of the F_0 , has been claimed to be of particular importance [3]. This so-called harmonic cancellation theory assumes that harmonicity as such enables the auditory system to segregate harmonic maskers from the target speech

signal. Here, we examined the role of harmonicity in complex tone maskers with static and dynamic F_0 contours, in which spectral rotation and spectral shifting eliminated periodicity. Importantly, while both the harmonicity and the periodicity of sounds can be thought to vary continuously, the current study was intended to answer the question if harmonicity and periodicity *per se* determine the effectiveness of a speech masker. Likewise, the pitch strength of a sound will vary along with the degree of harmonicity and periodicity, rather than in a categorical manner.

For the complexes with static F_0 contours, all three versions (harmonic, shifted, and rotated) have very similar modulation spectra, with modulation power primarily in the F_0 region, as well as the same spacing of the spectral components. For the dynamic complexes, spectral components sweeping in frequency create low frequency modulations in addition to those in the F_0 region. The harmonic and shifted stimuli contain similar modulations, but rotated stimuli have very different modulation patterns.

Harmonic stimuli have the clearest pitch, but the F_0 contour can be clearly heard in the shifted and rotated sounds too. However, the rotated stimuli have a notably rougher timbre, particularly with dynamic F_0 contours.

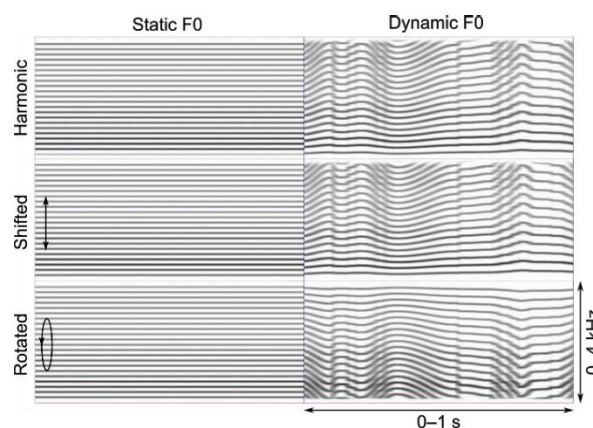


Figure 1. Maskers. The maskers were harmonic complex tones and inharmonic versions thereof, where the component tones were either shifted in frequency or spectrally rotated.

2. METHODS

Twelve normal-hearing listeners were tested (mean age = 22.8 years). The ABC sentences spoken by an adult male

Southern British English talker (median F_0 ~150 Hz) were used as target materials.

The maskers were harmonic complex tones that either had static or dynamically varying F_0 contours. Inharmonic equivalents of these maskers were produced by either shifting the component tones up or down by 25% of the median F_0 or by rotating the spectra around 2 kHz (Fig. 1). The resulting six masker types were each presented with a low (100 Hz), mid (150 Hz), or high (225 Hz) median F_0 level. For comparison, speech-shaped noise was included as additional masker, resulting in 19 stimulus conditions in total.

Speech reception thresholds (SRTs) at the 50%-correct level were determined for each condition using 20 ABC sentences. Additionally, SRTs were estimated using the sEPSM^{corr2} speech intelligibility model [4].

The sEPSM^{corr2} is characterised by the combination of a modulation filterbank in the front end and a correlation-based decision back end. An evaluation of several intelligibility models [4] has shown that this model performed best in estimating the intelligibility of speech in the presence of harmonic complex maskers.

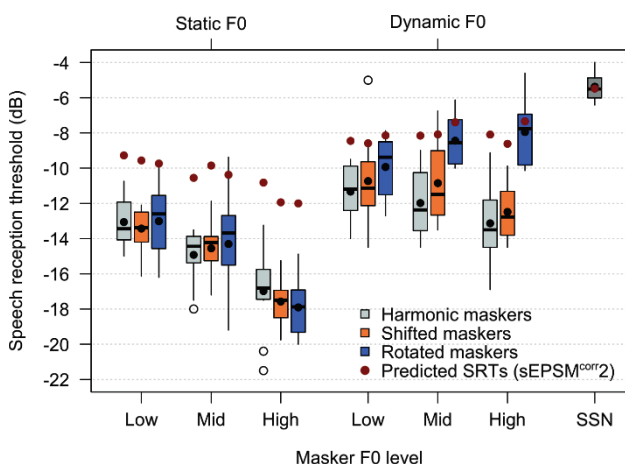


Figure 2. Results. The behavioural data are represented by the boxplots and model results are indicated by the superimposed red dots.

3. RESULTS

The behavioural data were analysed by fitting a general linear mixed-effects regression model in a top-down manner. The main effects of masker harmonicity [$F(2,176.24) = 31.78, p < 0.001$], masker F_0 level [$F(2,172.03) = 66.96, p < 0.001$] and masker F_0 contour [$F(1,174.09) = 545.19, p < 0.001$] were all highly significant.

For the static F_0 maskers, there was no significant main effect of masker harmonicity [$F(2,81.77) = 0.20, p = 0.821$], but a highly significant main effect of masker F_0 level [$F(2,75.91) = 102.32, p < 0.001$].

For the dynamic F_0 maskers, a highly significant main effect of masker harmonicity [$F(2,80.15) = 71.91, p < 0.001$], a significant main effect of masker F_0 level [$F(2,73.10) = 3.35, p < 0.04$], and also a highly significant interaction of the two factors [$F(4,74.28) = 8.83, p < 0.001$] were observed.

The modelling results are plotted together with the behavioural data in Fig. 2. While the sEPSM^{corr2} could account for some of the main trends in the data, the effectiveness of the maskers was over-estimated considerably.

4. DISCUSSION

The results for the harmonic and shifted complexes imply no role for harmonicity *per se*, irrespective of whether the maskers had static or dynamic F_0 contours. This is contrast to the harmonic cancellation theory [3].

For the spectrally rotated maskers, the results are more complicated. While performance was similar for the rotated, shifted, and harmonic maskers when their F_0 contours were static, the SRTs for the rotated maskers showed a different pattern when the F_0 contours were dynamic. Here, higher F_0 levels led to poorer performance, while speech intelligibility increased with F_0 level for the other two masker types.

One explanation for this may be that the spectral rotation eliminates the advantage of the wider spacing of spectral components caused by increases in F_0 (cf. Fig. 1). Secondly, the rotated maskers have a very unusual timbre which may be a limiting factor in stream segregation. Thirdly, the modulation filterbank output of the sEPSM^{corr2} model showed more modulation power in low-frequency auditory filters for this masker type, an effect that increased with masker F_0 level.

Regarding the predictions of the sEPSM^{corr2}, it appears that the main limitation of the model is its inability to account for pitch-related streaming effects. In line with previous evaluations of the model [4], speech intelligibility in the presence of complex tone maskers was underestimated throughout.

5. REFERENCES

- [1] K. Steinmetzger and S. Rosen: "The role of periodicity in perceiving speech in quiet and in background noise," *J. Acoust. Soc. Am.*, 138, pp. 3586–3599, 2015.
- [2] M. Deroche, J. Culling, M. Chatterjee, and C. Limb: "Speech recognition against harmonic and inharmonic complexes: Spectral dips and periodicity", *J. Acoust. Soc. Am.*, 135, pp. 2873–2884, 2014.
- [3] A. de Cheveigné, S. McAdams, J. Laroche, and M. Rosenberg: "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement" *J. Acoust. Soc. Am.*, 97, pp. 3736–3748, 1995.
- [4] K. Steinmetzger, J. Zaar, H. Relañó-Iborra, S. Rosen, and T. Dau: "Predicting the effects of periodicity on the intelligibility of masked speech: an evaluation of different modelling approaches and their limitations," *J. Acoust. Soc. Am.*, 146, pp. 2562–2576, 2019.