Redundant information is sometimes more beneficial than spatial information to understand speech in noise

Benjamin Dieudonné, Tom Francart

Experimental Oto-rhino-laryngology, Department of Neurosciences, KU Leuven –

University of Leuven, Belgium.

Address correspondence to:

Tom Francart, Experimental Oto-rhino-laryngology, Department of Neurosciences, KU Leuven – University of Leuven, Herestraat 49 bus 721, 3000 Leuven, Belgium.

E-mail: tom.francart@med.kuleuven.be, benjamin.dieudonne@med.kuleuven.be

Abstract

Objectives: To establish a framework to unambiguously define and relate the different spatial effects in speech understanding: head shadow, redundancy, squelch, spatial release from masking, etc. Next, to investigate the contribution of interaural time and level differences to these spatial effects in speech understanding, and how this is influenced by the type of masking noise.

Design: In our framework, spatial release from masking is uniquely characterized as a linear combination of head shadow, binaural redundancy and binaural squelch. The latter two terms are combined into one binaural term, which we define as binaural contrast: a benefit of interaural differences. In this way, spatial release from masking is a simple sum of a monaural and a binaural term. We used the framework to quantify these spatial effects in ten normal-hearing listeners. The participants performed speech intelligibility tasks in different spatial set-ups. We used head-related transfer functions to manipulate the presence of interaural time and level differences. We used three spectrally-matched masker types: stationary speech-weighted noise, a competing talker, and speech-weighted noise that was modulated with the broadband temporal envelope of the competing talker.

Results: We found that (1) binaural contrast was increased by interaural time differences, but reduced by interaural level differences, irrespective of masker type, and (2) large redundancy (the benefit of having identical information in two ears) could reduce binaural contrast, and thus also reduce spatial release from masking.

Conclusions: Our framework yielded new insights in binaural processing in speech intelligibility. Firstly, interaural level differences disturb speech intelligibility in realistic listening conditions. Therefore, to optimize speech intelligibility in hearing aids, it is more beneficial to improve monaural signal-to-noise ratios rather than to preserve interaural level differences. Secondly, although redundancy is mostly ignored when considering spatial hearing, it might explain reduced spatial release from masking in some cases.

Keywords: speech intelligibility; binaural hearing; spatial release from masking; head shadow; binaural intelligibility level difference; binaural squelch; binaural redundancy; binaural contrast; binaural interaction; interaural time differences; interaural level differences; binaural information; binaural cues; normal-hearing

I   Introduction

The importance of binaural hearing to understand speech in complex listening environments is well-established (e.g., Blauert, 1997, 2013; Dillon, 2001). One can test the use of binaural cues by measuring speech intelligibility in noise in different spatial set-ups. For example, one can compare intelligibility in a set-up with target speech and masking noise spatially collocated in front of the listener, with the intelligibility with the masker separated at the right ear. Spatial release from masking (SRM) is then defined as the improvement in speech intelligibility due to the spatial separation of target and masker. This is an often-used measure to quantify the performance of the binaural system with or without auditory prostheses (e.g., Litovsky, 2005; Arbogast et al., 2005; Veugen et al., 2016).

SRM is however hard to interpret quantitatively, because different mechanisms are involved (Plomp, 1976). First, monaural cues change: due to the head shadow effect, the signal-to-noise ratio (SNR) in one ear increases while the SNR in the other ear decreases. Second, binaural cues arise: interaural level differences (ILDs) and interaural time differences (ITDs) provide the listener with spatial information about the sound sources, which can be used to differentiate between speech and noise. The interpretation of SRM becomes even more difficult yet of greater importance when hearing aids are involved, since their signal processing might distort these cues.

Therefore, many measures have been introduced to gain more insight in SRM and the different mechanisms behind it: head shadow, redundancy, squelch, binaural interaction, binaural interference, binaural intelligibility level difference (BILD), binaural advantage, binaural summation, etc. (Plomp, 1976; Dillon, 2001; Van Deun et al., 2010). Unfortunately, the exact definitions of these phenomena vary across different papers. For example, BILD is sometimes defined as the difference in speech understanding between diotic and dichotic stimulation (Dillon, 2001), while otherwise defined as the difference in speech understanding between monaural and binaural stimulation (Peissig and Kollmeier, 1997). These ambiguities result in discussions that are unintentionally semantic and can lead to poorly drawn conclusions about the binaural system.

A Theoretical framework to disentangle spatial release from masking

Here, we disentangle SRM into the following effects by including monaural speech intelligibility tests: (monaural) head shadow, (binaural) redundancy and (binaural) squelch (Van Deun et al., 2010). For this purpose, we established a new framework which enables us to unambiguously define and relate these spatial effects in speech understanding, as depicted in Figure 1.

Head shadow is a purely physical effect: the increase in SNR in one ear due to spatial separation of speech and noise results in a monaural improvement in speech intelligibility. For example, when the masker is separated towards the right ear, the SNR in the left ear increases due to the head shadow; the left ear is then called the better ear.

Redundancy (also called binaural summation[1]) is the benefit of listening with two ears with identical information, as compared to listening with only one ear; this effect is measured with target and masker collocated in front of the listener. It can be interpreted as internal noise cancellation due to the summation of signals arriving in both ears.

Squelch is the benefit of listening with two ears as compared to listening with the better ear only (the ear with higher SNR); this effect is measured with spatially separated target and masker. This is a remarkable effect, since it is in improvement in speech intelligibility by supplying inferior information (an ear with a lower SNR) than the information that is already available monaurally. However, by adding this worse ear, spatial information (interaural differences) is supplied to the listener, which makes it easier to extract and concentrate on the target speech (either by peripheral noise cancellation or more central attentional mechanisms). Moreover, the worse ear also supplies redundant information which enables internal noise cancellation. The latter is an often-ignored effect in the discussion of squelch.

Finally, we define **binaural contrast** as the difference between squelch and redundancy [2]. In this way, squelch is the sum of binaural redundancy and binaural contrast,

---

[1] We prefer the term redundancy over summation, as it emphasizes the cue that is used in this effect (redundant information), rather than speculating about the mechanism behind it (summing of signals).

[2] This effect has also been referred to as binaural interaction by Hawley et al. (2004). However, the term "binaural interaction" is often used in more general contexts (e.g., Colburn and Durlach, 1978) and can therefore be confusing. The term "contrast" emphasizes the cue that is used (interaural differences), as with the term "redundancy".

such that it is clear that squelch is both due to interaurally identical information and interaural differences. Note that there is only a benefit of binaural contrast when squelch is larger than redundancy: only then it is certain that there is true binaural processing due to interaural differences. If squelch is smaller than or equal to redundancy, the squelch benefit might be due to redundant information only.

If we consistently apply these definitions, we can easily disentangle SRM:

$$\text{spatial release from masking} = \text{head shadow} + \text{squelch} - \text{redundancy}$$

$$= \text{head shadow} + \text{binaural contrast} \tag{1}$$

In words, SRM is the sum of head shadow and binaural contrast, i.e., the sum of a monaural term and a binaural term.
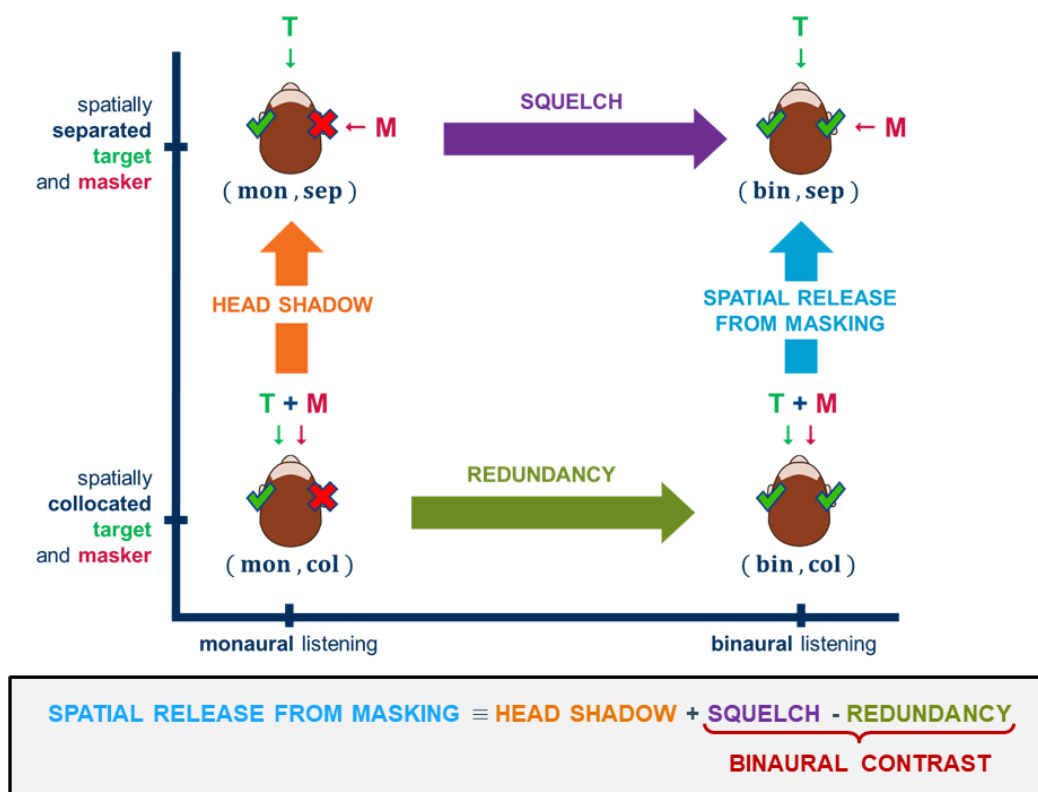

Figure 1: Framework to disentangle spatial release from masking. Within this framework, SRM is uniquely characterized by head shadow, redundancy and squelch. True binaural processing can be investigated by measuring the difference between squelch and redundancy, which we refer to as binaural contrast. With these definitions, SRM equals the sum of head shadow and binaural contrast, i.e., the sum of a monaural and a binaural component.

B   Behind the separate spatial effects

Although the usage of binaural cues has been investigated by many researchers, there still appears to be discussion on how these cues contribute to SRM (Glyde et al., 2013). The separate contribution of ILDs and ITDs to head shadow and redundancy is trivial: head shadow is only mediated by (interaural) level differences, since time differences have no meaning in a monaural set-up; redundancy does not require any spatial cues, since it is defined for spatially collocated target and masker. Therefore, it remains to be discussed how binaural cues contribute to squelch (as SRM and binaural contrast follow directly from head shadow, redundancy and squelch).

Some suggest that SRM is merely due to physical parameters induced by spatial separation, as opposed to the spatial locations themselves: a monaural component (head shadow) mediated by ILDs and a binaural component (binaural contrast) mediated by ITDs (Bronkhorst and Plomp, 1988; Edmonds and Culling, 2005). Due to ITDs, a linear combination (conducted in the auditory periphery) of left and right ear signals can cancel out masking noise, while preserving the signal of interest. Bronkhorst and Plomp (1988) were indeed not able to observe squelch (and therefore binaural contrast) when ITDs were not present: adding a worse ear with only ILDs did not result in better speech intelligibility, whereas adding a worse ear with ILDs and ITDs did result in better speech intelligibility. Edmonds and Culling (2005) investigated whether SRM could be reduced by conflicting spatial cues: the ILDs suggested that speech was coming from the right, whereas the ITDs suggested that speech was coming from the left. While Bronkhorst and Plomp (1988) only investigated an energetic masker (speech-weighted noise), Edmonds and Culling (2005) did their experiment with an energetic (Brown-noise) and an informational (competing-speech) masker. For an informational masker, one could expect the perceived direction to be more important (Noble and Perrett, 2002). However, they could not find a significant difference between the conditions with conflicting and non-conflicting spatial cues for either masker type. This suggests that spatial release from masking is not driven by focusing attention on the perceived direction of target and masker – in contrast to what the name suggests – irrespective of the masker type. However, as the authors stated themselves, this reasoning is not completely solid, since conflicting spatial cues mostly lead to a perceived direction dominated by the ITD (Wightman and Kistler, 1992).

Others believe indeed that higher-level mechanisms are involved in SRM, and the benefit of interaural differences (what we call binaural contrast) also occurs by focusing attention to a perceived direction (Shinn-Cunningham et al., 2005; Gallun et al., 2005). Shinn-Cunningham et al. (2005) tested whether the amount of SRM depends on the realism of the spatial cues for an informational masker. They minimized energetic masking by vocoding speech and noise with different sinusoidal carriers, such that there was very little spectral overlap between speech and noise. They presented either all spatial cues, or only ILDs and ITDs in the envelope (i.e. zero interaural phase difference in the carrier), or only ITDs of the carrier (i.e. realistic interaural phase differences). For the same (broadband) SNR in one ear (the better ear), a condition with a worse SNR in the other ear yielded better speech intelligibility than a condition with the same SNR in the other ear, irrespective of the spatial cues transmitted. This means that squelch was larger than redundancy in all conditions; in other words, they were able to measure binaural contrast in all conditions. However, in each condition ITDs were presented, either in the envelopes or in the carrier; although it is unlikely with only envelope ITDs, the binaural contrast may still have been mediated by these ITDs. Gallun et al. (2005) did a similar experiment with vocoded target and masker to minimize energetic masking. They did report a benefit of adding a worse ear with only ILDs, and again squelch was larger than redundancy. However, the ILDs were non-realistic, broadband ILDs (as opposed to realistic, frequency-dependent ILDs).

These results are often cited as evidence that ITDs and ILDs both contribute to squelch (and thus binaural contrast), without mentioning the specific conditions in which this was shown (e.g., Dillon, 2001). Both Shinn-Cunningham et al. (2005) and Gallun et al. (2005) made use of artificial stimuli to show higher-level mechanisms in squelch: sounds were vocoded, and target and masker had exactly the same sentence structure (the coordinate response measure of Bolia et al., 2000). Moreover, they only measured percentages of correct responses at fixed SNRs, which makes it hard to compare their results with elsewhere reported squelch benefits measured in dB (e.g. Bronkhorst and Plomp, 1988). Finally, while Gallun et al. (2005) were able to measure squelch and binaural contrast in a situation with only ILDs, they made use of non-realistic, broadband ILDs. Altogether, it is difficult to judge how these results can be translated towards more realistic situations.

The question remains to what extent a benefit of binaural contrast may occur due to head-induced ILDs for non-vocoded stimuli. This is important in the decision whether novel signal processing strategies should preserve (or enhance) ILDs considering speech intelligibility. If not, it might be more interesting to optimize monaural SNRs rather than to preserve faithful ILDs.

We present an experiment in which we investigate (1) the contribution of ILDs and ITDs to the different spatial effects involved in speech understanding, and (2) how these spatial effects are influenced by the type of masking noise. We use our theoretical framework of SRM to quantify the spatial effects unambiguously. We make use of realistic (non-vocoded) stimuli and realistic binaural cues.

II Materials and Methods

A Participants

Ten normal-hearing participants were recruited, aged between 23- and 29-years-old. Their pure-tone hearing thresholds were better than 20 dBHL at all octave audiometric frequencies from 250 to 8000 Hz. The study was approved by the Medical Ethical Committee of the University Hospital Leuven (S58970).

B Apparatus

All tests took place in a double-walled soundproof booth. We presented our stimuli using the software platform APEX 3 (Francart et al., 2008). The stimuli were presented through Sennheiser HDA200 over-the-ear headphones via an RME Hammerfall DSP Multiface soundcard.

C Spatial simulation

We simulated spatial hearing using head-related transfer functions (HRTFs). Spatial conditions are coded according to Figure 1: stimulation is either monaural (mon) or binaural (bin), with target and masker being either collocated (col) or separated (sep). By manipulating the HRTFs, we were able to choose which spatial cues to deliver: only ILDs (sep_{ILD}), only ITDs (sep_{ITD}), or all cues (sep_{all}) (Bronkhorst and Plomp, 1988).

We measured the HRTFs of a human-like acoustical manikin (head and torso) in a localization arc, for sounds coming from 0° and 90° in the azimuthal plane (i.e., in front and at the right of the manikin respectively). Loudspeakers were placed at a distance of 1 m from the center of the manikin's head, at the same height as its ears. The room had sound-dampening curtains to reduce reflections. The reverberation time (time for the reflections to decay 60 dB, measured between 160 Hz and 8000 Hz per 1/3 octave and then averaged with a speech intelligibility index (SII) weighting (ANSI, 1997)) was 0.15 s. The HRTFs were equalized with respect to the acoustical path from headphone transducer to microphone of the manikin, such that the ear-canal resonance would only occur once in the eventual presentation of the stimuli. The frequency-dependent ILDs and ITDs resulting from these HRTFs are plotted in Figure 2.
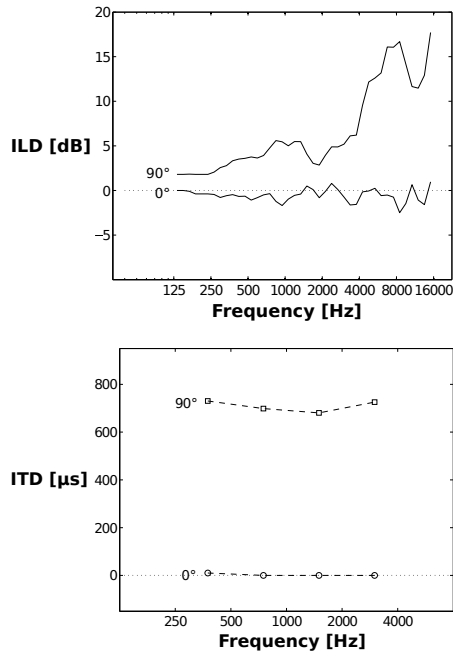

Figure 2: Frequency-dependent ILDs and ITDs for the HRTFs recorded from the manikin.

From these two HRTFs we derived two artificial HRTFs, containing only ILDs (i.e., ITD = 0 for all sounds) or only ITDs (i.e., ILD = 0 dB for all sounds) respectively. For the ILD-only HRTF we used the magnitude response of the 90° HRTF and the the phase response of the 0° HRTF; for the ITD-only HRTF we used the phase response of the 90° HRTF and the the magnitude response of the 0° HRTF.

In the end, we had 4 different HRTFs, corresponding to 4 different spatial conditions: no cues (noise at 0°, i.e., col), only ILDs (sep$_{\text{ILD}}$), only ITDs (sep$_{\text{ITD}}$) and all cues (noise

9

at 90°, i.e., sep$_{all}$). Target speech was always presented at 0°.

D   Stimuli

We used the Flemish Matrix sentence test as target speech (Luts et al., 2015). It consists of 13 lists of 20 sentences uttered by a female speaker. Each sentence has the same grammatical structure (name, verb, numeral, adjective, object).

We compared three different masker types: speech-weighted noise (SWN), a competing talker and modulated SWN. The SWN was a stationary noise constructed from the Matrix sentences, such that it had the same long-term average spectrum (Luts et al., 2015). The competing talker was a Swedish story uttered by a female speaker. We chose a Swedish story because like Dutch it is a Germanic language, while our participants could not understand it. In this way, we could investigate the effect of temporal fine structure, while avoiding informational masking due to failure of object selection (Francart et al., 2011). We chose the masker such that the fundamental frequency was comparable to the one of target speech: around 200 Hz (calculated using Praat (Boersma et al., 2002)). We reduced silent gaps longer than 100 ms to 100 ms, flattened the time-dependent root-mean-square level, and filtered the signal to obtain the same average spectrum as the SWN. The story still sounded like natural speech after these operations. The modulated SWN was constructed by modulating the SWN with the broadband envelope of the competing talker. We determined this envelope by full-wave rectification and low-pass filtering of the speech signal. We chose this masker to investigate the effect of temporal gaps separately, without having the effect of temporal fine structure.

In the end, all maskers had the same long-term average spectrum and the same root-mean-square level.

E   Procedure

For each condition, we measured the speech reception threshold (SRT), defined as the SNR at which 50% of speech could be understood. We did this according to the adaptive procedure as described by Brand and Kollmeier (2002). Each measurement consisted of 20 trials. For each trial, the magnitude of the SNR adaptation depended on the participant's score of the previous trial and on the trial number, with SNR adaptation getting smaller towards the end of the procedure. The masker was presented continuously at a level of

65 dB SPL during each run, while the speech level was set according to the presented SNR. For each measurement, we estimated the SRT as the SNR that was calculated based on the response on the last trial.

A condition was defined by three factors: noise type, ears (monaural (mon) or binaural (bin)) and spatial cues (collocated (col) or separated (sep), with sep being sep_{ILD}, sep_{ITD} or sep_{all}). For each noise type, conditions were coded as SRT(ears,spatial cues). An overview of all conditions (per noise type) is given in Table 1. Note that for monaural SRTs, ITDs have no influence on the HRTFs:

$$SRT(mon,col) \equiv SRT(mon,sep_{ITD}) \tag{2a}$$

$$SRT(mon,sep_{all}) \equiv SRT(mon,sep_{ILD}) \tag{2b}$$

Therefore, we end up with 6 different conditions per noise type. For each participant, we performed each measurement twice to test for learning effects and to reduce random variability in the results. We ended up with a total of 3 (noise types) $\times$ 6 (SRTs(ears,spatial cues)) $\times$ 2 (repetitions) = 36 measurements for each subject. We performed the tests in blocks per noise type, while randomizing the order of these blocks and randomizing the conditions within each block. Before each block, we did 2 training measurements to get used to the respective noise type.

Table 1: An overview of the different conditions that were tested for each noise type. Note that $SRT(mon,col) \equiv SRT(mon,sep_{ITD})$ and $SRT(mon,sep_{all}) \equiv SRT(mon,sep_{ILD})$ (because ITDs have no influence on monaural SRTs), such that there were only 6 unique conditions to be tested.

	monaural	binaural
collocated	SRT(mon,col)	SRT(bin,col)
separated (all cues)	$SRT(mon,sep_{all})$	$SRT(bin,sep_{all})$
separated (only ILD)	$SRT(mon,sep_{ILD})$	$SRT(bin,sep_{ILD})$
separated (only ITD)	$SRT(mon,sep_{ITD})$	$SRT(bin,sep_{ITD})$

F   Calculation of spatial effects

We calculated the magnitude of each spatial effect according to our theoretical framework of SRM (Figure 1). An overview of all definitions is given in Table 2. Each effect is quantified as the additive inverse of the difference between two SRTs, such that a positive value corresponds to an improvement in speech intelligibility due to the respective spatial effect.

Table 2: An overview of all spatial effects, according to the theoretical framework of Figure 1. Note that a positive value corresponds to an improvement in speech intelligibility due to the respective spatial effect

	Effect size [dB SNR]
head shadow	$= -[\text{SRT}(\text{mon,sep}_{\text{all}}) - \text{SRT}(\text{mon,col})]$
redundancy	$= -[\text{SRT}(\text{bin,col}) - \text{SRT}(\text{mon,col})]$
squelch$_{\text{all}}$	$= -[\text{SRT}(\text{bin,sep}_{\text{all}}) - \text{SRT}(\text{mon,sep}_{\text{all}})]$
squelch$_{\text{ILD}}$	$= -[\text{SRT}(\text{bin,sep}_{\text{ILD}}) - \text{SRT}(\text{mon,sep}_{\text{ILD}})]$
squelch$_{\text{ITD}}$	$= -[\text{SRT}(\text{bin,sep}_{\text{ITD}}) - \text{SRT}(\text{mon,sep}_{\text{ITD}})]$
binaural contrast$_{\text{all}}$	$= \text{squelch}_{\text{all}} - \text{redundancy}$
binaural contrast$_{\text{ILD}}$	$= \text{squelch}_{\text{ILD}} - \text{redundancy}$
binaural contrast$_{\text{ITD}}$	$= \text{squelch}_{\text{ITD}} - \text{redundancy}$
SRM$_{\text{all}}$	$= -[\text{SRT}(\text{bin,sep}_{\text{all}}) - \text{SRT}(\text{bin,col})]$
SRM$_{\text{ILD}}$	$= -[\text{SRT}(\text{bin,sep}_{\text{ILD}}) - \text{SRT}(\text{bin,col})]$
SRM$_{\text{ITD}}$	$= -[\text{SRT}(\text{bin,sep}_{\text{ITD}}) - \text{SRT}(\text{bin,col})]$

III   Results

All analyses were done by means of linear mixed models with subject as random factor. We report our data as mean values $\pm$ 1 standard deviation (s.d.). A difference is considered significant if $p < 0.05$.

A SRTs

We started with investigating a possible learning effect in our data, i.e., a significant difference between repetitions for the same condition. For this purpose, we investigated the SRTs with a linear mixed model with variable SRT and factors noise type (3 levels), SRT-condition (6 levels) and repetition (2 levels). Both noise type and SRT-condition were significant predictor variables [$F(2, 335) = 2710.72, p < 0.001$ and $F(5, 335) = 241.08, p < 0.001$ respectively]. We also found a significant learning effect of 0.76 ± 1.32 dB [$F(1, 335) = 39.94, p < 0.001$]. Therefore, we continued our analysis by including interaction terms of repetition with noise type and repetition with SRT-condition in our model. None of these interactions were significant [$F(2, 335) = 1.86, p = 0.16$ and $F(5, 335) = 0.96, p = 0.44$ respectively], so we concluded that we could average both repetitions for all other analyses. The resulting SRTs are reported in Table 3.

Table 3: Mean and standard deviation (s.d.) of speech reception thresholds (SRTs) for the different spatial conditions.

	SWN				Modulated SWN				Competing talker			
	mon		bin		mon		bin		mon		bin	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
col	−9.6	1.0	−10.4	1.0	−15.0	1.8	−16.3	1.7	−20.8	2.0	−23.5	2.0
sep$_{\text{all}}$	−14.1	1.2	−16.7	1.2	−18.9	2.3	−21.5	1.7	−23.4	2.5	−26.6	2.3
sep$_{\text{ILD}}$	−14.1	1.2	−14.5	1.3	−18.9	2.3	−19.7	1.8	−23.4	2.5	−25.0	2.2
sep$_{\text{ITD}}$	−9.6	1.0	−13.8	1.3	−15.0	1.8	−19.0	2.0	−20.8	2.0	−24.5	2.0

Next, we investigated the effect of masker type on SRT. To be able to compare our results with literature, we only considered SRT(mon,col) in this analysis. We calculated a linear mixed model with variable SRT and factor noise type. Noise type was a significant predictor of SRT in our model [$F(2, 18) = 524.18, p < 0.001$]. SRTs were -9.6 ± 1.0 dB, -15.0 ± 1.8 dB and -20.8 ± 2.0 dB for SWN, modulated SWN and the competing talker respectively (Figure 3).
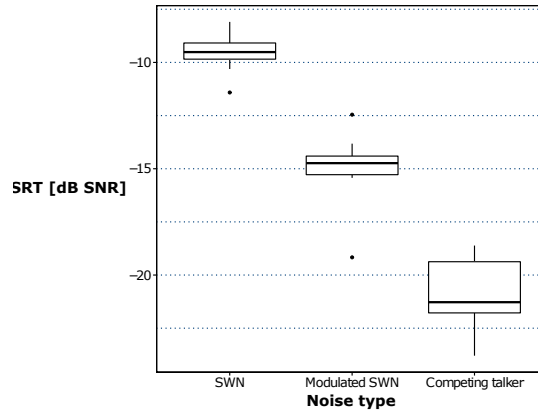
Figure 3: Monaural speech reception thresholds (SRTs) with collocated target and masker for different noise types. It can be seen that both temporal gaps (modulated speech-weighted noise (SWN) and competing talker) and temporal fine structure (competing talker) were beneficial for speech understanding, as lower speech reception thresholds (SRTs) correspond to better intelligibility.

B   Spatial effects

For all following analyses, we transformed our SRT data to the different spatial effects according to our framework (Figure 1) with the equations of Table 2. The results are displayed in Table 4 and Figure 4. We discuss the separate effects in increasing order of complexity: head shadow, redundancy, squelch and binaural contrast, SRM.

Table 4: Different spatial effects according to our framework (Figure 1) obtained with the equations of Table 2.

|	SWN		Modulated SWN		Competing talker	
	mean	s.d.	mean	s.d.	mean	s.d.
head shadow	4.6	0.6	3.9	0.9	2.5	1.3
redundancy	0.9	0.4	1.4	0.6	2.6	0.8
squelch$_{all}$	2.5	0.3	2.6	1.1	3.2	1.4
squelch$_{ILD}$	0.4	0.3	0.8	1.0	1.7	1.2
squelch$_{ITD}$	4.2	0.7	4.0	0.8	3.7	1.1
binaural contrast$_{all}$	1.7	0.4	1.2	1.2	0.6	2.1
binaural contrast$_{ILD}$	−0.5	0.5	−0.6	0.7	−0.9	1.8
binaural contrast$_{ITD}$	3.4	0.6	2.7	0.7	1.1	1.1
SRM$_{all}$	6.2	0.4	5.2	0.6	3.1	1.5
SRM$_{ILD}$	4.1	0.7	3.3	0.7	1.6	1.4
SRM$_{ITD}$	3.4	0.6	2.7	0.7	1.1	1.1
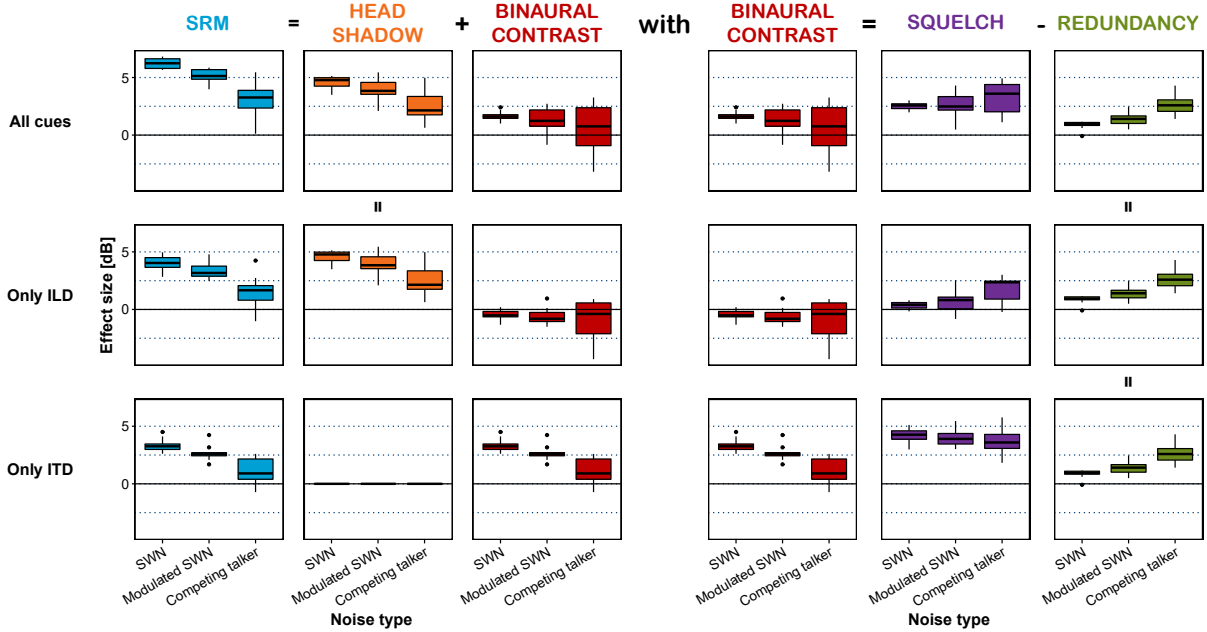
Figure 4: Every column corresponds to a spatial effect according to our framework (Figure 1), calculated for each participant with the equations of Table 2. As such, the effect of binaural cues and masker type on spatial release from masking (SRM) can be disentangled into a monaural and a binaural component: head shadow and binaural contrast. Our most important findings were that (1) binaural contrast was improved by interaural time differences, but disturbed by interaural level differences, and (2) large redundancy could reduce binaural contrast, and thus also reduce SRM.

Head shadow

As explained in the introduction, head shadow is merely determined by the presence of ILDs in the HRTF. Therefore, the effect is equal when presenting either all cues or only ILDs, and zero when only ITDs are presented. It only remains us to investigate the effect of noise type on head shadow. We found a significant effect of noise type on head shadow $[F(2, 18) = 16.11, p < 0.001]$. The head shadow effect was not significantly different for modulated SWN as compared to stationary SWN $[t(18) = -1.70, p = 0.11]$; it was significantly smaller for the competing talker as compared to stationary SWN $[t(18) = -5.54, p < 0.001]$.

Redundancy

Redundancy is per definition measured for spatially collocated speech and noise. There-fore, no spatial cues are presented, and their effect should not be investigated. As for head shadow, it only remains us to investigate the effect of noise type. We found a significant effect of noise type on redundancy $[F(2, 18) = 21.76, p < 0.001]$. Redun-dancy was not significantly different for modulated SWN as compared to stationary SWN $[t(18) = 1.83, p = 0.084]$; it was significantly larger for the competing talker as compared to stationary SWN $[t(18) = 6.40, p < 0.001]$.

Squelch and binaural contrast

For squelch, we were particularly interested in the effect of the spatial cues (i.e., whether there is a benefit of binaural contrast), the effect of noise type and the interaction of both factors. In the analysis of squelch, we also included redundancy as a baseline condition and coded it as "squelch with no spatial cues". In this way, we could investigate binaural contrast within the same statistical model: significant binaural contrast is implied by a significant difference between squelch and redundancy.

We found a significant interaction of spatial cues and noise type $[F(6, 99) = 3.91, p = 0.0015]$. This could have been expected from Figure 4: the effect of noise type seems absent for the condition with only ITDs. Therefore, we split our squelch data in two sub-sets (while including redundancy in both subsets): (1) a subset for the spatial condition with only ITDs, and (2) a subset with all other spatial conditions (i.e., all cues and only ILDs).

In the first subset (squelch$_{\text{ITD}}$), there was still a significant interaction of spatial cues (only ITDs or no cues) and noise type $[F(2, 45) = 12.00, p < 0.001]$. This means that the difference between squelch$_{\text{ITD}}$ and redundancy ("squelch with no spatial cues") depended on the noise type; in other words, binaural contrast was dependent on noise type when only ITDs were presented. For another subset with only ITDs (without the condition with no spatial cues), only the noise type remains as a predictor variable. We found no significant effect of noise type on squelch with only ITDs $[F(2, 18) = 0.89, p = 0.43]$.

In the second subset (squelch$_{\text{all}}$ and squelch$_{\text{ILD}}$), there was no more interaction of spatial cues (all cues or only ILDs or no cues) and noise type $[F(4, 72) = 1.12, p = 0.35]$. This means that binaural contrast was independent of masker type when all spatial cues

or only ILDs were presented. For the following analysis, we left out the interaction term. Noise type and spatial condition were both significant predictors [$F(2, 72) = 19.73, p < 0.001$ and $F(2, 72) = 40.36, p < 0.001$ respectively]. As for redundancy, squelch was not significantly different for modulated SWN as compared to stationary SWN [$t(76) = 1.55, p = 0.13$]; it was significantly different for the competing talker as compared to stationary SWN [$t(76) = 6.01, p < 0.001$]. Moreover, we found that squelch with all spatial cues was significantly larger than redundancy [$t(76) = 5.57, p < 0.001$], and squelch with only ILDs was significantly smaller than redundancy [$t(76) = -3.26, p < 0.001$]. In other words, when ILDs and ITDs were presented, we were able to detect a benefit of binaural contrast; when only ILDs were presented, binaural contrast was negative.

Spatial release from masking

For SRM, we also investigated the effect of spatial cues, the effect of noise type and the interaction of both factors. Since the interaction was not significant [$F(4, 72) = 0.82, p = 0.51$], we left out the interaction term in our further analysis. We found that both spatial cues and noise type were significant predictors of SRM [$F(2, 76) = 83.36, p < 0.001$ and $F(2, 76) = 91.50, p < 0.001$ respectively].

SRM was significantly smaller for modulated SWN as compared to stationary SWN, and also significantly smaller for the competing talker as compared to stationary SWN [$t(76) = -4.16, p < 0.001$ and $t(76) = -13.23, p < 0.001$ respectively]; SRM was also smaller when only ILDs or only ITDs were presented as compared to all spatial cues [$t(76) = -9.30, p < 0.001$ and $t(76) = -12.41, p < 0.001$ respectively].

IV   Discussion

Spatial release from masking (SRM) is an often-used measure to quantify the performance of the binaural system. However, it is hard to interpret quantitatively. Therefore, we developed a theoretical framework to quantify SRM as a linear combination of different spatial effects (see Figure 1): (monaural) head shadow, binaural redundancy, binaural squelch and binaural contrast. We used the framework to investigate the contribution of binaural cues to these spatial effects, and how this is influenced by the type of masking

noise. This is important in the decision whether novel hearing aid processing should preserve (or enhance) binaural cues to optimize speech intelligibility.

A   SRTs

The monaural SRT in SWN agreed very well with the expected SRT of -9.0 dB SNR (Luts et al., 2015).

We found better speech understanding when temporal gaps were present in the masker than when these gaps were not present. Speech understanding was even better when temporal fine structure was present in the masker, even though target speech and masker had the same gender and approximately the same fundamental frequency. Temporal gaps and temporal fine structure in a masker are well-known to be beneficial for speech understanding in normal-hearing listeners (Festen and Plomp, 1990; Francart et al., 2011).

Note that for the competing talker, speech was always presented at a very low level: an SNR of -20 dB corresponds to a speech level of 45 dB SPL. This is however well above the SRT in quiet for the Flemish Matrix sentence test (18.7 dB SPL). Moreover, if we would have chosen to adapt the noise level instead of the speech level in our procedures, the noise would have become uncomfortably loud.

B   Spatial effects

Head shadow

We measured a head shadow effect of 4.5 dB, which is smaller than the benefit of 8.2 dB reported by Bronkhorst and Plomp (1988). This can be understood by comparing our ILD curves (Figure 2) with the normalized response curves reported by Bronkhorst and Plomp (1988) (Figure 2 in their paper): spatial separation towards 90° yields a smaller SNR increase in our set-up than in theirs. This might be because our localization arc was not located in a perfectly anechoic room.

Although both the spectrum and the temporal envelope of the modulated SWN and the competing talker were matched, we found head shadow to be smaller for the competing talker. However, we should take into account that (1) head shadow results in a frequency-dependent SNR-change, and (2) the importance of different frequency bands in speech understanding depends on target audibility and also masker type (ANSI, 1997;

Ma et al., 2009). The latter can be understood from the fact that although the spectra of both maskers are matched, the variation of the spectra over time is of course not matched.

Redundancy

We measured a redundancy benefit of 0.9 dB in SWN, which corresponds well with the value of 1.3 dB reported by Bronkhorst and Plomp (1988).

We measured a significant effect of noise type on redundancy. This has also been reported by Plomp (1976), and the data presented by Hawley et al. (2004) might also suggest this effect (Figure 1 in Hawley et al., 2004).

One can consider any **binaural benefit** (the benefit of listening with two ears instead of one, i.e., redundancy and squelch) as a result of a linear combination (conducted in the auditory periphery) of left and right ear signals to obtain an overall SNR improvement, as depicted in Figure 5: the right ear signal is scaled with a factor A and delayed by T, after which it is added to the left ear signal (we chose this asymmetric approach for simplicity of the model). This model is similar to the speech intelligibility model of Beutelmann et al. (2010), who implemented the equalization-cancellation model of Durlach (1963) as a binaural stage. However, contrary to their model in which only subtraction of left and right ear signals is allowed, we allow any linear combination. Estimates for the optimal scaling factors A and delays T for redundancy and squelch are given in Table 5.

Within this rudimentary model, the redundancy benefit (ILD = 0 dB, ITD = 0) can be modeled as internal noise cancellation by simple addition ($A = 1$, $T = 0$), i.e., binaural summation. If the signals (and external noises) in the two ears are in phase, they gain 6 dB after summation; if the internal noise in both ears is independent, it will only gain 3 dB after summation. Overall, this means a maximum of 3 dB noise cancellation.

The resulting net SNR improvement depends on the relative contribution of the internal noise to the total noise (i.e., the sum of internal and external noise). E.g., the internal noise might be higher in low sound levels (as was the case for target speech when the masker was a competing talker), resulting in a higher redundancy benefit. Or the internal noise cancellation might be frequency-dependent, also resulting in an effect of noise type. Note that the effect of noise type on redundancy is opposite to its effect on head shadow. Although these are two completely different effects, this might yield more insight in the frequency-dependent behavior of redundancy: while head shadow is an

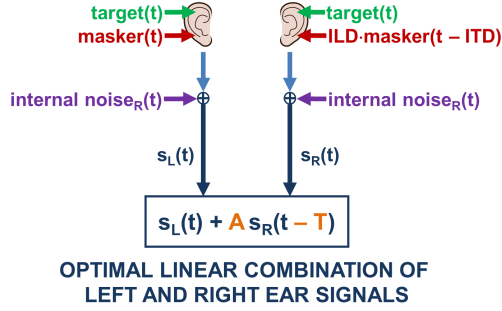SNR improvement mostly for high frequencies, it might be the opposite for redundancy.

Figure 5: This rudimentary model of binaural processing might explain the intelligibility benefits in redundancy and squelch. Instead of the equalization-cancellation model of Durlach (1963) in which only subtraction is allowed, we propose to allow any linear combination of left and right ear signals. The optimal delay T and scaling factor A can be frequency-dependent. Note that the ILD depicted in the figure is on a linear scale instead of logarithmic scale, and that we chose this asymmetric presentation of left and right ear signals for simplicity of the model.

Table 5: Estimates for the optimal scaling factors A and delays T for redundancy and squelch, according to our rudimentary model of binaural processing of Figure 5.

	A	T
redundancy	1	0
squelch$_{\text{all}}$	$\left[-\frac{1}{\text{ILD}}, 1\right]$	-ITD
squelch$_{\text{ILD}}$	$[0, 1]$ (or $-\frac{1}{\text{ILD}}$)	0
squelch$_{\text{ITD}}$	$-\frac{1}{\text{ILD}} = -1$	-ITD

Another reason for the effect of noise type might be that it is easier to pay attention to the target speech with two ears than with one ear, and that this attentional benefit is higher for informational maskers or for low sound levels.

We believe this is the first time that the effect of masker type and/or sound level on redundancy has been discussed.

Squelch and binaural contrast

When discussing squelch, we are particularly interested in which information yields the benefit in speech intelligibility: redundant information, spatial information (interaural differences) or a combination of both? Only when there is positive binaural contrast (i.e., when squelch is larger than redundancy), we are sure that the spatial information contributes to the binaural benefit.

With only ITDs (ILD = 0 dB, ITD $\neq$ 0), we can be sure that spatial cues did indeed add an extra benefit, as squelch$_{\mathrm{ITD}}$ was larger than redundancy for all masker types (i.e., significant binaural contrast). Moreover, in squelch$_{\mathrm{ITD}}$, we observed no effect of noise type, whereas we did in redundancy. We consider the binaural benefit again as an optimized linear combination of left and right ear signals (see Figure 5): the optimum is now to delay and subtract one signal w.r.t. the other to cancel out the external noise ($A = -1$, $T = -\mathrm{ITD}$), similar to the equalization-cancellation model of Durlach (1963)[3]. The amount of external noise cancellation is only dependent on how well the external noises in both ears are in phase, such that this benefit would indeed not depend on masker type. In SWN, we measured benefit of 4.2 dB, which is smaller than the benefit of 6.0 dB reported by Bronkhorst and Plomp (1988); this can again be explained by different room characteristics of our set-up.

With only ILDs (ILD $\neq$ 0 dB, ITD = 0), there was negative binaural contrast: the spatial cues did not add an extra benefit, as squelch was now smaller than redundancy for all masker types (with our result in SWN being in good correspondence with Bronkhorst and Plomp (1988)). Negative binaural contrast means that the binaural contrast counteracts a redundancy benefit; in other words, the spatial cues (ILDs) deteriorated speech perception. In squelch$_{\mathrm{ILD}}$, we observed the same effect of noise type as in redundancy, which suggests that both redundancy and squelch were based on a similar mechanism in this case. We consider again the addition of left and right ear signals without delay ($T = 0$, see Figure 5). However, left and right ear signals now have to be scaled by a

---

[3]Considering this mechanism, what we call binaural contrast could also be called binaural subtraction, similar to binaural summation. As for binaural redundancy, we prefer the term contrast over subtraction, as it emphasizes the cue that is used, rather than speculating about the mechanism behind it. Moreover, while squelch is definitely a result of both redundant information and interaural differences (as reflected by squelch = redundancy + contrast), we would not suggest that it is a summation and subtraction at the same time.

factor $A \neq 1$, to find the optimal trade-off between internal noise cancellation (maximal if worse ear signal is scaled with a factor $A = 1$) and external noise amplification (minimal if worse ear signal is scaled with a factor $A = 0$). As internal noise cancellation always trades off with external noise amplification, it does indeed make sense that squelch$_{\text{ILD}}$ is similar to but smaller than redundancy. Note that subtraction of left and right ear signals might also be beneficial to cancel external noise with only ILDs (by having $A = -\frac{1}{\text{ILD}}$), which might be the underlying mechanism of binaural masking level differences (Dillon, 2001); however, because head-induced ILDs are relatively small in the frequency region that is important for speech understanding (see Figure 2), subtraction also results in the cancellation of an important part of target speech.

With all spatial cues (ILD $\neq 0$ dB, ITD $\neq 0$), we measured significant positive binaural contrast for all masker types (with our result in SWN again in good correspondence with Bronkhorst and Plomp (1988)). Moreover, we measured the same effect of noise type on squelch$_{\text{all}}$ as in redundancy. If we consider again the binaural benefit as an optimized linear combination of left and right ear signals (see Figure 5), the signals have to be scaled to compensate for the ILD of the noise ($A \neq 1$, $T = -\text{ITD}$). The optimal scaling is again a trade-off between minimization of internal and external noise, resulting in the same effect of noise type as in squelch$_{\text{ILD}}$ and redundancy (note that the optimal scaling can be either positive or negative in this case).

We have to make some final remarks considering our interpretation of squelch, if attentional benefit might also play a role. Firstly, the extra spatial information that is obtained by listening with two ears instead of one, might contribute less in the conditions with low speech levels, since loudness is already an important segregation cue. Secondly, squelch$_{\text{ILD}}$ might have been smaller than squelch$_{\text{ITD}}$ and squelch$_{\text{all}}$, because the external noise was not perceived as far from the mid-line when only ILDs were presented. Therefore, we cannot conclude whether squelch was mostly mediated by top-down or bottom-up processes in our experiments; both theories can be supported by reasonable arguments.

Spatial release from masking

As expected, SRM was largest when all spatial cues were presented (Edmonds and Culling, 2005; Bronkhorst and Plomp, 1988). Moreover, we found smaller SRM for an

informational masker than for a (modulated) energetic masker. The latter is not agreed upon in literature: e.g., Edmonds and Culling (2005) reported the same effect, while Noble and Perrett (2002) reported the opposite. Note however that the spatially separated condition varies across papers when discussing SRM.

Irrespective of interpretation of masker type or spatial cues, it is important to note that SRM is uniquely defined by head shadow, squelch and redundancy (or – equivalently – by head shadow and binaural contrast) according to our definitions. This means that either of these effects might improve or deteriorate SRM.

For example, our smaller SRM with all spatial cues in SWN of 6.2 dB compared to the benefit of 9.4 dB reported by Bronkhorst and Plomp (1988) can be attributed to a smaller head shadow effect in our case. As another example, in the case with only ITDs, we observe that SRM becomes smaller because of a large redundancy effect. We emphasize this case, since redundancy is an effect that to our knowledge is never taken into account explicitly when considering SRM.

C General discussion

Our first objective was to investigate the contribution of ILDs and ITDs to the different spatial effects involved in speech understanding. As explained in the introduction, the contribution of ILDs and ITDs to head shadow and redundancy is trivial, since head shadow is only dependent on level differences and redundancy is defined without interaural differences. The reason why SRM remains such a puzzling effect, can therefore be attributed mostly to the complex nature of squelch. We found squelch irrespective of the spatial cues that were presented; the magnitude of the benefit was however dependent on the spatial cues. On the one hand, both in the cases with or without ITDs, the presentation of ILDs always resulted in a reduction of binaural benefit: $\text{squelch}_{\text{ILD}}$ was smaller than redundancy, and $\text{squelch}_{\text{all}}$ was smaller than $\text{squelch}_{\text{ITD}}$, irrespective of noise type. On the other hand, both in the cases with or without ILDs, the presentation of ITDs always resulted in an improvement of binaural benefit: $\text{squelch}_{\text{ITD}}$ was larger than redundancy, and $\text{squelch}_{\text{all}}$ was larger than $\text{squelch}_{\text{ILD}}$, again irrespective of noise type.

Thus, we were not able to measure the use of ILDs in speech understanding with realistic stimuli, which is important considering the signal processing of auditory prostheses: it suggests that it makes more sense to improve monaural SNRs, rather than

trying to preserve ILDs, to optimize speech intelligibility. However, in cases where ITDs are not or poorly perceptible, it does make sense to preserve or enhance ILDs for localization of sounds. Improved localization might also result in a release from informational masking due to failure of object selection (Shinn-Cunningham et al., 2005; Gallun et al., 2005) – which is an issue for e.g. bilateral cochlear implant (CI) listeners who cannot use temporal fine structure to segregate speech sources – and audio-visual speech intelligibility benefits (van Hoesel, 2015). It might be interesting to use our framework in specific hearing-impaired populations to quantify the trade-off between negative binaural contrast, spatial release from informational masking due to failure of object selection and audio-visual speech intelligibility benefits; this can help in the decision where the focus should be in new signal processing strategies.

We want to emphasize that the discussion about squelch is often implicitly a semantic discussion: if one measures a benefit of listening with two ears as compared to listening with the better ear only, that benefit cannot automatically be contributed to the spatial cues that are supplied. In our case with only ILDs, the squelch benefit appeared to be dominated by a redundancy benefit, while it was actually counteracted by the spatial cues (ILDs). In other words, squelch does not imply the utility of spatial cues; only binaural contrast – defined as the difference between squelch and redundancy – is able to imply the utility of spatial cues.

Our second objective was to investigate how these spatial effects are influenced by the type of masking noise. We found that the noise type had an influence on each spatial effect, even if spectrum and/or temporal characteristics were matched; even head shadow did not appear to be uniquely characterized by the spectrum of the noise. Although spatial effects are often attributed to attentional mechanisms (top-down processes), we were also able to support our results with a rudimentary model of binaural processing (bottom-up processes, see Figure 5). It might be interesting to extend existing speech intelligibility models (e.g., Beutelmann et al., 2010) with this adapted binaural stage.

To conclude, SRM is an often-used measure to summarize all binaural mechanisms in speech understanding. However, the measurement of a reduced SRM is hard to interpret, since it does not inquire which spatial mechanism is affected. The measurement of small or absent SRM is sometimes attributed to not being able to process spatial information (Ihlefeld and Litovsky, 2012). However, one should be aware of the fact that head shadow

is a purely physical effect, and this SNR improvement (even a small gain in low frequency regions) should always be beneficial. Instead of trying to explain reduced SRM by means of complex attentional mechanisms, it is also possible that the benefit of head shadow is canceled out by the reduced benefit of redundancy when speech and noise are spatially separated. We did indeed find that large redundancy can result in reduced SRM. Therefore, when in doubt about the utility of spatial cues in speech understanding for a certain population, it is advised to investigate monaural conditions as well. While redundancy is mostly relatively small for normal-hearing listeners, it does play an important role in e.g. bimodal cochlear implant users, for whom both ears deliver complementary information resulting in large binaural benefits.

D   Conclusions

We established a theoretical framework to unambiguously define and relate spatial effects in speech understanding. Then, we applied the framework to characterize speech understanding of 10 normal-hearing listeners, tested in different spatial set-ups and with different masker types. This was able to give us new insights in binaural processing in speech intelligibility:

1. Head-induced ITDs always improve binaural benefit (benefit of listening with two ears as compared to listening with one ear), while head-induced ILDs always disturb binaural benefit, irrespective of noise type. Therefore, to optimize speech intelligibility in hearing aids, it is more beneficial to improve monaural signal-to-noise ratios rather than to preserve interaural level differences.

2. Although redundancy is mostly ignored when considering spatial hearing, it might be able to explain reduced SRM in some hearing-impaired populations.

V   Acknowledgements

Author contributions: Benjamin Dieudonné and Tom Francart designed experiments, analyzed data and wrote the paper. Benjamin Dieudonné performed experiments.

References

ANSI, A. (1997). S3. 5-1997, methods for the calculation of the speech intelligibility index. *New York: American National Standards Institute*, 19:90–119.

Arbogast, T. L., Mason, C. R., and Kidd Jr, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 117(4):2169–2180.

Beutelmann, R., Brand, T., and Kollmeier, B. (2010). Revision, extension, and evaluation of a binaural speech intelligibility model. *The Journal of the Acoustical Society of America*, 127(4):2479–2497.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Blauert, J. (2013). *The technology of binaural listening*. Springer.

Boersma, P. P. G. et al. (2002). Praat, a system for doing phonetics by computer. *Glot international*, 5.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, 107(2):1065–1066.

Brand, T. and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, 111(6):2801–2810.

Bronkhorst, A. W. and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83(4):1508–1516.

Colburn, H. S. and Durlach, N. I. (1978). Models of binaural interaction. *Handbook of perception*, 4:467–518.

Dillon, H. (2001). *Hearing aids*, volume 362. Boomerang press Sydney.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35(8):1206–1218.

Edmonds, B. A. and Culling, J. F. (2005). The role of head-related time and level cues in the unmasking of speech in noise and competing speech. *Acta Acustica united with Acustica*, 91(3):546–553.

Festen, J. M. and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88(4):1725–1736.

Francart, T., Van Wieringen, A., and Wouters, J. (2008). Apex 3: a multi-purpose test platform for auditory psychophysical experiments. *Journal of Neuroscience Methods*, 172(2):283–293.

Francart, T., Van Wieringen, A., and Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International journal of audiology*, 50(1):2–13.

Gallun, F. J., Mason, C. R., and Kidd, G. (2005). Binaural release from informational masking in a speech identification task. *The Journal of the Acoustical Society of America*, 118(3 Pt 1):1614–1625.

Glyde, H., Buchholz, J. M., Dillon, H., Cameron, S., and Hickson, L. (2013). The importance of interaural time differences and level differences in spatial release from masking. *The Journal of the Acoustical Society of America*, 134(2):EL147–EL152.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America*, 115(2):833–843.

Ihlefeld, A. and Litovsky, R. Y. (2012). Interaural level differences do not suffice for restoring spatial release from masking in simulated cochlear implant listening. *PLoS One*, 7(9):e45296.

Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America*, 117(5):3091–3099.

Luts, H., Jansen, S., Dreschler, W., and Wouters, J. (2015). Development and normative data for the flemish/dutch matrix test. Technical report.

Ma, J., Hu, Y., and Loizou, P. C. (2009). Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *The Journal of the Acoustical Society of America*, 125(5):3387–3405.

Noble, W. and Perrett, S. (2002). Hearing speech against spatially separate competing speech versus competing noise. *Perception & psychophysics*, 64(8):1325–1336.

Peissig, J. and Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of the Acoustical Society of America*, 101(3):1660–1670.

Plomp, R. (1976). Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acta Acustica united with Acustica*, 34(4):200–211.

Shinn-Cunningham, B. G., Ihlefeld, A., Larson, E., et al. (2005). Bottom-up and top-down influences on spatial unmasking. *Acta Acustica united with Acustica*, 91(6):967–979.

Van Deun, L., Van Wieringen, A., and Wouters, J. (2010). Spatial speech perception benefits in young children with normal hearing and cochlear implants. *Ear and hearing*, 31(5):702–713.

van Hoesel, R. J. (2015). Audio-visual speech intelligibility benefits with bilateral cochlear implants when talker location varies. *Journal of the Association for Research in Otolaryngology*, 16(2):309–315.

Veugen, L. C., Chalupper, J., Snik, A. F., van Opstal, A. J., and Mens, L. H. (2016). Matching automatic gain control across devices in bimodal cochlear implant users. *Ear and hearing*, 37(3):260–270.

Wightman, F. L. and Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–1661.