A Binaural Cochlear Implant Coding Strategy inspired by the Medial Olivocochlear Reflex:
Intelligibility and Listening Effort with Moving Sources

University of Groningen
Major Thesis as part of the BCN Research Master Program
Henning Schulte

Supervisors: Enrique A. Lopez-Poveda (University of Salamanca), Anastasios Sarampalis (University of Groningen)

Acknowledgements: This work was conducted at the Neuroscience Institute of Castilla y León, University of Salamanca (Spain). I thank Almudena Eustaquio-Martín for help with the technical aspects of the work, and María Milagros Jerónimo-Fumero for help with data collection. Work funded by MED-EL GmbH (grant to Enrique A. Lopez-Poveda).
Abstract

It has been demonstrated that advanced binaural signal processing strategies inspired by the medial olivocochlear reflex have the potential to improve speech reception for cochlear implant users [Lopez-Poveda, Eustaquito-Martín, Stohl, Wolford, Schatzer, and Wilson (2016). Ear Hear. 37, e138–e148]. Here, we evaluate the performance of several alternative implementations of the novel processing strategies for moving sound stimuli using word triplets in a vocoder simulation. In addition, we look at listening effort using the single-task oral response time. We found that, in the tested scenarios, performance with the advanced processing strategies was similar as for a ‘standard’ processing strategy involving two independently functioning cochlear implants. Indeed, performance was slightly worse for one of the novel strategies than for the standard strategy. Further tests are needed to explain these results. Response times were reliably affected by the signal-to-noise ratio and the inter-stimulus interval (rapidity of presentation), but were approximately equal for all processing strategies. Assuming that response times reflect listening effort, this suggests that listening with the more advanced strategies requires similar effort as listening with the standard strategies.

Keywords: Bilateral CI, medial olivocochlear efferent, speech in noise, listening effort, response time
Abbreviations and acronyms

AIC: Akaike Information Criterion.
CI: Cochlear Implant.
FSF: Front-Side-Front.
GLME: Generalized Linear Mixed-Effects model
HL: Hearing Level.
ISI: Inter-Stimulus Interval.
LME: Linear Mixed-Effects model.
ML: Maximum Likelihood.
MOC1: Medial Olivo-Cochlear processing strategy #1.
MOC3: Medial Olivo-Cochlear processing strategy #3.
MOCR: Medial Olivo-Cochlear Reflex.
NH: Normal Hearing.
REML: Restricted Maximum Likelihood.
RT: Response (or Reaction) Time (oral).
SFS: Side-Front-Side.
SNR: Signal-to-Noise Ratio.
SPL: Sound Pressure Level.
SRM: Spatial Release from Masking.
SRT: Speech Reception Threshold (for word triplets).
SSN: Speech-Shaped Noise.
STD: ‘Standard’ processing strategy.
STOI: Short-Time Objective Intelligibility.
Introduction

Cochlear implants (CIs) are auditory prostheses employed to partially restore hearing for people with severe-to-profound sensorineural hearing loss. CIs have become a routine choice to treat profound sensorineural hearing loss and are regarded as the most advanced and most successful neuro-prosthesis to date. Nowadays, children are safely implanted as early as 6 months (Miyamoto, Colson, Henning & Pisoni, 2018) and the sinking costs of devices, the surgical routine and the consistent development make CIs increasingly available for more and more people, including the elderly and people with residual hearing (Brown, Hullar, Cadieux & Chole, 2011). Nonetheless, the hearing experience that users gain from cochlear implantation remains limited.

One main shortcoming of CIs is that they provide users with little spectral information (Wang, Zhou & Xu, 2011, Gfeller et al., 2007). The pattern of electrical stimulation delivered by CIs to the auditory nerve is only an approximation of the sound’s waveform envelope and disregards most of the temporal fine structure. The spatial matching of the tonotopical representation of the nerve fibers is very coarse and imprecise (Caldwell, Jiam & Limb; 2017). It is the spectral information that is perceived as pitch, which is one of the key variables in music and contains information about the speaker and the emotional content of speech (Looi, McDermott, McKay & Hickson, 2008). Above that, the spectral information carries implicit social cues, such as the speaker’s perceived sexual orientation (Munson, Jefferson & McDonald, 2006). In contrast, CI users who have no residual hearing might struggle to guess the sex (Winn, Rhone, Chatterjee & Idsardi, 2013) of a speaker or to recognize a simple melody (Looi et al, 2008). CI users often learn to adapt and rely on visual cues more than normal hearing (NH) people do to identify e.g. the speaker’s sex or their sexual orientation.
(Winn et al., 2013). Unfortunately, though, such cues are often not redundant across sensory modalities and cannot be accessed by CI users.

Arguably, a most limiting shortcoming of CIs in daily life is the difficulty for CI users to understand human speech in background noise. While it is known that it can already be difficult for CI users to understand and speak tonal languages in silence (Wang, Zhou & Xu, 2011), intelligibility is mainly affected by background noise in all languages.

While this shortcoming certainly depends on the sheer amount of spectral information that is available to the CI user (Friesen, Shannon, Başkent & Wang, 2001), there are several other factors that contribute to this difficulty. One factor is the saturation of neural responses. The dynamic range over which the auditory nerve fibers encode the stimulus is restricted when the speech signal is embedded in a noisy background. With the nerve fibers close to saturation, the CI user is less capable of retrieving the speech content from the pattern of electrical stimulation. Increasing electric spread across nerve fibers for higher stimulation levels increases the effect.

Another factor is poor spatial release from masking (Loizou et al., 2009, compare e.g. Adiloğlu et al., 2015). Spatial release from masking refers to an increased perception of a target sound with increasing the spatial separation between the target and the masking sound source, relative to the position of the listener.

One additional shortcoming relates to the way CIs are implemented. Cochlear implants bypass the mechanical parts of the human hearing apparatus and electrically stimulate the auditory nerve. Unfortunately, the substituted parts of the early auditory system do not restrict themselves to a bottom-up organization. There are several efferent effects that regulate auditory function already at a mechanical level, one of the most
important and best studied of which is the medial olivocochlear reflex (MOCR; Guinan, 2018; Lopez-Poveda, 2018).

**The medial olivocochlear reflex**

The MOCR is thought to be involved in speech-in-noise perception in healthy humans (Kirk & Smith, 2003; Mishra & Lutman, 2014) by modulating the responsiveness of the basilar membrane to sound. It has been assumed that the MOCR may be involved in spatial release from masking (SRM; Kim, Frisina & Frisina, 2006).

Cooper and Guinan (2006) observed the functioning of the MOCR in vivo, supporting earlier hypotheses that the reflex acts upon the motion of the basilar membrane via frequency specific modulation of the so-called cochlear amplifier as facilitated by the disengagement of the outer hair cells (Dallos et al, 1997). MOC efferent activation linearizes the compression curve of the basilar membrane, meaning that the transmission of low- and medium-level sounds are suppressed, while high-level sounds are less affected (see Figure1). The basilar membrane, which is usually able to respond to a wide range of sound levels thanks to the cochlear amplifier, responds less to sounds when the MOCR is active (particularly to low- and mid-level sounds). The effect can be thought of as a frequency-specific mild temporal hearing loss, introduced by the brain in adaptive response to the soundscape in order to avoid saturation of the auditory system, especially on the level of the nerve fibers. While the observation of Cooper and Guinan was done in guinea-pigs, it is assumed that the MOCR works similarly in humans.
Figure 1: Figure 2A from Cooper and Guinan (2006). Blue lines show amplitude growth functions of the basilar membranes response to sound stimulation at the characteristic frequency (CF=18 kHz, mid panel), as well as below (15 kHz, top panel) and above (20 kHz, bottom panel). When the efferent MOC fibres are stimulated electrically, the slope steepens (green line), and low-level input is suppressed, especially for tones at the CF and above.

Froud et al. (2015) found that the MOC innervation is controlled by spiral ganglion cells. It can therefore be assumed to depend on the early neural representation of the contralateral signal rather than the signal itself. While there are some indications that the MOCR might be modulated by attentional processes (Garinis, Glattke & Cone, 2011), it is a subconscious and relatively fast process. Backus and Guinan (2006) found the onset delay of the reflex to be about 25ms and the exponentially build up and decay to take 277±62ms and 159±54ms, respectively. Importantly, the buildup function can be described by a combination of fast, medium and long time constant components (70ms, 330ms, 2500ms).

When speaking of the MOC system, there are ipsilateral and the contralateral effects. While the ipsilateral pathway contains more MOC efferent fibres than the contralateral MOC pathway (in cats; Liberman & Brown, 1986), there is to date conflicting evidence about the relative strength of their effects (Lopez-Poveda, 2018). While the strength of the benefits of the MOCR on speech-in-noise intelligibility is still under debate today (Wersinger & Fuchs, 2011; Lopez-Poveda, 2018), Lopez-Poveda
and colleagues have translated parts of its features into a novel bilateral CI sound processing strategy design that shows great potential for improving the perception of speech in noise by CI users. The strategy in question is referred to as “the MOC strategy”.

The MOC processing strategy

The MOC strategy is designed to mimic the effect of the contralateral MOCR on basilar membrane compression and thus to equip CI users with the effects and (hopefully) the benefits of the MOCR. The MOC strategy can improve the recognition of speech in steady state noise for bilateral CI users and increase spatial release from masking (Lopez-Poveda et al., 2016). While an extensive description of the MOC strategy can be found in the literature, we will briefly outline its main features and functioning here.

Figure 2: Flowchart showing the MOC strategy. Adapted from Lopez-Poveda et al. (2016)
The MOC strategy can be understood as an interactive coupling between two similar basic standard coding strategies (see Figure 2), resembling those often used in commercially available devices (hereafter referred to as standard strategy or STD). The findings regarding the performance of the MOC strategy are therefore to be interpreted in comparison to the STD strategy.

Both strategies are based on Wilson’s method of continuous interleaved sampling of compressed envelopes (Wilson et al., 1991) and contain a 3dB high pass pre-emphasis filter at 1.2 kHz (1st-order Butterworth), followed by a 3dB band-pass filter-bank (number of filter bands equal to the number of electrodes), its cut-offs logarithmically distributed between 100 and 8500 Hz (6th-order Butterworth). The envelopes are estimated via full-wave rectification and 3dB low-pass filters (4th-order Butterworth) and finally compressed logarithmically across all filter-bands, according to

$$y = \frac{\ln(1+c \cdot x)}{\ln(1+c)}$$

where $x$ and $y$ are the instantaneous input and output amplitudes, respectively (with minimum and maximum values within 0 to 1) and $c$ denotes the compression coefficient. For the STD strategy, $c$ is fixed at 1000 while for the MOC strategy, the compression for each spectral channel dynamically depends on the time-weighted output energy from the equivalent filter band in the contralateral device, as estimated over a sampling window. For no contralateral output, there is no change in (ipsilateral) compression ($c = 1000$; equivalent to STD), while for nonzero outputs the equation becomes more linear, which causes inhibition in the corresponding frequency channel. This type of dynamic on-frequency compression fitting is intended to prevent saturation in the respective spectral cochlear regions and thus to aid the intelligibility of complex stimuli over rivalling stimuli.
To assess whether there would be benefits of the MOC coding strategy also in fluctuating maskers, Lopez-Poveda et al. (2017) tested the processing strategy with speech maskers. They again found that the MOC coding strategy outperformed the use of two independent devices. They reported similar benefits for the fluctuating masker in bilateral CI users as they had found for the steady-state masker.

In order to further improve the proposed coding scheme, Lopez-Poveda & Eustaquio-Martín (2018) designed several modified versions of the MOC strategy and compared them using an objective index of speech quality: the short-time objective intelligibility (STOI; Taal, Hendriks, Heusdens & Jensen, 2011). Their main goal was to assess whether there was a benefit in setting the processing strategy parameters closer to those of the natural MOC reflex. Four different MOC strategies were tested against a STD processing strategy. Note that all implementations are equally well depicted by Figure 2. There are solely differences in how the parameters are set.

The authors found that there was indeed an increased benefit for the implementation that modelled the MOCR the closest. The best strategy (referred to as MOC3) differed in two ways from the previously tested strategy (MOC1):

1. The second time constant of the sampling function was set to 300ms, (instead of 2ms), in accordance with the medium time constant length of the MOCR in humans. Indeed, this delay is closer to the natural functioning of the MOCR (Backus and Guinan, 2006).

2. Contralateral inhibition was stronger (greater linearization of the compression function) for lower frequency channels as compared to higher frequency ones (see Aguilar, Johannesen & Lopez-Poveda 2015).
It was found that the MOC3 (longer time-constants) strategy generally outperformed the MOC1 (shorter time-constants) strategy. The authors argued that the stronger fluctuation of the signal sampled at shorter time constants (2ms) probably caused distortions to the envelope in the single contralateral frequency channels. The longer time constants (300ms) cause less fluctuation in the contralateral compression rate and therefore elicit a “smoother” effect. The authors additionally compared the STOI for different target and masker sound locations across processing strategies to experimental data from NH and bilateral CI listeners. They observed that the SRM benefit of the STD processing strategy closely followed the data for bilateral CI users, while the MOC3 strategy scored similar to the NH sample.

**This investigation**

As described, the MOC strategy is expected to improve intelligibility for sounds sources over all azimuth angles and a wide range of noise levels (e.g., Lopez-Poveda & Eustaquio-Martín, 2018). However, the evaluations of the MOC strategy to date have focused on combinations of *fixed* spatial arrangements of the target and interferer sound sources; i.e., on using spatially static sound sources. In natural listening, however, sound sources (or the speaker) are seldom static. To better predict the benefits of the MOC strategies in realistic situations there is an additional dimension that needs to be incorporated: movement. The present study is a first exploratory attempt to evaluate the MOC strategy with moving sound sources.

The MOC coding strategy differentiates itself from STD processing strategies in that it samples the output energy from the contralateral device over time to regulate the compression parameters in the ipsilateral device: the higher the output energy, the less the compression. This implies an interaction between timing and inter-aural level difference (or spatial location), which may affect the functioning of the MOC strategy.
and its potential benefits. For quickly moving sound sources, the interaction between
the devices is expected to be complex and demands investigation. Here, we tested two
different implementations of the MOC strategy (MOC1, MOC3; described above)
against the STD processing strategy for moving target speech sources.

For two independently functioning CIs, the relative energy output across
frequency bands mainly depends on the spectral energy distribution in the soundscape,
the diffraction effects of the head shadow and the device´s specifications. With the
MOC strategy, it can additionally depend on the energy output from the corresponding
frequency band of the contralateral device. As this energy output is estimated over a
time window by an exponentially decaying sampling function, changes in the
contralateral stimulus can influence the ipsilateral compression functions at a temporal
delay. For the MOC1 strategy, this delay is short (2 ms) and might be negligible, but for
the MOC3 strategy the delay is substantial (300 ms).

NH listeners seem to deal very well with quickly moving sound sources; we are
not aware of any findings that show the contrary. In addition, the motion of the stimulus
sound at moderate speed can even improve intelligibility in NH listeners (Allen, Charlie
& Alais, 2008; Davis, Grantham & Giffort, 2016). Likewise, both CI users and NH
people seem to benefit from motion of the noise masker (Weissgerber, Rader &
Baumann, 2015). The findings of Davis and colleagues, indicate though, that the benefit
of motion does depend on a sustained continuity of that motion.

We decided to employ a vocoder simulation study to test the MOC strategies in
a perceptually very challenging setup designed to reveal the time delay effects of the
MOC strategies on the speech reception threshold. We chose to test the speech reception
threshold for word triplets that change position in saccadic motion across different
azimuth positions. The stimulus location changed for every word, from lateralized (60°)
to centralized (0°) and back (60°), or vice versa (0°-60°-0°). The noise masker was static at 30° azimuth. This implies that the stimulus crisscrosses the spatial location of the masker.

The listener was required to report all three words correctly. To differentiate between the processing strategies, we presented the three words at different inter-stimulus-intervals (ISI), the idea being that longer ISIs (300 ms) would allow the MOC3 strategy to fit its spectral channel compression to the background noise while for the shorter ISIs (2 ms) the fitting would still be that of the preceding word and the background noise. Additionally, an intermediate ISI (100 ms) was tested as well.

We expected to observe an interaction between the different inter-stimulus-intervals and the time course of activation and deactivation of contralateral inhibition due to the MOC strategy. In addition, we expected a main effect of ISI, with longer time gaps between the words aiding intelligibility. Due to the nonlinear nature of the MOC strategy, it was hard to make more specific predictions on the results.

Knowledge on the caveats or benefits of different strategies towards sound movements might prove valuable in finding the most effective parameter distributions for the MOC strategy as well as for an enhanced understanding of the importance of the MOCR in natural hearing.

**Listening effort**

We also looked into the response times to the stimulus. This part of the investigation was solely exploratory. We aimed at assessing the utility of the single-task oral response time as a measure for listening effort in the broad context of parameter fitting for CIs. We understand listening effort to be the “attention and cognitive resource required to understand speech” (McGarrigle et al., 2014). The above-described study
offered a chance to assess the usefulness and convenience of the single-task response time.

Listening to degraded speech signals puts a high strain on the available cognitive resources (Stenfelt & Rönnberg, 2009). When these signals happen to be embedded in high levels of noise, the available cognitive resources might be reduced even further. It might not only make listeners feel drained or stressed after longer auditory engagements, but also affect them in making use of perceived auditory information (compare Hicks & Tharpe, 2002 on the effect of mild-to moderate hearing loss on listening effort in schoolchildren, reflected in reaction time) and have consequences on the long run (Bess, Dodd-Murphy & Parker, 1998).

Already in 1992, Feuerstein noted that there seems to be a systematic difference between subjective and objective measures of listening effort. We disregarded self-report measures of effort as they are hardly correlated to psychophysical and physiological measures (e.g. McGarrigle et al., 2014).

Attempts to measure listening effort in an objective manner are mostly based on the notion of a single, limited cognitive resource (Kahneman, 1973). There are different approaches to implement Kahneman’s attention model and his theory of a limited cognitive capacity. One way is the dual-task response time measure, in which the effort spent on a first task is rated by the reaction times on a secondary task. It is assumed that there is a fixed amount of cognitive resources that can be distributed and has additive effects on reaction time. As the first task takes up more effort, this shows in longer reaction times in the secondary task. The additional task is needed to assure that the cognitive capacities are properly engaged, so that the variability on the second task reflects the variability in effort for the first task.
While there is not yet a final consensus on what is actually being measured (McGarrigle et al., 2014), reaction time (RT) measurements have been found potent in discriminating between more and less challenging listening situations (e.g. added noise; Huckvale & Leak, 2009; Sarampalis et al, 2009). This also applies to high SNR, when intelligibility is already at a ceiling level (Houben, van Doorn-Bierman & Dreschler, 2013).

The dual-task paradigm is relatively easy to implement and well tested but introduces practical restraints to the experiment design.

Pals, Sarampalis, van Rijn and Başkent (2015) have shown that the RTs on the first task can also be taken as a measure of difficulty, correspondingly to the RTs of the secondary task. They suggest to concentrate on evaluating the effort of the first task directly. Further, they argue that, especially in the medical context, a single task paradigm would be easier to administer and the results more straightforward to interpret.

Houben and colleagues (2013) have also shown that both the primary and secondary tasks (in this case an identification and an arithmetic task, respectively) reflect the effect of added noise on response time. While the RTs to the primary task were found to show a lower variability, the RTs to the secondary task showed a larger difference between SNR levels. All in all, the primary task was slightly more powerful in discriminating between higher SNR and the secondary task was more suited for lower SNR. Houben and colleagues advise that the choice of tasks should depend on the research question in mind and especially the SNR region of interest.

Van den Tillaart-Haverkate, de Ronde-Brons, Dreschler & Houben (2017) found that in a dual task paradigm, RTs to a secondary task provide an objective measure for
noise levels. Their first task, a simple identification task, was not sensible to the stimulus conditions. It can be assumed, that a simple identification task did not reflect the differences in the noise level, as the task itself was not requiring for the use of enough cognitive resources, itself. We consider this notion in our experimental setup.

Based on a pilot study, we had reason to believe that the chosen speech-in-noise task (as explained below) alone was challenging enough to raise systematic differences in RT between the SNR conditions. The subsequent, remaining questions were, how sensitive the single-task approach would be in differentiating between the chosen experimental conditions and in what way the conditions would differ.

**Methods**

The experiment consisted of a speech-in-noise intelligibility task for word triplets that were presented at an adaptive SNR corresponding to the 50% intelligibility level. Eighteen conditions were tested, following a 2x3x3 design. The conditions differ in the processing strategy (STD, MOC1, MOC3), in the temporal interval between the three words (inter-stimulus interval [ISI]; 2 ms, 100 ms, 300 ms) and in the movement scheme of the stimulus (front-side-front [FSF], side-front-side [SFS]). The study was approved by the Ethics Review Board of the University of Salamanca.

**Participants**

Nine (four male) native Spanish young adults participated in the experiment. Seven of them were tested thrice in three distinct sessions, while two of them were tested only once. All participants had normal pure-tone audiograms (< 20 dB HL) and reported no hearing difficulties. Participants signed an informed consent to participate in the study. They were volunteers and received no compensation for their service.
**Stimulus**

Participants were presented with recordings of words triplets in speech-shaped noise (SSN) that they are asked to reproduce correctly and promptly. For each word triplet, recordings of common Spanish disyllabic words were randomly chosen from (Cardenas & Marrero, 1994) by a custom make Matlab script (The Mathworks, Inc., R2014a). The word recordings and the noise masker were then processed via a head-related transfer function (HRTF) for each ear in order to gain a perceived presentation direction. For the spatial arrangement of the target words and the noise masker see Figure 3. The single words were concatenated at an inter-stimulus-interval (ISI) of 2, 100 or 300ms, according to condition. The ISI was always kept the same within conditions. The stimulus streams containing the word triplets were thereon combined with the noise masker signal. The noise masker started 500 ms earlier and ended 100 ms later than the stimulus. The noise was gated with 50-ms onset/offset raised-cosine ramps. The two sound streams were further processed by one of the three strategies (STD; MOC1; MOC3), using 12 frequency channels with cutoff frequencies at approximately logarithmic spacing between 100 and 8500Hz.

The compressed envelope structure of all channels were then recombined into a vocoded sound stream following the procedure described by Shannon, Zeng, Kamath, Wygonski & Ekelid (1995) and subsequently presented to the participants.
**Presentation**

The apparatus was controlled by custom Matlab software (the Mathworks, Inc., R2014a) running on a Windows operating system. The sound settings were controlled by a Fireface 400 mixer (RME/Audio AG). Stimuli were presented using insert earphones (ER-2, Etymotic Research, Inc.) at a fixed level of 65 SPL. Testing took place in a soundproof semi-anechoic booth. Each session took about 90 minutes and was split in four blocks, separated by a few minutes of rest. Presentation of the single conditions was counterbalanced across sessions and across subjects.

For each word triplet, the SNR was reduced for every correct response and increased for every incorrect response following an adaptive one-down-one-up staircase procedure. Starting at an SNR of 20 dB, the succeeding word triplet was presented a lower SNR for correct and at a higher SNR for incorrect responses. For the first 14 word triplets the step size was fixed at 4 dB SNR, then it narrowed to 2 dB SNR. In total, 25 items were presented, of which the SNR for the last 17 word triplets was averaged to determine the Speech Reception Threshold (SRT). Responses were recorded with a
simple clip-on microphone and presented to the experimenter outside the booth via headphones. A response was scored as ‘correct’ only when all three words were repeated perfectly correct and in the right order. A separate computer was simultaneously recording the stimulus and response streams using Adobe Audition (Adobe Systems, CC 2017). The RT was measured from the offset of the noise masker to the onset of the response. The word triplets are thus treated like a single stimulus and the beginning of the response is assumed to reflect the combined listening effort of all three words.

The choice of the experimental conditions bases itself partly on the research question, partly on experience and partly on a pilot study in five NH participants that was set up to test different parameters and validate the methodology.

**Analyses and Results SRT**

We measured SRTs for 18 different conditions, as described above. Separate statistical analyses were conducted for spatial arrangements FSF and SFS. The SRT data consist of a mean score over the last 17 word-triplets per session (414 observations). We performed a very conservative outlier rejection, removing SRTs above 16 dB SNR. This amounted to removing less than 1% of the data. We assume that a lack of attention caused these unusually high scores.

1. **Data inspection**

For convenience, we will first describe the results visually and then follow up with the statistical evaluation of the question whether there are differences between the MOC strategies and STD. We believe that this structure will help to make the data and the results more accessible. Moreover, it helps to remember that the figures (4-10) show
the uncorrected data (if not otherwise specified), while some of the statistical analyses work with transformed data and are therefore not tantamount to the visual interpretation.

We observe a large variance of SRT between and within subjects. We decided not to include error bars in most figures to avoid clutter. In addition, there are little consistent effects across subjects in Figure 4, although most subjects tend to underperform (i.e., show higher SRTs) with the MOC1 strategy.

![Figure 4: Mean SRT (in dB SNR) per subject across ISI for the different processing strategies. Left and right panels show results for the FSF and SFS moving directions, respectively. Each point is the mean of three SRT measurements per ISI per processing strategy, except for subjects six and seven (indicated by asterisks) who were tested only once per ISI per strategy.](image)

We were interested in testing for potential interactions between the MOC strategies and ISI in comparison to the STD strategy (Figure 5). Therefore, we subtracted the mean SRT values for the MOC strategies from the corresponding STD for each ISI. We observe a consistent underperformance of MOC1 (except for FSF at 2ms, where MOC1 is exactly identical to STD), while MOC3 seems to outperform STD in FSF, while it performs poorer in SFS.
Figure 5: The left panels show the mean SRT for the different ISI levels with the STD strategy (FSF top, SFS bottom). The middle and right panels show the deviation of the corresponding mean SRT for the MOC strategies. Herein, we subtracted the SRT_{MOC} from the SRT_{STD}, so that positive values denote a better (lower) SRT_{MOC} and negative values a better performance with the STD strategy. There is a large variance for all strategies. We omitted the error bars in the right panels to avoid clutter.

When looking at the single subjects (Figure 6), there is little consistency across participants for most conditions.

Figure 6: The difference between the mean SRT for STD and MOC strategies per subject. Positive values denote a better performance (lower SRT) with the corresponding MOC strategy. Again, note that subjects six and seven were only tested once per ISI and strategy.
From the size of the effects we can already anticipate that there were no significant interactions between processing strategy and ISI. The mean divergence of the MOC strategies from the STD strategy were small compared to the standard deviation of the STD strategy. Nonetheless, the deviations were relatively consistent, so there might be a main effect for strategy.

2. Statistical evaluation

To quantitatively assess the strength of this effect, we generated a linear mixed effects (LME) model design based on a maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013), including Subject, Session nested within Subject, SNR level and Number of correct responses as random effects. We then compared the full model to the next smaller option based on the Akaike information criterion (AIC) and chose the better model. It is important to note that the AIC penalizes for the number of parameters, weighting between the increase in model fit and the number of introduced parameters.

We used the maximum likelihood (ML) estimator function for comparison between two nested models that differ in their fixed factors and the restricted maximum likelihood (REML) for comparison between models that only differ in their random factors part. We use the REML to report the results for the identified best model (Zuur, Ieno, Walker, Saveliev & Smith, 2009). As it can be observed in the results, the interaction between the fixed effects were not included in the final model, due to lack of significance.

We performed several model diagnostics (Appendix A-C) and observed a violation of the normality assumption. As transformation of the data did not lead to improvements and we did not want to exclude larger amounts of data, we decided to
ignore the issue. After all, LMEs are generally regarded to be robust to violations of the normality assumptions.

2.1 Model

We started with a maximal random effects model including all potentially meaningful random effects as random slopes for both fixed effects, but found that the best fit (AIC) was reached with the same small random intercept model, for FSF as for SFS. In Wilkinson-Rogers notation (Wilkinson & Rogers, 1973), the model spells out as:

\[
SRT \sim 1 + STRATEGY + ISI + (1|SUBJECT) + (1|SUBJECT:SESSION)
\]

The formula declares that the data \(SRT\) was best described by an additive combination of the intersect, the fixed factors \(Strategy\) and \(ISI\) and a random intersect model of the factor \(Subject\) and \(Session\) nested within the factor \(Subject\). In the process of recursively choosing the better model (as described above), the interaction term between \(Strategy\) and \(ISI\) was removed. In turn, this means that there was no benefit of including e.g. the interaction term between the two fixed factors in the model. When nevertheless including the interaction term (out of curiosity), none of the interactions turned out to be statistically significant. If we had chosen a less concise model, the surplus factors would have influenced the outcomes (Barr, Levy, Scheepers & Tily, 2013) due to the way shared variabilities are allocated. Now, that the model is specified that explains the observed variability the best, we can calculate and interpret its results. All results are reported at \(\alpha=.05\).

2.2 Results

For the FSF data, we found a main effect for MOC1 \([p=.009]\) and MOC3 \([p=.030]\) with respect to the grand mean, but no main effect for the ISI levels. The grand
mean denotes the overall average of all factor levels (2 ms, 100 ms and 300 ms for STD, MOC1 and MOC3). Comparing against the grand mean (main effect) is a good first measure for difference between the factor levels, in that it does not require for the choice of a reference level. We chose to first look at the main effects in order to get an overview of all statistically relevant effects. In a second step, we tested on our research question, that is, whether the MOC strategies do perform differently from the STD strategy.

Using a Bonferroni corrected planned comparison (MOC1 vs. STD; MOC3 vs. STD), we did not find MOC1 and MOC3 to significantly differ from STD ($p=.152$ and $p=.641$, respectively). This indicates that there were indeed differences between the processing strategies (see main effects), but when directly comparing MOC1 or MOC3 against STD at a corrected $\alpha$, these differences were not statistically significant. Note that there are other, less conservative methods that could have been used to deal with the problem of multiple comparisons, that would potentially have led to different conclusions.

We continued by repeating the analysis (as described above) for the SFS data. Here, we found a significant main effect of processing strategy MOC1 ($p \leq .001$) and STD ($p \leq .001$) over the grand mean. In addition, we observed a main effect for the shortest ISI level (2 ms) [$p=.017$], over the grand mean. In the (corrected) planned comparison, only MOC1 differed significantly from STD ($p \leq .001$), while MOC3 did not ($p=.168$). In light of the performed experiment, this means that in the SFS arrangement, performance was worse with the MOC1 than with the STD strategy across all levels of ISI (compare: Figure 5), while it did not differ significantly for MOC3 and STD.
Analysis and Results RT

The RT data contain the actual response times to the first word for all word triplets (10350 observations). The RT was scored manually in Adobe Audition by visually assessing the difference between the offset of the noise masker and the onset of the response (see Appendix D). All scoring was done by the same rater. The responses were additionally evaluated acoustically to safely separate responses from other recorded sounds (coughing, rustling, expression of contemplation, etc.). Three randomly selected small subsets of the data were given to a second rater to assess the inter-rater variability. For all three tests, the correlation between rater and second rater was very high (Pearson’s r: .987, .989 and .985).

1. Outliers

We excluded RTs below 0 seconds and above 1.615 seconds (<5% of the data). The low cutoff was chosen because of technical restrictions. For higher SNRs, some participants managed to begin to respond within the 100 ms noise-only period succeeding the stimulus signal. These responses were unexpectedly early, but valid within the instructions the participants received. The high cutoff was specifically selected to exclude the same amount of data as the zero-cutoff to keep the median untouched. Response times longer than 1.5 - 2 seconds are often assumed to not reflect the effort connected to the task difficulty, but rather e.g. the attentional state of the participant (Ratcliff, 1993). We chose to remove only this small portion of what could potentially be assumed to be outliers, based on recommendations made by Whelan (2008).
Houben and colleagues (2013) observed that there is a systematic difference in RT length between correct and incorrect responses, the latter being longer. To make the analysis ecologically valid, we did not differentiate between correct and incorrect responses. We were explicitly interested in comparing the reaction times at the 50% SRT level, so that by design half of the responses are expected to be erroneous. Likewise, responses like “Nothing understood”, “Nothing” or “I don’t know” were counted as valid (although incorrect) responses. Only when the participant failed to respond or when the response time was outside the cutoff values, the corresponding data point was excluded.

2. Data inspection

We are first interested in corroborating that RT co-varies with SNR, which we interpret as an indication that it reflects listening effort (in the same way that comparable methods are interpreted to measure listening effort). Figure 7 shows that RT increased (non-linearly) with decreasing SNR. Importantly, the curvature of the decline was similar for each processing strategy, our main factor of interest. This indicates that apparently there was no interaction between SNR level and processing strategy, which could have complicated the statistical interpretation.

Next, we check the relation between RT and SNR level for every ISI (Figure 8). We immediately see a strong effect of ISI on RT, but this effect seems to be relatively stable over SNR levels. This is important to check not to misinterpret potential interactions between processing strategy and ISI.
Figure 7: The mean RT data for all three processing strategies sorted for SNR level. We do not include those SNR levels that hold fewer than 100 observations for each strategy, which sum up to about 5% of observations. The reason for the exclusion is the high variability of the RT scores and thus the low reliability of the average for small samples. Note that the narrowing of step-size during the adaptive-SNR method (four and two decibels, respectively) is mirrored in the relative number of observations for each SNR level.

Figure 8: Mean RTs sorted per SNR level, this time for the different ISI levels. The ISI seems to have a relatively stable effect on RT across SNR.
Subsequently, we compare the MOC strategies to the STD strategy for every ISI. The following graphs depict the simple mean RTs. Note, that as the RT data will probably be strongly right skewed (for statistical analyses, see below), the simple mean is strongly influenced by a small portion of the data (leverage of the tail) and can therefore be unreliable. Nonetheless, Figure 9 is a helpful depiction of the general direction and magnitude of the effects.

![Figure 9: The left panels depict the mean RT duration of STD for each of the three ISI levels. The right panels show the deviation of the MOC strategies’ means from that of STD. Herein, positive values signal an improvement (shorter RT) of MOC over STD. Negative values represent a shorter RT for STD.](image)

Next, we inspect the difference between the MOC strategies and STD per subject (Figure 10), to get an idea about size and regularity of the effects across our sample.
3. Statistical evaluation

As expected, RTs were right-skewed (towards higher values; Figure 11; compare Ratcliff, 1993; Whelan, 2008). With the normality assumption violated, we were left with the choice between transforming the RT data or choosing a statistical analysis that did not assume normality of the residuals.

Based on the recommendations of Bolker and colleagues (2009), as well as Lo and Andrews (2015), we decided to employ a generalized linear mixed-effects model (GLME), instead of transforming the data upfront and using a linear mixed-effects model. Repeated-measures analysis of the variance (RM-ANOVA) is somewhat robust to violations of the normality assumptions but was disregarded because it simply averages over random factors (Whelan, 2008; Speelman & McGann, 2013). Generalized
estimating equations models (GEE) were not an option due to the small sample size (Teerenstra et al., 2010). We employed a Gamma distribution (Figure 11) for the variance model and used an identity link function in order to meet the distribution of the data in the most conservative way (Lo & Andrews, 2015).

3.1 Evaluation of distribution and link function

We plotted the observed quantiles versus the predicted quantiles for the GLME of FSF data (Appendix E; Figure E1) to quantify the distribution model fit (Pearson’s $r = .973$). The fit for the next best distribution, an Inverse Gaussian distribution model was slightly worse (Pearson’s $r = .967$), which supported our model choice. Correspondingly, for SFS data, we plotted the observed quantiles versus the predicted quantiles for the GLME (see Appendix E, Figure E2) to demonstrate the distribution model fit (Pearson’s $r = .961$). Again, using the Inverse Gaussian distribution we obtained a slightly worse fit (Pearson’s $r = .956$). These fits were decent but not perfect. Unfortunately, as demonstrated by Baayen and Milin (2010), there is a tradeoff between stricter outlier rejection and a worse model fit.

3.2 Model choice

Once the distribution and link functions were specified, we progressed to the level of the single factors. Analogous to the analysis of the SRT data, we used maximum pseudo-likelihood estimations for model comparison (fixed factors) and reported the restricted maximum pseudo-likelihood estimations of the best model.

Following the guidelines of Barr and colleagues (2013), we again started with the maximal random effects model and employed a backward stepwise selection procedure, based on the Akaike information criterion (AIC). In the case that the model did fail to converge, we also reduced the random effects structure. For the selection of
the fixed-effects, we followed the guidelines by Zuur and colleagues (2009). As for the analysis of the SRT data, we included various random effects parameters that did not make it into the final model, like SNR and Number of words correct. We did not attempt to model the single words used in the word triplets though, due to their sheer number and the resulting degrees of freedom that would be required.

Figure 11: We modelled a gamma distribution to the two data sets and observed a reasonably good visual fit for both, FSF and SFS. The data are presented in 150 bins.

3.3 Results

The best model (for the notation, again see Wilkinson & Rogers, 1973) emerged to be:

\[ RT \sim 1 + ISI \times STRATEGY + (1|SUBJECT) + (1|SUBJECT:SESSION) \]

The RT data were thus best described by a model that included the intercept, ISI, Strategy and their interaction as fixed effects and the random intercept effects Subject as well as Session nested within Subject (Note that * is the crossing operator). The random factors SNR and Number of words correct (as an indicator of how little was perceived in case of an incorrect response) did not improve the model fit. Likewise, all random slopes structures produced a worse fit than the simpler random intercept structures, as based on the AIC. For a visual check of the statistical assumptions, see Appendix E-G.
Analogous to the analysis for the SRT data, we first compared all factor levels to the grand mean, then continued with the direct comparisons of the MOC strategies with STD.

For the FSF condition, we found MOC1 to show significantly longer RTs as compared to the grand mean ($p=.030$). We observed significant interaction between ISI = 2 ms and STD ($p<.001$) and MOC1 ($p=.040$). There was also an interaction between ISI = 300 ms and MOC3 ($p=.026$). We employed a Bonferroni-corrected planned comparison and found MOC1 to significantly differ from STD ($p=.044$), while MOC3 did NOT differ ($p=.875$) from STD. This indicates that MOC1 underperformed in comparison with STD and that this effect was especially strong for the ISI = 2 ms condition.

For the SFS condition, we found a significant effect of MOC3 above the grand mean ($p=.046$). There were no significant interactions. When directly comparing the MOC strategies to STD (Bonferroni-corrected), we found no significant differences in RT.

**Discussion**

We employed a vocoder simulation study to assess the effect of different processing strategies on word-triplet perception in noise. For this, we measured SRTs and RTs over 18 conditions. In the following, we will discuss the results regarding the differences between strategies, the methodology and potential shortcoming of the experiment.
Speech reception thresholds

As expected, we observed SRTs to vary largely between and within subjects and sessions. The SRT depends not only on the objective quality of the signal (i.e., the SNR in the delivered signal) and the abilities of the subjects to exploit them. Multiple other factors can have an influence on a time-variant basis, such as the attentional state (Koelewijn, Versfeld & Kramer, 2017), learning and fatigue effects (Borragán, Slama, Bartolomei & Peigneux, 2017) or the varying difficulty of combinations of words. From the pilot study we conducted to test the methodology, we know that there is also an effect on variability in SRT scores based on the choice of the noise masker type. Steady state noise (SSN) caused the scores to vary less than speech babble maskers (see also Lopez-Poveda et al., 2017).

The large variance in SRT was the reason to employ a repeated measures design, to be able to assess the effect of processing strategy on the level of an individual subject. The benefit of this design becomes obvious when comparing the results for subject 6 and 7 (who were tested only once per condition) to those of the other subjects. Subject 6 and 7 seem to raise the most extreme results when looking at Figure 6 and Figure 10, while in fact, their responses did not stand out from those of the other subjects when compared at the level of the individual sessions.

We observe only minor difference between the processing strategies at the group level as well as at the level of individual subjects. Comparing these results to the earlier work of Lopez-Poveda and colleagues (2016, 2017, 2018), we believe that the spatial and temporal parameters, as well as the type and difficulty of the task are possible factors that can explain the decline of the differences in SRT introduced by the different processing strategies. At this point, it must be stressed that we used vocoder simulations
with NH listeners for the present study, thus the results may differ from those that would have been obtained with actual CI users.

The spatio-temporal setup of the task was intentionally chosen to be particularly unfavorable to the MOC strategies. As shown by Lopez-Poveda and colleagues (2017), the benefit of the MOC strategies depends on the spatial relation of stimulus to noise, with stronger lateralization generally leading to a better SNR. The MOC1 strategy can underperform the STD strategy for unfavorable spatial setups (Lopez-Poveda and Eustaquio-Martín, 2018). For our experiment, signal and noise sources were always 30° apart, with the signal either on the same side as the masker (60° and 30°, respectively) or straight ahead (0°). This alone does not allow the MOC strategies to unfold their full potential.

In addition, we argue that the task itself has hampered the MOC processing strategies from outperforming the STD strategy. Word triplets are different from sentences in that they do not allow for the use of any contextual information. They are very short, but dense in factual information when compared to sentences. This makes them a more potent stimulus for psychometric testing, but also raises the bar for the subjects. Smits, Kapteyn & Houtgast (2004) have shown that the SRT for sentences is highly correlated with the SRT of word triplets in a similar task (r=.85).

However, repeating word-triplets at the 50% intelligibility threshold as quickly as possible and over an extended time period is very challenging. Repeated perception and reproduction of three completely unconnected words under time pressure is a demanding cognitive task in itself, and might introduce a higher variability in performance than sentences, dominating the overall variability and thereby potentially obscuring the effects of processing strategy.
Next to that, the challenging task has caused the mean SRT scores to be unexpectedly high for all participants, possibly due to memory and attention rather than auditory perception issues. From the way the MOC strategies work, they best differentiate themselves from the STD strategy in rather low SNR conditions. The relatively high SRTs might partially explain why there is overall little difference between the MOC and the STD strategies.

We expected to observe ISI to influence the SRT score (thus resulting in a statistically significant interaction between ISI and Strategy), particularly for the MOC3 strategy. Due to the long-time constants there is a difference in how the compression functions are set for different ISIs. For an ISI of 300 ms, compression parameters are effectively ‘reset’ between each single word in a triplet to reflect the influence of the noise masker alone. By contrast, for the shorter ISIs, especially the 2 ms ISI, the compression function for each word in the triplet is affected by the preceding word. Nonetheless, the only reliable pattern we observed was an underperformance of MOC1 as compared to STD (Figure 5). This is confirmed when looking at the single subjects (Figure 6). MOC1 stands out, at varying levels, to result in slightly higher SRTs, across most participants. For the SFS condition, this is supported statistically by the LME model. Assuming that this pattern was entirely due to the difference between the strategies, we are faced with the question why MOC1 underperformed and why MOC3 performed similarly to STD.

We attribute the slightly worse performance of MOC1 to the hypothesized effect of excess temporal variability on the compression function that might cause the envelope of the MOC1 output signal to “jitter” (Lopez-Poveda & Eustaquio-Martín, 2018). We have no explanation for why we observed similar SRT scores for MOC3 over all ISI levels.
In conclusion, there is no reliable and convincing difference in performance between the different strategies that would be of clinical significance.

**Response times**

The single-task RT seems to be highly sensitive to the SNR (Figure 7) and the ISI (Figure 8). When looking at the difference between the strategies and the interaction with ISI (Figure 9), we found that while for the 100 ms and the 300 ms ISIs there was not much difference between the strategies, for the 2-ms ISI, MOC1 seems to induce slightly longer response times. The GLME-fitted analysis supported this visual observation, showing that MOC1 performed slightly worse than STD in the FSF condition, on average. The effects depicted in Figure 9 are far from convincing, especially when looking at the contradicting direction of effects across subjects (Figure 10). Nonetheless, in the statistical analysis we found MOC1 to differ significantly from STD in the FSF moving direction. As mentioned before, we did not have prior assumptions on what we expected the effect of the different strategies on listening effort to be. One possible explanation for the small increase in RT mirrors that one already mentioned for the SRT results. The stronger fluctuation in the compression rate of MOC1 could have made word comprehension more effortful, as compared to the smooth compression fitting in MOC3 and the fixed fitting in STD. This hypothesis however fails short to explain why the effect occurs mainly for the 2-ms ISI.

The difference induced by the ISI is in line with the conceptual model of listening effort. We assume that the more information has to be processed in a given time span (speed of presentation), the higher the cognitive effort that is required, (compare: Piquado, Benichov, Brownell & Wingfield, 2012), which is reflected in longer response times.
The observation that RT is sensitive to SNR is reflected by earlier studies (Houben et al, 2013; Pals et al, 2015). This observation confirms us in the notion that the single-task response time is a valid measure for listening effort.

Besides the already mentioned effects, the data suggest that later word-triplets within the same condition elicit longer response times. There are generally higher mean values in RT for those word-triplets that are separated by 2-dB steps than those that are separated by 4-dB steps, even at high SNR levels. However, as the quantification of this effect would require a completely different reprocessing of the output data, we can only speculate. Systematic slowing of the RT has been found in a comparable dual task study by Hornsby (2013) who interpreted this as a fatigue effect (see Borragán, Slama, Bartolomei & Peigneux, 2017). The main difference is that we observed a slowing within the blocks of 25 word-triplets (roughly 4-5 min of sustained effort), while Hornsby found a moderate slowing between consecutive stimulus blocks (roughly 1 h of sustained effort, including brakes).

**Methodology**

It has been difficult to collect all the data rigorously in complete sessions. Many sessions were split over different days. This introduces variability in the factor *Session* that cannot completely be accounted for by the statistical models. The introduced error is nevertheless assumed to be random, and so it does not undermine the results.

The experiment was conducted by different experimenters. Different experimenters, even though all reported to have scored the responses in a very strict way, might nonetheless differ slightly in what they attribute to individual pronunciation or even dialect and what they regard a wrong answer. For example, the distinction between singular and plural Spanish nouns (suffix -s as in pollo/pollos) can be difficult
to detect in noise. However, in spoken Spanish such suffixes are often dropped. This has led to several discussions about pronunciation and the most conservative rating method but did not endanger the validity of the experiment. It was made sure that the data of a single session were always rated by the same experimenter, so that any potential misalignment of rating behavior was modelled as variability between sessions, not affecting the factors of interest.

Computational load was much larger for the MOC3 than for the STD or MOC1 strategies due to the longer time constants in the MOC3 strategy, which amount to integrating over many more data points. This caused a delay in the presentation of the word-triplets of several seconds. To make presentation times similar across strategies, we took two measures. First, we clipped parts of the tail of the integration window for the MOC3 strategy. In other words, those parts of the contralateral output data that had only a marginal contribution to the ipsilateral compression coefficient, were disregarded. The remaining 2.1 seconds in excess processing time were accommodated for by introducing a corresponding time pause before presentation of the STD and MOC1 stimuli.

When participants are given conflicting instruction experimenters often observe that some parts of the instruction are neglected in favor of others. Here, we instructed participants to be as accurate as possible while being as quick as possible. We tried to account for both instructions in a balanced way, by correlating the corresponding outcome measures. There was no correlation (Pearson’s’ $r = .1$; Figure 12) between SRT and the average of the corresponding RT scores (average of the last 17 word-triplets). This indicates that there was no detectable systematic penalty on the RT for a low SRT, nor did faster responses systematically come at the price of accuracy.
We measured RT manually as the duration from the offset of the stimulus to the onset of the response all by hand (see Appendix D). This potentially leads to systematic effects of different first phonemes on the response onset time. For example, we observed that the onset time was more straightforward to determine for plosive sounds than for fricative sounds. The manual method is still preferred over an automatic approach due to the difficulty in implementing such a system and the little improvements that can be expected from it. While algorithms could lower the random error due to human inaccuracy they would probably show similar systematic errors between different phonemes and might above that be influenced by other factors, like speech volume and speaker. The recordings would nonetheless have to be screened by a human rater for incorrectly labeled non-responses. The random error in the human rating is above that marginal, with high correlations between raters. The potential systematic error between different phonemes in rating does was unlikely to have an

Figure 12: SRT data versus RT data. There is no observable tradeoff between RT and SRT scores.
effect on the results due to random picking of the stimulus words and the high number of trials.

**Single task versus dual task paradigms**

We believe, that in contrast to dual-task response times, single-task response times can be modelled in a more convincing way, when studying auditory perception by itself. Lo & Andrews (2015) argue that much of the cognitive research is based on the assumption that the chronometry of mental processes is inherently additive (compare Sternberg, 1967; van Zandt & Ratcliff, 1995). This assumption forms a theoretical basis for the model choice (in our case, the “identity” link function in the GLME). In the way the dual-task paradigm is often implemented this assumption is stretched across several modalities and various cognitive levels at the same time, inducing more unknown factors of variability that might ultimately disguise the effect of interest (van Zandt & Ratcliff, 1995). Additionally, it is to note that while both, single and double task paradigms seem to be valid in quantifying effort across different tasks, the single task paradigm avoids the problem of the unknown ratio of effort allocation between first and second task (McGarrigle et al., 2014).

**Limits of using vocoders**

Based on earlier research, CI users are expected to be able to exploit the benefits of the MOC strategy effects even better due to the fact that they do not have a natural MOCR. In contrast, it is assumed that the performance of CI users might be biased towards the processing strategy that they are used to (or a resembling one), based on the increase in performance that is usually observed to result from practice with a new device. Vocoder simulations in NH listeners do not have this problem, as the participants can be assumed to be equally naïve to all the processing strategies. Vocoder simulations cannot reliably predict the benefits of the CI device in the clinical
population. There are a lot of factors that vocoders do not account for, like the neural stochasticity, the varying refractory period durations over the stimulated neural populations and the spectral smearing for higher electrical intensities (El Boghdady, Kegel, Lai & Dillier, 2016). Nonetheless, vocoder simulation investigations in NH people offer a labor- and cost-effective starting point for further research.

**Further analysis**

We have not yet discussed the difference between FSF and SFS moving directions. This is because we believe that the spatial arrangements can be best analyzed using a completely different set of analysis methods. For example, we can quantify the relative number and position of the correctly identified words within each word-triplet. While this is beyond the scope of this project, knowledge on the spatial distribution of the correct/incorrect responses could help identifying possible reasons for the differences in the overall SRT and RT distributions between conditions and in specific between FSF and SFS. In addition, knowledge on the distribution of the STOI for the single strategies (Lopez-Poveda & Eustaquio-Martín, 2018) could be collated with the observed results. On top of that, the reverse presentation orders of FSF and SFS allow to differentiate position-related effects and temporal sequence position effects (e.g. see Golob & Starr, 2000) on the single words. The analyses that have been done so far only scratched the surface of what could be investigated in this rich dataset.

Future investigations should test the processing strategies in more realistic hearing situations, especially at more adverse conditions (lower SNR, syntactically more complex stimuli and in combination with other tasks).
Conclusion

We conducted a vocoder simulation study to investigate the effects of fast changes in stimulus direction at different inter-stimulus interval durations on the performance of different simulated CI processing strategies. In contrast to earlier findings by Lopez-Poveda and colleagues, the overall results show that the MOC strategies perform at a par with a STD strategy in our particular setup. The only difference was found in the SFS condition, where the MOC1 strategy underperformed significantly. There were no noteworthy effects of ISI, nor interactions between the two factors. It must be stressed that these findings were obtained at mainly positive SNR levels and using word-triplets. Further studies are needed to evaluate the MOC strategies with moving stimuli at lower SNR levels and for different spatial arrangements.

Besides the SRT, we also analyzed the oral response time for the word-triplets. We observed some similarities to the SRT outcome in the results. This time, only in the FSF condition was the MOC1 found to differ from the STD strategy, especially for the 2-ms ISI. In addition, we found a strong and reliable effect of ISI itself, with shorter ISIs producing longer RTs. Additionally, the SNR level moderates the RTs in a reliable fashion, supporting the notion that the single-task oral RT may be used to assess listening effort.

Based on the presented results we argue for further study of the MOC processing strategies, especially the advanced MOC3. The current assessment has not been able to find any difference in performance between the MOC3 and STD strategies due to even the most adverse conditions.
References


Appendix A

Visual check for linearity and homogeneity of variance in the SRT data.

When comparing the SRT model predictions with the SRT residuals the Figures A1 and A2 show a wide and random spread of the data, indicative of a proper fit of the linear model to the data. The data itself can therefore be assumed to adhere to the linearity assumption. There is also no indication of overall heteroscedasticity.

Figure A1: Predicted versus residual SRT for FSF. Figure A2: Predicted versus residual SRT for SFS.
Appendix B

Check for heteroscedasticity, primarily due to random factors subject and session.

The plots show relative homogeneity in the variance. Only for the FSF data (Figure B1), there is a neglectable relative increase in variance towards the right part of the data cloud. As values were fitted in the order of data collection, the x-axis is analogous to the random factor subject (further subdivided into session). The single participants with their individual capabilities are the major threat to homoscedasticity in this investigation. It is unlikely that another minor factor introduced substantial amounts of systematic difference in the variance that we were not able to detect in either Appendix A or Appendix B. We therefore conclude that homoscedasticity is given.

![Figure B1: Fitted values versus residual SRT for FSF.](image1)

![Figure B2: Fitted values versus residual SRT for SFS.](image2)
Appendix C

Check for normality of the SRT data

Looking at the QQ plots for FSF (Figure C1) and SFS (Figure C2), we are confronted with a moderate (but uniform) violation of the normality assumption in the upper quartiles. After several failed attempts to correct for the offset values by data transformation, we had to accept the shortcoming. Moreover, we did not want to exclude more data than necessary.

*Figure C1:* Standard normal vs. SRT quantiles for FSF. *Figure C2:* Standard normal vs. SRT quantiles for SFS.
Appendix D

Definition of the individual RT scores.

The RT scores were measured from the end of the noise ramp of the presented stimulus (lower channel) to the beginning of the response (upper channel). The validity of the response was checked acoustically to account for non-responses, coughing, etc. Whenever the beginning of the response preceded the end of the stimulus, the RT was kept at 0ms.

Figure D1: Screenshot of the RT measurement in Adobe Audition. The stimulus is in the lower, the response in the upper channel. The RT is the relative distance between the two signals.
Appendix E

Evaluation of the gamma distribution fit.

Observed versus predicted quantiles of the RT data for FSF (Figure E1) and SFS (Figure E2). Note that the line approximates a linear function at roughly 45°. This indicates a reasonably good fit of the distribution model.

Figure E1: Observed vs. predicted quantiles for FSF. Figure E2: Observed vs. predicted quantiles for SFS.
Appendix F

Check for linearity and homogeneity of variance for the RT data.

When plotting the predicted RT scores versus the residuals (Figure F1, F2), we can clearly see the effect of the artificial zero cutoff in the sharp diagonal edge, which hints at problems with homoscedasticity. Other than that, there is no clear pattern and the data forms a uniform cloud. This indicates that the choice of a linear link function can be considered justified. If we, e.g. would have needed a quadratic link function to fit the data, we would expect to observe remnants of the quadratic distribution in the plots (that is, given that the model distribution fit is correct. See Appendix E).

*Figure F1:* Predicted RT values vs. residuals for FSF.  
*Figure F2:* Predicted RT values vs. residuals for SFS.
Appendix G

Check for the homogeneity of variance in the RT data.

The variance is relatively homogeneous. Only for a small portion of the data (roughly between 700 and 1300), there seems to be a reduced variance. This corresponds to the data of subject two, who continuously showed reduced RT scores. Fortunately, the GLME is relatively robust to moderate violations of homoscedasticity.

Figure G1: Fitted RT values vs residuals for FSF.  Figure G2: Fitted RT values vs residuals for SFS.