UCL

Lucile Belliveau

The role of spatial cues for processing speech in noise

Nicholas Lesica (UCL)
Michael Pecka (LMU)

PhD thesis
September 2013 - 2017

LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

I, Lucile Belliveau confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

LBelliveau

Abstract

How can we understand speech in difficult listening conditions? This question, centered on the ‘cocktail party problem’, has been studied for decades with psychophysical, physiological and modelling studies, but the answer remains elusive. In the cochlea, sounds are processed through a filter bank which separates them in frequency bands that are then sensed through different sensory neurons. All the sounds coming from a single source must be combined together again in the brain to create a unified speech percept. One of the strategies to achieve this grouping is to use common sound source location. The location of sound sources in the frequency range of human speech in the azimuthal plane is mainly perceived through interaural time differences (ITDs). We studied the mechanisms of ITD processing by comparing vowel discrimination performance in noise with coherent or incoherent ITDs across auditory filters. We showed that coherent ITD cues within one auditory filter were necessary for human subjects to take advantage of spatial unmasking, but that one sound source could have different ITDs across auditory filters. We showed that these psychophysical results are best represented in the gerbil inferior colliculus when using large neuronal populations optimized for natural spatial unmasking to discriminate the vowels in all the spatial conditions. Our results establish a parallel between human behavior and neuronal computations in the IC, highlighting the potential importance of the IC for discriminating sounds in complex spatial environments.

Table of Contents

I. Introduction

How can we understand speech in difficult listening conditions? This question, centered on the 'cocktail party problem', has been studied for decades with psychophysical, physiological and modelling studies, but the answer remains elusive. In the ear, sounds are processed through a filter bank and all the sounds coming from a single source must be combined together again in the brain to create a unified speech percept. One of the strategies to achieve this grouping is to use common sound source location. The location of sound sources in the frequency range of human speech in the azimuthal plane is mainly perceived through interaural time differences (ITDs): sounds from a source on the side of the head arrive at one ear before the other. These ITD cues are important for understanding speech in noise. We aim to study the integration of ITDs across frequencies by comparing vowel discrimination performance in noise with coherent or incoherent ITDs across frequencies. A discrimination task was designed to be applicable to humans and to an animal model, allowing for collection of psychophysical and physiological data. Our results will help to develop an integrated model of ITD processing, hopefully providing insight into strategies for speech processing in difficult listening conditions.

1. Motivations

Understanding speech in a complex environment is a challenging task that becomes especially difficult with ageing and hearing loss. People over 65 years old with normal hearing thresholds have more difficulty understanding complex sentences in noise than people under 44 years old. People with mild to moderate hearing loss also have more difficulty understanding speech in noise than their normal hearing counterparts of the same age group (Dubno, Dirks, and Morgan 1984). It is therefore important to understand the brain mechanisms underlying this perception to develop more targeted treatments, for example implementing efficient binaural listening in hearing aids.

Bilateral cochlear implant users also have a reduced performance for understanding speech in noise. They do benefit from the spatial separation of sound sources, but it is mainly due to monaural better ear effects (Loizou et al. 2009). It was shown that if subjects have a post-lingual deafness onset, they can sense ITDs if they are applied directly to electric pulses sent through the implants' electrodes (Litovsky et al. 2010). However, they are unable to take advantage of ITD cues for binaural release for masking (see below for definition) or lateralization (Hoesel and Tyler 2003). Studying how ITD cues are integrated in the brain might provide an efficient method of conveying these signals through bilateral cochlear implants.

2. Mechanisms of cross-frequency grouping

Sound processing by the cochlea can be roughly approximated by filtering through a bank of bandpass filters. As a first approximation, sounds are processed tonotopically through the brain, with each auditory structure organized in a gradient of neurons sensitive to different frequencies. Yet, when we listen to a complex auditory scene, we do not perceive sounds segregated in frequency bands but rather relevant auditory objects integrated over frequency. How is this integration achieved by the auditory system?

Integration of auditory information across frequencies is possible because in most cases all the frequency components of natural sounds have common properties. The components of a single sound stream have common onset time and source location. If the stream is a vocalization, the components can also be harmonically related with a common fundamental frequency. The influence of these cues on cross-frequency grouping has been extensively studied in psychophysical experiments.

We will use the terms 'component' or 'small frequency band' to refer to sounds that have a bandwidth that does not exceed the bandwidth of one auditory filter. We acknowledge that this definition is vague given the complexity of defining monaural and binaural auditory filter bandwidths. We will assume that pure tones and bands of noise of 150Hz bandwidth as used in some experiments discussed below comply with this criterion (Sondhi and Guttman 1966; Glasberg and Moore 1990), at least well enough to justify the conclusions drawn from these experiments.

a. Fundamental frequency, harmonicity and onset time

Fundamental frequency (F0), harmonicity and onset time contribute to grouping or segregating single tones from harmonic complexes and to grouping or segregating several complex sounds. Indeed, a pure tone is more likely perceived as separated from a harmonic complex if it begins at a different time than the harmonic complex (Dannenbring and Bregman 1978). Changing the onset time or mistuning one harmonic within a vowel changes the vowel identity in a direction consistent with removing the modified harmonic (C. J. Darwin and Hukin 1998). The same cues are used to separate two groups of sounds: two harmonic complexes are more likely grouped into a single vowel if they have a common fundamental frequency (Broadbent and Ladefoged 1957). The intelligibility of two simultaneously presented vowels is higher if the vowels have distinct F0s (Culling and Darwin 1993). And the intelligibility of two simultaneously presented sentences is higher if the they have distinct F0s (Darwin, Brungart, and Simpson 2003). Fundamental frequency, harmonicity and onset time are thus strong cues for cross-frequency grouping.

b. Sound source location

The effect of common source location on sound perception is more complex and controversial. Sound location is perceived through three main cues:

- Interaural level differences (ILDs): when the sound source is on one side of the head, the ear further away from the source receives the sound at lower intensity due to damping by the head. ILDs provide information on the azimuthal position of high frequency sounds (>2kHz for humans, because the head does not attenuate low frequency sounds that have a wavelength comparable to its size).
- Interaural time differences (ITDs): when the sound source is on one side of the head, the ear further away from the source receives the sound with a time delay due to the distance between the two ears. This causes an onset time difference and a continuous phase difference between sounds reaching the two ears. ITDs provide information on the azimuthal position of low frequency sounds (<2kHz for humans) and of high frequency complex sounds based on their envelope (McFadden and Pasanen 1976).
- Spectral cues: sounds coming directly from the source interact with their reflections from the pinna, head and torso at the ear drum, creating peaks and troughs at certain frequencies. Spectral cues provide information on elevation and help resolve the front/back ambiguity.

Constructing artificial sounds using only these cues is sufficient to make them lateralizable by a human listener (Wenzel et al. 1993), showing their behavioral relevance.

Separating sound sources spatially is very effective for segregating them into different streams and increasing their intelligibility. We saw that spatial release from masking was extensively studied (Bronkhorst 2000; C.J Darwin 2008 for reviews on speech) and depends on many different factors. However, the problem of across frequency grouping is slightly different as it is concerned with the formation of a single sound stream from frequency components rather than the segregation of two complex streams.

Culling and Summerfield (1995) were the first to test across frequency grouping by ITDs and ILDs explicitly. They presented 4 bands of noise of 150Hz bandwidth, of which every pair was identified as a different vowel. The subjects were asked to report on the vowel they heard, in conditions where pairs of noise bands shared a common ITD or ILD. If two bands of noise shared the same ILD, the subject could identify correctly the vowel formed by the pair. If they shared the same ITD, the subjects were unable to identify the vowel. This is evidence that subjects were able to group simultaneous bands of noise relying on ILDs, but not on ITDs. However, a later study showed that subjects could be taught to perform this grouping by ITD if they were extensively trained (W. R. Drennan, Gatehouse, and Lever 2003).

Hukin and Darwin (1995) tested the segregation by ITD by applying an ITD to a single harmonic composing a vowel and measuring the phoneme boundary. They found that applying an ITD to a single harmonic did not change the perception of the vowel identity. Interestingly, they found that if the same harmonic with the same ITD was presented on its own before the vowel, it did change the perception of the vowel identity (Darwin and Hukin 1997). Hence, ITDs seem unable to segregate a pure tone from a harmonic complex if they are presented simultaneously, but if the pure tone is already perceived as a separate stream, the distinction is maintained.

These results lead to the generally accepted idea that cross-frequency grouping does not rely on ITD: ITDs are processed separately for each frequency band and are only merged into a single location perception after the auditory object is defined (Darwin and Hukin 1999).

Recently, more evidence was gathered on cross-frequency grouping by ITD using full speech samples. Edmonds and Culling (2005) studied the intelligibility of a target sentence masked by another sentence or by brown noise (broadband noise that approximates the power spectrum of speech). The target sentence and the masker were split at 750Hz in a low frequency band and a high frequency band. After checking that both parts of the target sentence were equally intelligible but less intelligible than the full sentence, they compared performance in three conditions (Figure 1):

- Baseline: whole target and masker at +500µs ITD (target and masker at the same location),
- Consistent: whole target at +500µs ITD and whole masker at -500µs ITD (target and masker at opposite locations relative to the head midline),
- Swapped: low frequency target at +500µs ITD, low frequency masker at -500µs ITD, high frequency target at -500µs ITD, high frequency masker at +500µs ITD (the two frequency bands of the target are at opposite locations, and for each frequency band the target and the masker are at opposite locations).

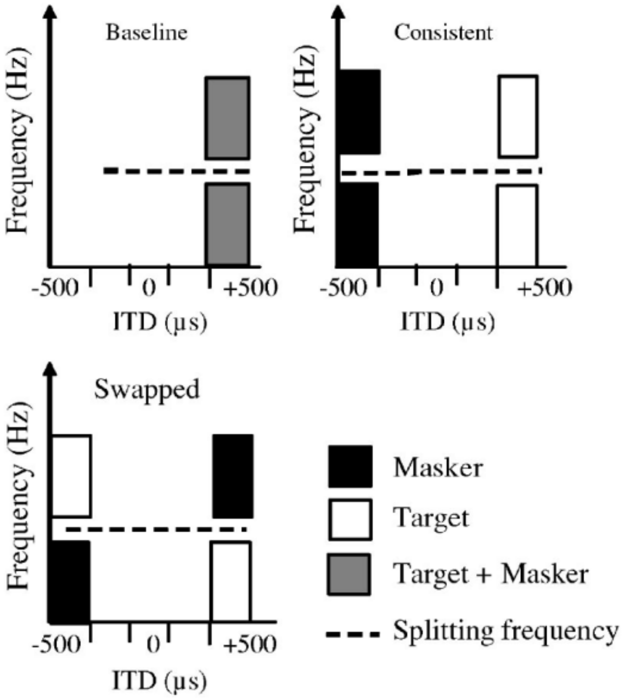

Figure 1: ITD conditions from Edmonds and Culling 2005. (Reproduced from Edmonds and Culling 2005)

In accordance with spatial release from masking results, target speech intelligibility was significantly higher in the Consistent condition than in the Baseline condition. Interestingly, the intelligibility was the same in the Swapped condition and in the Consistent condition. This confirms the absence of across frequency grouping by ITD for speech processing.

It is worth noting that the mechanisms underlying sound localization seem different than the ones underlying binaural release from masking (see I.3). Indeed, sound localization relies on grouping ITDs and ILDs across frequencies (Stern, Zeiberg, and Trahiotis 1988; Shackleton, Meddis, and Hewitt 1992) while will see in the next section that binaural release from masking seems to happen independently within each auditory filter.

3. Spatial release from masking

In a complex auditory environment, we are able to distinguish different sound sources on the basis of many properties such as the fundamental frequency or the spatial location of the sound. It is well known that separating two sound sources spatially improves their intelligibility and perceptual separation (A. W. Bronkhorst 2000). This intelligibility improvement, often called spatial release from masking, relies on binaural and monaural cues, and depends in a complex way on other factors such as the type of signal and masker and the room acoustics. We will give an overview of the psychophysical studies of spatial release from masking, with an emphasis on binaural mechanisms.

We will call ‘target' the sound that subjects have to attend to, either to detect its presence (detection task) or to understand its content (discrimination task). We will call ‘masker' the interfering sound that subjects don't have to attend to, which can be noise, speech or other material as we will discuss below.

a. Interaural level differences and better ear effects

When a sound comes from a source on the side of the head, the head shadow effect creates interaural level differences (ILDs). Indeed, the sounds arriving at the ear further away from the source are attenuated by the head, and have a lower intensity than the sounds arriving at the ear closest from the source. This affects mostly high frequency sounds as low frequency sounds are not significantly attenuated by the head. If a target and a masker are presented from different spatial locations, the ear closest to the target location will have a better signal to noise ratio (SNR) than the other ear, which is called the better ear effect.

The effects of these monaural cues on sound perception were investigated by measuring speech intelligibility in presence of a masker with a distinct ILD. Speech reception thresholds can be measured by finding the signal level for which the subjects can understand a fixed percentage of the words in the sentence (usually 50%), relative to a fixed masker level. Bronkhorst and Plomp (1988) showed that for a sentence presented in a white noise masker, the threshold was -6.4dB if both sounds had the same ILD and -14.3dB if the masker had an ILD corresponding to a location at 90° from the head midline while the target had no ILD (which corresponds to a location on the midline). This monaural release from masking is efficient only if the masker and target have energy in the same frequency bands (Gerald Kidd et al. 1998).

If the target is presented with several complex maskers at different locations, the ear with the best SNR will vary over time and frequency. The listener could take full advantage of the better ear effect if they were able to selectively attend to the ear with the best SNR for each frequency band and at each point in time (Paul M. Zurek 1993). Brungart and Iyer (2012) tested this hypothesis by measuring speech intelligibility in presence of two speech maskers coming from symmetrical locations relative to the head. In this condition, the ear with the best SNR varies with time and frequency. They also reconstructed their stimulus such that all the fragments with best SNR were presented to one ear, and all the other fragments to the other ear, which did not improve the performance. They hence concluded that listeners are indeed able to take full advantage of better ear cues in complex auditory environments.

b. Interaural phase and binaural masking level differences

In a seminal study for spatial release from masking, Licklider (1948) tested the intelligibility of speech presented in a white noise masker when inverting the polarity of the speech and/or masker at one ear. This polarity inversion gives rise to a phase shift of π of the sound at one ear, which can be detected only by binaural listening.

He found that the target speech intelligibility was the same when both the target and the masker were diotic (same sounds presented at both ears, referred to as N_0S_0 condition) and when both were inverted at one ear ($N_\pi S_\pi$). He found that the intelligibility increased when only the signal or masker was inverted at one ear (N_0S_π or $N_\pi S_0$). He also showed that the intelligibility decreased if both sounds were presented only at one ear (N_mS_m for monaural presentation). These intelligibility differences were later called binaural intelligibility level differences (Bronkhorst and Plomp 1988).

This phenomenon was extensively studied using a simpler paradigm where a single pure tone has to be detected in a white noise masker. Hirsh established this order of increasing detection performance: N_mS_m; N_0S_0 and $N_\pi S_\pi$; $N_\pi S_0$; N_0S_π (Hirsh 1948a, 1948b). The differences in performance between $N_\pi S_0$ or N_0S_π and N_0S_0 were termed binaural masking level differences (BMLD). A lot of models were developed to explain these differences, which we will discuss in a later section. The strength of the BMLD also depends on various other factors such as masker intensity or masker type, reviewed in Blauert (1997).

c. Interaural time differences

The BMLD paradigm is very useful to understand sound processing in the brain, but it does not model a real world situation. Indeed, the ear further away from a sound source receives the sound with a time delay compared to the closest ear. This interaural time difference (ITD) is present at the onset of the sound but also throughout the sound presentation, which gives rise to an interaural phase difference (IPD). A natural sound source and the reflexions on the head and torso will give rise to ITDs that vary slowly with frequency (Algazi et al. 2002), which correspond to IPDs that vary much faster with frequency.

Langford and Jeffress (1964) showed that BMLDs can be observed by applying a single ITD to the masker, which is equivalent to delaying the masker signal at one ear. For a pure tone presented diotically (S_0) in a white noise masker of varying ITD (N_τ), the BMLD was maximal when the ITD of the noise gave rise to a phase shift of π at the pure tone frequency. Levitt and Rabiner (1967a) studied the effect of applying a single ITD to a sentence presented in white noise on its detectability and intelligibility. Compared to the N_0S_0 condition, they observed that ITDs produced a detectability increase and a smaller intelligibility increase. They also found that these increases were smaller than those observed in the N_0S_π condition.

This implies that the subjective spatial lateralization of a sound does not play an important role in binaural release from masking, as applying a single ITD to a sound gives rise to a lateralized perception whereas inverting the signal at one ear gives rise to a diffuse perception. Other studies tested the intelligibility of sentences in white noise when the sentences were presented with opposite ITDs in adjacent frequency regions (for example Edmonds and Culling 2005a; Beutelmann, Brand, and Kollmeier 2009), and these manipulations did not affect the discrimination performance.

We explained previously that listeners could take advantage of monaural cues that vary in time and frequency. Even when binaural and monaural cues indicate opposite spatial locations of the sound source, the performance is not affected (Edmonds and Culling 2005b). Hence, it seems that listeners can take full advantage of binaural and monaural cues even if they lead to a diffuse and non-lateralizable perception of the sound.

Spatial cues can also be applied to sounds presented over headphones using a head related transfer function (HRTF), which models the effects the head and torso have on the sounds reaching the ears. Naturally, ITDs coming from a single sound source vary with frequency (Algazi et al. 2002), which is represented in the time delay component of the HTRF. The effect of using fixed or naturally varying ITDs across frequency seems small for binaural release from masking (Bronkhorst and Plomp 1988), so the effects observed using a single ITD value are probably a good estimate of the effects that would be observed using the time delay component of the HRTF.

The study of spatial release from masking using more natural spatial configurations can also be done by presenting sounds in free field, coming from speakers placed around the subject's head. In that case, binaural and monaural cues will be available. The contribution of binaural cues can be estimated by subtracting the performance in a monaural condition or the calculated estimate of the head shadow effect to the actual performance. For example, Dirks and Wilson (1969) studied the intelligibility of single words in white noise and found that subjects performed better in binaural than monaural listening conditions, even when using the ear with the highest SNR. Gerald Kidd et al. (1998) found that the masking of a pure tone sequence by multiple other tone sequences could not be accounted for by the head shadow effect only. This increase in intelligibility when sound sources are separated spatially was termed binaural release from masking, and can be considered as a generalization of BMLDs in more natural conditions.

d. Interaural correlation

The BMLD paradigm can be approached in a different way if we consider interaural correlation (for example N. I. Durlach et al. 1986): white noise presented diotically (N_0) is perfectly correlated at both ears (correlation coefficient c=1), and adding a pure tone with a phase shift of π (S_π) will decrease the interaural correlation at the frequency of the pure tone. This is also valid for the $N_\pi S_0$ stimulus with perfectly anti-correlated noise (c=-1) decorrelated by the diotic pure tone. Indeed, it was shown that BMLDs depend on the interaural correlation of the noise: BMLDs are maximal for fully correlated noise (which is the only case we considered until now) and decrease as the noise is decorrelated between the ears (Wilbanks and Whitmore 1968). This is consistent with the idea that detecting the decorrelation created by the pure tone is harder if the noise is less correlated overall, but does not prove that human subjects are sensitive to interaural correlations.

Pollack and Trittipoe (1959a; 1959b) measured human discrimination performance between bands of noise with varied interaural correlation. The subjects were indeed able to discriminate changes in interaural correlation, with better sensitivity to changes near perfect correlation (c=1 or -1) than near total decorrelation (c=0). This study was extended by Culling, Colburn, and Spurchise (2001), showing that this nonlinearity was lessened if the bands of noise were presented in broadband diotic noise. Hence, the auditory system seems to sense changes in interaural correlation, which supports their putative role in BMLD.

Interaural correlation also seems to have an effect even in bands very remote from the signal in the frequency domain. Marquardt and McAlpine (2009) tested the detectability of a 500Hz pure tone in the presence of one band of noise of various bandwidths centered on 500Hz and two independent flanking bands of noise. They showed that the detection performance was degraded if the masker configuration resulted in flat noise interaural correlation functions at any frequency. In other words, if the noise interaural correlation function was flat as far as 400Hz away from the pure tone frequency, it still had a detrimental effect on the detection performance.

e. Models

Different models have been developed to account for binaural and monaural effects in spatial release for masking and BMLDs (see Blauert (1997) for a review) but is still unclear how to model more complex issues such as room acoustics or type of interferer.

One of the most successful models in psychophysics is the equalization cancellation model developed by Durlach (Durlach 1963; Durlach 1972). This model processes sounds in two steps: the equalization step where sounds arriving at one ear are modified such that the noise coming from both sides is equal, which can be done by a time shift and/or amplitude modification the sounds; and the cancellation step where the equalized sounds from one ear are subtracted from the original sounds from the other ear, which if the process was perfect would cancel the noise entirely. The performance of the model is defined as the signal to noise ratio in the output. In the original implementation of the model, it is assumed that the equalization step is a noisy process, which is in fact necessary for agreement with psychophysical data. It is also assumed that sounds are first processed through a bank of bandpass filters at both ears.

The equalization cancellation model was applied to standard BMLD protocols (pure tone detection in white noise), using a single bandpass filter centered at the target tone frequency. This accounts well for the psychophysical data (for example Heijden and Trahiotis 1999), and offers an explanation for the fact that N_0S_π yields better performance than $N_\pi S_0$. Indeed, there is no need for internal delays to equalize the noise in the N_0S_π condition so the processing can be 'perfect'. In the $N_\pi S_0$ condition, internal delays are required to equalize and cancel the noise and this process is modelled as being noisy.

Heijden and Trahiotis (1999) also measured the discrimination performance when applying a single ITD to the noise ($N_\tau S_0$ condition for τ between 0 and 4000µs) and found that performance decreased for τ>750µs. They explain it by the existence of large internal delays (up to 4000µs) for which the equalization step is noisier. However, physiological data suggests that internal delays are confined within the π-limit: a range of delays between $-\frac{1}{2*F}$ and $\frac{1}{2*F}$ for a center frequency F within which each time delay corresponds to a single phase difference (David McAlpine, Jiang, and Palmer 2001). Marquardt and McAlpine (2009) developed a model using a bank of cross correlation detectors with time lags within the π-limit. In their scheme, signal to noise ratios are computed as the ratio of the cross correlation of the signal over the cross correlation of the noise for each frequency and time lag (within the π-limit). The best time lag is chosen for each frequency and a global SNR is computed using neurons which have the best SNR for each frequency channel. This model can account fairly well for Heijden and Trahiotis' data, so the existence of large internal delays doesn't seem necessary. Moreover, this model can also account for results from more complex stimuli where the interaural correlation in frequency bands remote from the target influences performance. It seems that models using interaural correlation could be a good generalization of the equalization cancellation model and be more applicable to physiological data and neural mechanisms.

An important result that emerged through the adaptation of the equalization cancellation model to complex tasks is that the equalization cancellation process takes place independently for each auditory filter (Culling and Summerfield 1995; Akeroyd 2004; Edmonds and Culling 2005a). This was termed the free equalization cancellation model, and is in keeping with the idea that lateralization is not important for spatial release from masking and that there is no across frequency grouping by ITDs, which we will study in a subsequent section.

f. Type of interferer

We saw that spatial release from masking could be studied using white noise or speech as a masker. This difference can be crucial for the masking effects, and a distinction is often made between energetic and informational masking. There is a lot of discussion on the exact definition of these terms (Kidd et al. 2007) so we only intend to give a broad understanding of the concept.

Energetic masking is traditionally thought to arise in the periphery of the auditory system when the target sound and the masker have power at the same frequencies. The target sound cannot be represented well by peripheral neurons and is more difficult to perceive. Informational masking is thought to depend on higher cognitive centers and arise when the masker can easily be confused with the target. For example, masking a target sentence with broadband noise would be energetic masking whereas masking a sentence with another sentence that the subject could mistakenly attend to would be, at least in part, informational masking.

It is difficult to construct stimuli that only give rise to informational masking because it requires the target and masker to have energy at distinct frequencies while remaining perceptually similar. Arbogast, Mason, and Kidd (2002) processed recorded speech through a bank of 15 butterworth filters of 1/3 octave bandwidth, and used a random subset of 6 frequency bands to construct target sentences. Subjects had to understand the target sentence in presence of different maskers:

- Same band noise: noise in the same frequency bands that were used to construct the target (energetic masking),
- Different band noise: noise in the frequency bands that were excluded from the target (not energetic, not informational),
- Different band sentence: a different sentence constructed using the frequency bands excluded from the target ('pure' informational masking).

They observed that when the target and masker were presented from the same spatial location, the performance was worse for the different band sentence than for the different band noise because the subjects reported words from the masker sentence instead of the target sentence. When the masker was moved to a different spatial location, they observed spatial release from masking in all conditions. With the same and different band noise, the effect could be accounted for using the head-shadow and binaural effects. With the different band sentence, the advantage due to spatial release from masking was larger and could not be explained by these acoustic properties.

These effects were observed in various other studies, including studies using tone sequences masked by other tone sequences or noise (Gerald Kidd et al. 1998) and birdsong masked by birdsong choruses or noise (V. Best et al. 2005), showing that these effects are not specific to speech. The authors suggest that the additional advantage of distinct spatial location using an informational masker is due to perception rather than acoustical properties: the subjects perceive the target and the masker as distinct auditory objects and can hence focus on the target better. This is contrary to the conclusions discussed before about spatial unmasking in noise where the lateralizability of sound sources did not seem to have an influence on perception, suggesting that mechanisms underlying informational and energetic unmasking are at least partially different.

g. Room acoustics

Most of the studies mentioned so far were conducted in anechoic chambers or over headphones modelling an anechoic environment, allowing no reflection or reverberation of the sounds. The effects of reverberant environments on sound perception are very complex and we will only give a brief overview.

The processing of reverberated sounds was often studied using delayed clicks: a first click is played from one speaker and a second click coming with a delay from a second speaker at a different spatial location, which models a reflection of the sound. If the delay between the two clicks is of 1 to 5ms, the sound is perceived as coming from the first speaker location. This led to the idea that the first (non-reverberated) segment of the sound to reach the ears determines more strongly our perception of the location of a sound (precedence effect, see Litovsky et al. (1999) for a review).

Using more complex stimuli, it was shown that reverberant environments impair spatial release from masking (Culling, Hodder, and Toh 2003) and that these effects also depend on target and interferer type (Kidd et al. 2005). These studies imply that speech reception thresholds in a reverberant environment can also be modelled using the equalization cancellation model (Zurek, Freyman, and Balakrishnan 2004; Beutelmann and Brand 2006).

4. Critical bandwidth of ITD processing

We have reviewed evidence showing that ITDs are processed in small frequency bands that presumably correspond to auditory filters, independently of ITDs at other frequencies. But what is the bandwidth of these binaural auditory filters? And are they the same as the monaural auditory filters?

a. Monaural filter bandwidth

Human auditory filter bandwidths are traditionally derived from pure tone detection thresholds in a notched-noise masker (Patterson 1976). Glasberg and Moore (1990) refined the bandwidth derivation process and applied it to several psychophysical data sets. They estimated values for filter equivalent rectangular bandwidth (ERB) in function of the filter center frequency Fc and found that $ERB = 24.7 * (4.37 * Fc + 1)$. This formula is widely used although there is still a controversy on the subject. For example, otoacoustic emission recordings yielded sharper filter estimates (Shera, Guinan, and Oxenham 2002).

While these results give a good approximation of monaural auditory filter bandwidths, they are not concerned directly with the bandwidths used for binaural information processing. It was shown that estimating auditory filter bandwidths using the same methods with the target tone inverted at one ear (N_0S_π instead of N_0S_0 or N_mS_m) gave a broader bandwidth filter estimate Hall, Tyler, and Fernandes (1983).

b. Binaural filter bandwidth

Sondhi and Guttman (1966) were among the first to estimate binaural filter bandwidths. They used a pure tone detection paradigm where a pure tone target was masked by a band of antiphasic noise of variable bandwidth centered on the pure tone frequency and flanked by two bands of homophasic noise ($N_{\pi 0 \pi}S_\pi$ or $N_{0 \pi 0}S_0$). They estimated bandwidth of a filter centered at 500Hz to be 200Hz, which is 2.5 times larger than the ERB estimate of Glasberg and Moore (1990).

Binaural bandwidths were also estimated using pure tone detection tasks with other masker configurations. For example, the masker can be composed of an antiphasic low frequency band and a homophasic high frequency band, with the distance from the pure tone to the frequency of the phase transition varied. Alternatively, the phase of the masker can vary according to a cosine function of varied period. Holube, Kinkel, and

Kollmeier (1998) tested these two paradigms along with the notched noise paradigm on the same subjects and used a single method to derive bandwidth estimates from the performance in the three paradigms. The monaural filter estimates were consistent across subjects and paradigms but the binaural bandwidth estimates were more variable. The latter were always larger than the monaural estimates, but were also a lot larger when using the masker varying according to a cosine function than the notched noise or single transition masker. The authors concluded that binaural processing may integrate information over several auditory filters, and that the variability between paradigms could be due to inappropriate bandwidth estimation methods.

Heijden and Trahiotis (1998) used a pure tone detection performance in a band of diotic noise of variable bandwidth and interaural correlation (N_0S_π with N at different correlation coefficients). They tried to model their results using independent binaural and monaural filter bandwidths, but the results could not account for the observed performance. They concluded against the necessity of having two different bandwidths for monaural and binaural processing.

Beutelmann, Brand, and Kollmeier (2009) estimated binaural filter bandwidth by testing speech intelligibility in complex binaural conditions and fitting the results to a model they previously developed Beutelmann and Brand (2006) that computes speech intelligibility after binaural processing through a free equalization cancellation model. They tested speech intelligibility in babble noise (a superposition of many sentences uttered by different talkers) while applying IPDs oscillating with different periods in the frequency domain to the target and masker. The period of the IPD oscillation was logarithmic in the frequency domain, to fit with the broader filter bandwidth observed at high frequencies, refining previous protocols where the IPDs varied cosinusoidally. They applied a continuum of IPD oscillations to the target speech ranging from slow (one half IPD cycle in 4 octaves: B=4) to fast oscillations (one half cycle in 1/8th of an octave: B=1/8), controlled by the parameter B (Figure 2A). They applied the same filtering process to the noise, either with IPDs of the same sign as the target (at each frequency, the IPD of the target is equal to the IPD of the masker: reference condition) or with IPDs of opposite sign (at each frequency, the IPD of the target is opposite to the IPD of the masker: binaural condition). They compared speech intelligibility in the alternating condition (speech IPDs between 0 and π/2, noise IPDs between 0 and –π/2) and in the non-alternating condition (speech and noise IPDs between –π/2 and π/2).

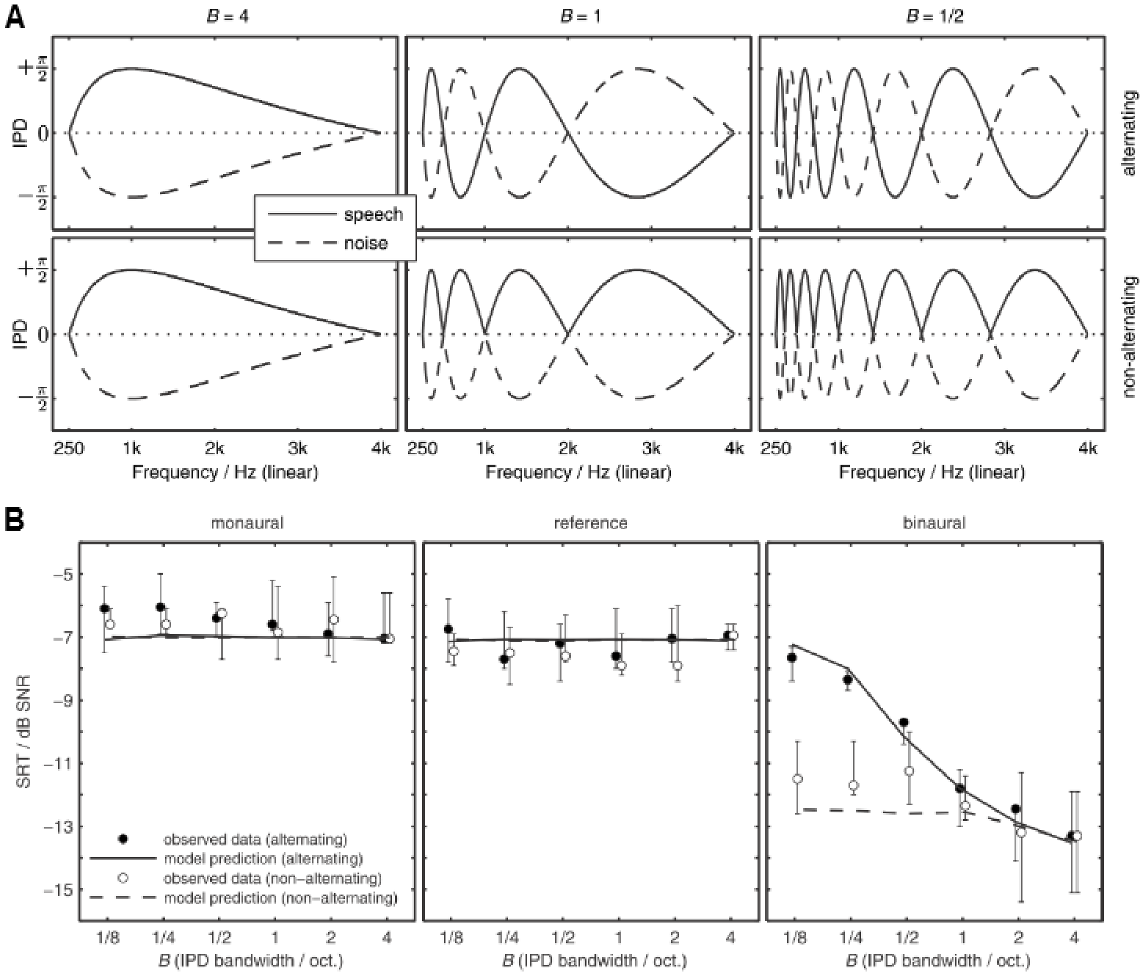

Figure 2: A. IPD conditions for speech (full lines) and noise (dashed lines). The IPD oscillation speed in the frequency domain is controlled by the parameter B. IPDs for the binaural condition (IPDs of speech and noise opposite at each frequency), in the alternating (top row) and non-alternating (bottom row) conditions. **B.** Speech reception thresholds (SRT) for all conditions for one sentence played in babble noise. SRTs are the speech intensity at which 50% of the words are intelligible in the presence of noise at a fixed intensity. Lower SRTs indicate better performance. In the binaural condition all sounds were presented as in A. In the reference condition the noise and speech IPDs were always equal. In the monaural condition one ear received the same sounds as in the binaural condition and the other ear received no sound. (Reproduced from Beutelmann, Brand, and Kollmeier (2009))

In the reference condition (Figure 2B), the noise and speech have the same IPD at all frequencies. The SRTs are high, consistent with the target and masker having the same binaural cues. In the monaural condition, sounds are presented only to one ear. The SRTs are again high, consistent with the absence of binaural cues. In the binaural condition speech and noise have opposite ITDs so binaural unmasking is possible. In the non-alternating condition they observe low SRTs for all B values, proving that binaural unmasking is possible for all the B values they used. In the alternating condition, SRTs are low for small and medium B values but become high for larger B values, showing that binaural unmasking is disrupted when the IPDs oscillate too fast.

These results are consistent with a model of binaural processing without cross-frequency integration and a bandwidth of 2.3*ERB (ERB as defined by Glasberg and Moore (1990)). This bandwidth estimation is in good agreement with previous studies (Hall, Tyler, and Fernandes 1983; Sondhi and Guttman 1966).

We could argue that the value of B should be large enough that distinct IPDs can be defined for the target and the masker within each auditory filter. Looking at the stimulus manipulations, we can infer that the interaural correlation is high for large B values and decreases with decreasing B values. We saw previously that binaural masking level differences were smaller in less correlated noise and non-existent in uncorrelated noise (Wilbanks and Whitmore 1968), so a similar phenomenon might be at play. In this study, the masker is presumably still correlated at minimal B values but the interaural correlation of the target also decreases, which might prevent any binaural intelligibility difference.

5. <u>Mechanisms of ITD processing in the mammalian brain</u>

a. Relays of ITD sensitivity in the auditory pathway

Sounds coming from the contralateral ear already have an effect on auditory nerve fibres responses through cochlear efferents (Warren and Liberman 1989), and binaural responses are already observed in the cochlear nucleus (Shore et al. 2003) and in the superior olivary complex (SOC). Most of the ITD sensitive cells are found in the medial superior olive (MSO) (J. M. Goldberg and Brown 1969; Yin and Chan 1990b), and some are in the low frequency part of the lateral superior olive (LSO) (Tollin and Yin 2005; Joris and Yin 1995). The MSO receives direct bilateral excitatory input from the cochlear nucleus (CN) and bilateral inhibitory input from the CN via the lateral nucleus of the trapezoid body (LNTB) for the ipsilateral CN and medial nucleus of the trapezoid body (MNTB) for the contralateral CN (Oliver 2000). All four ascending inputs are phase locked to sounds up to 2kHz, meaning that the neurons discharge at higher probability at specific phases of the stimulus. Temporal precision is key here as neurons have to resolve very small time differences (ITDs of 30µs to 660µs for humans).

The next major station in the primary ascending auditory pathway is the inferior colliculus (IC), with most ITD sensitive cells present in the central nucleus (ICC). The binaural sensitivity arises from direct excitatory input from bilateral MSO and contralateral

LSO. The ICC also receives direct inhibitory input from the ispilateral LSO and indirect inhibitory input from the dorsal nucleus of the lateral lemniscus (DNLL) that receives excitatory and inhibitory input from the LSO and MSO (Oliver, Beckius, and Shneiderman 1995; Jeffery A. Winer and Schreiner 2005a). The binaural information is then transmitted to the medial geniculate body (MGB) and to the primary auditory cortex, the IC being the principal source of ascending input to the MGB. In this project we investigated the mechanisms of ITD cues processing in the IC with a particular focus on cells with preferred frequencies lower than 2kHz in the dorsal part of the ICC.

b. Response properties of ITD sensitive neurons in the inferior colliculus

The ITD sensitivity of neurons in the inferior colliculus was probed by playing binaural stimuli to anesthetized animals. Rose et al. (1966) observed that some neurons in the IC had a cyclical discharge rate as a function of the ITD applied to one pure tone, and that the properties of this ITD tuning curve could change in function of the pure tone frequency. When ITD tuning curves are measured systematically with pure tones of different frequencies, the relationship between the pure tone frequency and the mean interaural phase at which the neuron responds can be modelled by a linear fit. The properties of these tuning curves can then be described in terms of characteristic delay (CD) and characteristic phase (CP) (Yin and Kuwada 1983; Kuwada, Stanford, and Batra 1987; Jeffery A. Winer and Schreiner 2005a). CD is defined by the slope of the linear fit between frequency and mean phase, and could represent the internal delay between sounds at one ear and the binaural cell. CP is defined as the phase intercept of the linear fit at 0Hz. It is a measure of the position of the intersection of the tuning curves at different frequencies relative to their peaks. These properties can be used to define three categories of neurons (Yin and Kuwada 1983):

- Peak type: CP near 0 or 1, the maximal firing rate is at the same ITD for all frequencies,
- Trough type: CP near 0.5, the minimal firing rate is at the same ITD for all frequencies,
- Intermediate type: CP near 0.25 or 0.75, maximal and minimal firing rates do not align with frequency.

Peak type neurons are thought to arise mainly from MSO input because they can be explained by two monaural excitatory inputs with a single time delay from one ear to the neuron. Conversely, trough type neurons are thought to arise from LSO input with one excitatory and one inhibitory monaural input with a single time delay on the inhibitory input. Intermediate type neurons could arise from convergent inputs from both structures.

This classification has been useful to characterize neurons properties but there seems to be a continuum between ITD tuning types rather than discrete categories in the IC, reflecting the convergence of inputs from different brainstem nuclei on individual IC cells.

A global best ITD (BD) across all frequencies can also be defined by averaging the tuning curves across frequency for each neuron. Neurons with low best frequencies (BF) have a wider range of BD that can exceed the physiological range while neurons with high BF have a narrow range of BD around 0µs ITD (David McAlpine, Jiang, and Palmer 1996). It seems that BDs are confined within the π-limit: the range between $-\frac{1}{2*BF}$ and $\frac{1}{2*BF}$ in which each time delay corresponds to a single interaural phase difference. This distribution of BDs in function of BFs allows the maximal slope of the ITD tuning curves to be in the physiological range. Indeed, if we consider ITD tuning curves measured at BF, they are periodic with period $\frac{1}{BF}$ which corresponds to a larger period for lower BFs. The maximal slope of the ITD tuning function will hence be further away from its peak for low BFs and having the peak further away from the physiological range will allow the slope to fall within it. This rationale led to the idea that the important variable for ITD coding is the variation of neurons firing rates and not whether they reach their maximal discharge rate or not (David McAlpine, Jiang, and Palmer 2001, 2).

c. Physiology of binaural masking level differences

We saw previously that BMLDs were extensively studied in psychophysical studies. This paradigm was also applied to physiological recordings, probing its neuronal mechanisms. We saw that BMLDs can be observed by applying an IPD or an ITD to the target or masker sounds, which can be sensed by ITD sensitive neurons.

The activity of IC neurons was recorded in response to pure tones masked by white noise in a classical N_0S_π paradigm. Neuronal BMLD was first measured as the increase in firing rate after the pure tone was added to the noise. It was shown that the best BMLD could be achieved for single neurons by playing the pure tone at their best frequency and best IPD for that frequency (David McAlpine, Jiang, and Palmer 1996; Caird, Palmer, and Rees 1991). The neurons showing the largest BMLDs were the ones that had the trough of their noise delay function near 0 ITD and hence did not respond a lot to the noise alone. For the best neurons, they observed a negative signal to noise ratio at threshold, which fits with the negative psychophysical thresholds.

Adding an antiphasic pure tone to diotic noise could also make the firing rate of neurons decrease. In a more general analysis, (Jiang, McAlpine, and Palmer 1997) showed that neurons had different behaviors in response to a 500Hz tone in function of their noise delay function and IPD tuning curve at 500Hz. They observed 2 categories of neurons:

- P-P: the neurons increase their firing rate in response to the tone in the N_0S_π and N_0S_0 configurations. If the firing rate increased faster with tone intensity in either condition, the neuron showed a BMLD.
- P-N: the neurons decrease their firing rate in response to the tone in the N_0S_π and increase their firing rate in the N_0S_0 configurations. If the firing rate decreased faster than it increased with increasing tone intensity, the neuron showed a BMLD.

This study shows that even without an optimized stimulus, BMLDs can be observed in many neurons. However, neurons with a best frequency near 500Hz are more likely to participate in the behavioral detection of the tone at threshold because the SNR at which they show a BMLD is smallest.

The same authors later showed that reducing the interaural correlation of the noise had the same effect on the firing rate of most neurons as adding an antiphasic tone to the noise (Palmer, Jiang, and McAlpine 1999). Namely, the noise delay functions became less modulated by the time delays, with lower peaks and higher troughs. This is consistent with the interaural correlation models of BMLDs.

d. Fine structure and envelope ITDs

BMLD paradigms use a single pure tone of various ITDs and hence rely on sensitivity to fine structure ITDs. While sensitivity to fine structure ITDs declines for frequencies higher than 1.4kHz for human subjects (Zwislocki and Feldman 1956), they are sensitive to ITDs in the envelope of high frequency complex sounds. We will discuss briefly the psychophysical and physiological evidence for envelope ITD sensitivity, concentrating on sensitivity to modulations around 60Hz of 1 to 2kHz carrier frequencies, because that is the most relevant for our study.

McFadden and Pasanen (1976) tested the lateralization performance of subjects presented with sinusoidally amplitude modulated (SAM) bands of noise of different bandwidths centered at 4000Hz. They showed that for bandwidths larger than 400Hz, the lateralization performance was similar to the performance for a 500Hz pure tone. The information contained in the envelope of the sound was hence sufficient to lateralize it.

Bernstein and Trahiotis (1985) tried to disambiguate the contribution of the fine structure and envelope ITDs on lateralization. They tested lateralization performance for SAM tones when the whole waveform was delayed by more than half the carrier period (and less than a full carrier period), which is less than half the envelope period. In that condition, the delay of the carrier and envelope point to the opposite sides of the listener's head. For carrier frequencies of 1kHz and modulations of 50 and 100Hz, they show that envelope cues do have an influence on lateralization, but don't override the fine structure cues completely.

Neurons in the IC are sensitive to envelope ITDs, but it was initially studied mostly for high carrier frequencies, which is not directly relevant to us (Batra, Kuwada, and Stanford 1989). Joris (2003) studied envelope sensitivity for low frequency carriers by comparing the ITD tuning curves for fully interaurally correlated noise and for the same stimulus with the signal inverted at one ear. Inverting the signal at one ear inverts the fine structure IPDs but does not modify the envelope. He observed neurons that had inverted ITD tuning curves in response to the latter stimulus, which means they are sensitive to fine structure ITD; neurons that had the same ITD tuning curves for both stimuli, which means they are sensitive to envelope ITD; and neurons that showed a combination of both effects. Neurons with characteristic frequencies (CFs) between 1 and 2kHz could belong to any of these categories, and Agapiou and McAlpine (2008) indeed observed envelope ITD sensitivity in neurons with BF below 1.5kHz.

Griffin et al. (2005) measured neuronal responses to SAM tones, which are closer to our stimulus than the broadband noise used in the previous study. They found that envelope ITDs could be predicted from single neuron activity, with the smallest just noticeable difference at around 600µs ITD for modulation frequencies of 100Hz. The carrier frequencies were fitted to neurons CFs so it is unclear how a population of neurons with different CFs would respond to a single SAM tone.

Although some work remains to be done to understand the complexity of this phenomenon, it is clear that sound lateralization depends on envelope ITDs and that neurons in the IC are sensitive to these cues.

e. Models of ITD sensitivity origin

A physiologically plausible model that accounts for how ITD sensitivity is created from binaural input and for the observed properties of ITD sensitive neurons has yet to be found. It is generally accepted that coincidence detector neurons receive inputs from both ears with various internal delays that compensate for the external ITDs, giving rise to neurons tuned to different ITDs. This idea takes root in the Jeffress model (Jeffress 1948) but several hypotheses exist to explain how the internal delay is generated (Joris and Yin 2007). It is worth noting that a simple coincidence detector model will fail to explain the dependency of BD on frequency so additional complexities will be necessary.

The historical hypothesis from the Jeffress model is that coincidence detector neurons receive input from axons of varying length which delay the arrival of the auditory signal. There is strong evidence for this hypothesis in birds, but not in mammals where no gradient of axonal length leading to the MSO was found.

More recently, it was suggested that coincidence detector neurons receive inhibitory inputs of varying strength and timing that delay the excitation (Brand et al. 2002). This hypothesis can explain the presence of BD outside the physiological range and is consistent with the concurrent emergence of BDs away from 0 ITD and inhibition during development. However the inhibition time constants required for this model are extremely fast and were not found so far in physiological recordings.

Coincidence detector neurons could also receive inputs from different regions of the two cochleae, which would create an internal delay (Shamma, Shen, and Gopalaswamy 1989). Indeed, low frequency sounds excite the apex of the basilar membrane which is distant from the tympani and are thus transmitted slower than high frequency sounds. The wiring precision from the coincidence detector neurons to the basilar membrane needed for such delays is plausible and its limitation could explain the similar BD distributions in mammals with big and small heads. However this hypothesis has not been tested extensively in mammals (only Joris et al. 2005 in auditory nerve fibers).

A combination of all these mechanisms could explain the observed properties of ITD sensitive neurons, but much experimental and modelling work has yet to be done.

f. Models of ITD population coding

The brain has access to a population of neurons with a wide range of ITD and frequency tuning. This information must be summarized in an ITD population code that indicates the ITD or the location of the sound. One prediction of the Jeffress model is the presence of an auditory spatial map which has not been found in the mammalian MSO, IC or primary auditory cortex. Another hypothesis is that ITD is coded by a two-channel model where the ratio of average activity in the two hemispheres is computed (David McAlpine, Jiang, and Palmer 2001). This hypothesis relies on neurons with a firing rate that varies approximately monotonically with ITD, hence that have the slope of their ITD tuning curve within the physiological range. As we saw previously, this is consistent with the dependence of BD on BF observed in *in vivo* recordings in the IC.

However, the two-channel model cannot account for the discrimination between multiple and single sound sources (Day and Delgutte 2013). Day and Delgutte (2013) hence suggested a pattern decoding model where the pattern of activity of all ITD sensitive neurons corresponds to a specific target and masker binaural configuration. This model could be implemented physiologically by an integration layer where each cell receives a weighted input of ITD sensitive cells. Such computation does not seem to happen in the tectothalamic circuit but the authors suggest it could happen in a higher auditory relay. Nonetheless, this model deals poorly with sound level changes while hemispheric models can take them into account (Stecker, Harrington, and Middlebrooks 2005).

6. Gerbils as an animal model

We are interested in probing the mechanisms of ITD processing at a neuronal level which forces us to use an animal model for single neuron activity recordings. We want to probe these mechanisms in the context of understanding speech in a complex acoustic environment so the animal model's audiogram and behavioral thresholds must be similar to human ones. Low frequency hearing is key because most of the power of speech is at low frequencies (<5kHz) and ITD cues are present only at low frequencies (<2kHz) for physical reasons.

Gerbils comply well with these criteria with a similar audiogram to humans and good low frequency hearing (Ryan 1976), making them more appropriate than other more common small animal models like mice or rats.

Our animal model must also be able to report detection and discrimination of speech-like sounds in a complex environment. Gerbils can detect vowels with similar thresholds as humans (Sinnott et al. 1997) and have successfully been trained to discriminate between 5 English vowels irrespective of the vocal tract length (Schebesch et al. 2010). They can also localize low-frequency sounds in the azimuthal plane in the presence of noise with the same acuity as humans when the difference in head size is taken into account (Lingner, Wiegrebe, and Grothe 2012a). In theory, they could therefore be trained to do a simple discrimination task with localized speech-like sounds in noise and that the results could be comparable to human performance.

Gerbils are also a suitable model because recording techniques developed for mice and rats are readily transferable to them. In fact, techniques for *in vivo* recordings and single unit isolation in the anesthetized gerbil IC are well established (Garcia-Lazaro, Belliveau, and Lesica 2013a). Techniques for recordings in awake behaving gerbils have been developed in the primary auditory cortex (A1) and in the IC (Ter-Mikaelian, Sanes, and Semple 2007a).

II. Psychophysical experiment: how do ITD cues influence vowel discriminability?

1. Probing the role of ITD cues for processing speech in noise in humans and gerbils

We will present and discuss the sound stimulus we used for both the psychophysical and physiological experiments.

a. Choice of speech and noise stimuli

To investigate the role of ITD cues for understanding speech in noise we needed to design a task that tested the intelligibility of speech in a complex auditory environment. This task had to be simple enough to be able to interpret neuronal activity in response to the stimulus, with the outlook that gerbils could eventually be trained to report on speech-like sound discrimination. It had to include different configurations of speech and masker locations with coherent and incoherent ITDs within one auditory filter so we could probe the influence of ITD cues on speech intelligibility and the mechanisms of ITD processing.

We chose to reduce human speech to isolated vowels. They are readily discriminable by gerbils (Schebesch et al. 2010) and humans. Pure tones can be localized if they have a sharp enough onset (Rakerd and Hartmann 1986) and single vowels can be localized by ferrets (Bizley et al. 2013) so vowels should be perceived as lateralized by humans and gerbils. Subjectively, we observed that applying ITDs to single vowels presented over headphones indeed gave rise to a lateralized perception (not shown).

Vowels can be approximated by a sum of sine waves (or harmonics) at different intensities and at frequencies that are multiples of the fundamental frequency. The maximum intensity peaks in their power spectra are called formants and define the vowel identity (Peterson and Barney 1952). We chose to reduce each vowel to two formants which makes the results more easily interpretable without reducing the amount of information available on vowel identity (Klatt 1980).

We chose to reduce each formant to only two harmonics (i.e. two sine waves) of frequencies centered on the formant's frequency. For example, a formant with a center frequency of 630Hz would be composed of one 600Hz and one 660Hz sine wave (Figure 3). If we consider a formant with the full harmonic spectrum in a noisy environment, the two center harmonics have the highest signal to noise ratio. Hence, simplifying our formants to contain only these two harmonics keeps the highest signal to noise ratio components and allows us to interpret the data with more confidence. For example, it will be easier to know whether a formant is within the receptive field of a neuron if it is only composed of two sine waves.

We chose to use a fundamental frequency of 60Hz for our vowels to be sure that each pair of consecutive harmonics would be unresolved (i.e. falling in the same auditory filter) for humans and for gerbils, even though such low fundamental frequencies are not typical for human speech. Our Reference vowel had one 630Hz formant and one 1230Hz formant (Figure 3). The human monaural ERB is estimated at 93Hz at a center frequency of 630Hz and 157Hz at 1230Hz. We saw in the introduction (I.4.b) that binaural bandwidths are estimated as the same or larger than monaural bandwidths. For gerbils, the auditory filters were estimated as broader than human ones (Kittel et al. 2002), so we indeed expect our harmonics to be unresolved for humans and gerbils.

We used babble noise as a masker, which consists of the superposition of sentences spoken by different speakers. It has the same power spectrum as speech (Figure 3), is not intelligible by humans and is more natural than white noise that has a flat power spectrum.

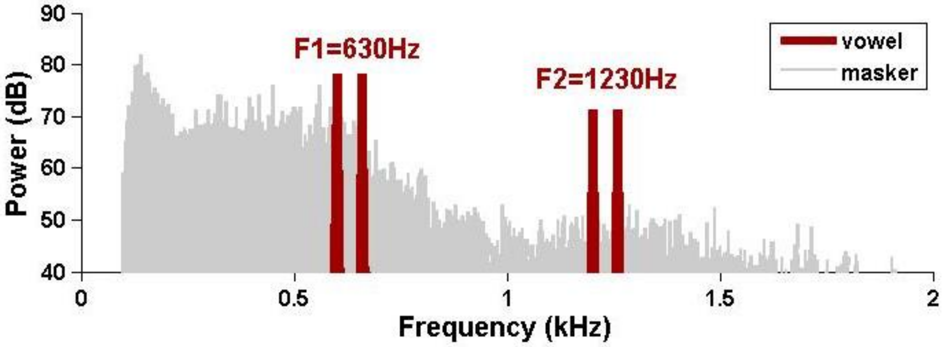

Figure 3: Frequency spectrum of the masker (babble noise) and of the Reference vowel. F1 is the center frequency of the first formant of the vowel; F2 is the center frequency of the second formant.

b. Structure of the discrimination task

Our stimulus was structured in successive trials where vowels were presented in pairs simultaneously with the masker. Each trial consisted of 750ms of masker alone, 250ms of masker with a first vowel, 350ms of masker alone, 250ms of masker with a second vowel and 350ms of masker alone (Figure 4A). The masker was ramped with a 50ms cosine ramp at the beginning and end of each trial. Each vowel had a 5ms cosine ramp at onset and offset. These sounds were presented to human and animal subjects through headphones. The psychophysical task was a Go/No-go task where the human subjects were instructed to press a button after trials where they heard a pair of identical vowels, and refrain from pressing the button if they heard two distinct vowels.

The first vowel presented in each trial was always the same vowel, which we will call Reference vowel. It was composed of a first formant of center frequency F1=630Hz (this formant was hence composed of two harmonics of frequencies 600Hz and 660Hz) and a second formant of frequency F2=1230Hz (composed of harmonics of frequencies 1200Hz and 1260Hz). The second vowel presented in each trial was chosen between (Figure 4B):

- the Reference vowel (R): F1=630Hz, F2=1230Hz;
- a Different vowel:
  - Different vowel 1 (D1): F1<630Hz, F2=1230Hz;
  - Different vowel 2 (D2): F1=630Hz, F2<1230Hz;
  - Different vowel 3 (D3): F1<630Hz, F2<1230Hz.

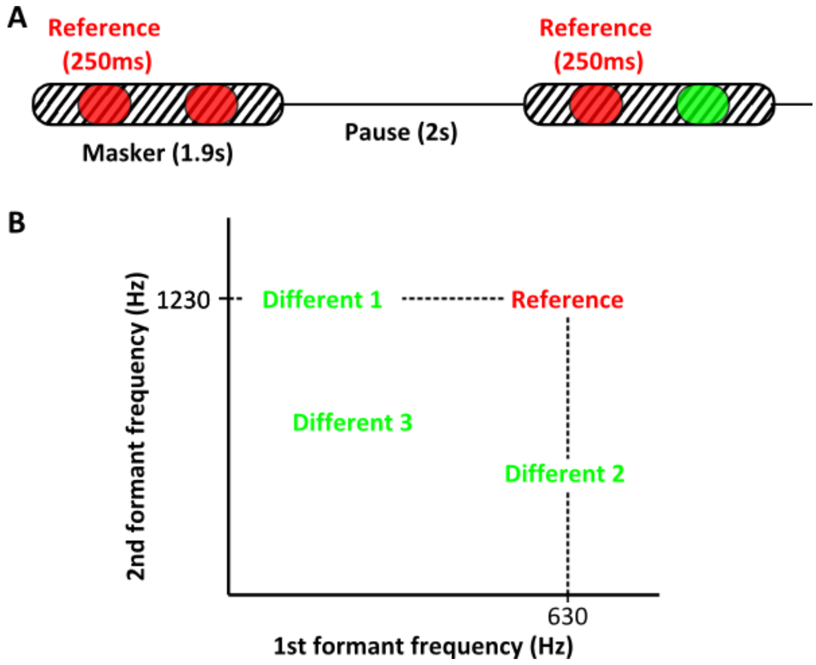

Figure 4: Structure of the stimulus. **A.** Structure of two example trials. The Reference vowel was always presented first and one of the four vowels (Reference or one of the three Different vowels) was presented second. A pause of 2s with no sound separated the trials. **B.** Center frequencies of the two formants of the four vowels.

We chose this frequency range for our vowels' formants to stay within the range of maximal sensitivity to fine structure ITDs which goes up to 1.4kHz for humans (Zwislocki and Feldman 1956). We chose the second formant of the Reference vowel close to the upper bound (F2=1230Hz), and hence had to use lower formant frequencies for all the other vowels. We chose the first formant frequency of the Reference vowel (F1=630Hz) at a plausible value for a vowel of F2=1230Hz (Peterson and Barney 1952) and still close to the human best frequency hearing range (Sivian and White 1933).

The Different vowel D1 was chosen to differ from the Reference vowel only by the first formant frequency. D2 differed from R by only the second formant frequency and D3 by both formant frequencies. For the psychophysical experiment, the exact formant frequencies for D1, D2 and D3 were adapted to each individual (see methods II.2.c) and they were fixed for the physiological experiments.

c. Spatial configurations of the vowels and masker

Our vowel discrimination task took place in presence of a masker, in five spatial conditions defined by the ITDs of the vowels and the masker (Figure 5). To facilitate comprehension, we will refer to sounds that are leading at the right ear as having a positive ITD and as 'coming from the right side of the head'. Conversely, sounds leading at the left ear will be referred as having a negative ITD and as ‘coming from the left side of the head’. We will refer to the different combinations of ITDs applied to the vowels and the masker as ‘spatial conditions’. The reader should remember that even though applying a positive ITD to a sound does create the perception that it is coming from the right side of the head (usually as an internalized perception on the right side inside of the head), we are using only ITDs as spatial cues and not the full head related transfer functions.

We used the following spatial conditions in our paradigm:

- Opposite (Figure 5A): the vowels are presented from the right side of the head (i.e at positive maximum ITD, +600µs for humans and +160µs for gerbils, giving rise to a perception at 90° from the midline of the head). The masker is presented from the left side of the head (i.e. at negative maximum ITD). All the vowel harmonics start in cosine phase (Figure 6A).
- Same (Figure 5B): the vowels and the masker are presented from the right side of the head.
- Split (Figure 5C): the vowels and masker are split in two wide frequency bands from 0Hz to 800Hz and 800Hz to 4000Hz. The low frequency band of the vowels (i.e. the first formant) is presented from the right side of the head while the low frequency band of the masker is presented from the left side. The situation is reversed for the high frequency band with the second formant of the vowels presented from the left side and the high frequency band of the masker from the right side. Hence, the vowels and the masked are each presented from two distinct locations but for each frequency band they are presented from opposite sides of the head.
- Alternating (Figure 5D): the vowels and masker have ITDs that change sign every 60Hz, which is the fundamental frequency of the vowels. The ITD of each vowel harmonic is opposite to the ITD of the noise at that frequency. For example, the Reference vowel has 4 harmonics at 600, 660, 1200 and 1260Hz. In the Alternating condition, the 600Hz and 1200Hz harmonics come from the right side of the head and the 660Hz and 1260Hz harmonics come from the left side. The bands of noise corresponding to these frequencies will come from the opposite side of the head. We note that the harmonics presented from one side of the head still start in phase, but out of phase with the harmonics presented from the other side (Figure 6B).

- Starting Phase: the vowels are presented from the right side of the head and the masker from the left side, like in the Opposite condition. The starting phases of the harmonics are the same as in the Alternating condition (Figure 6C).

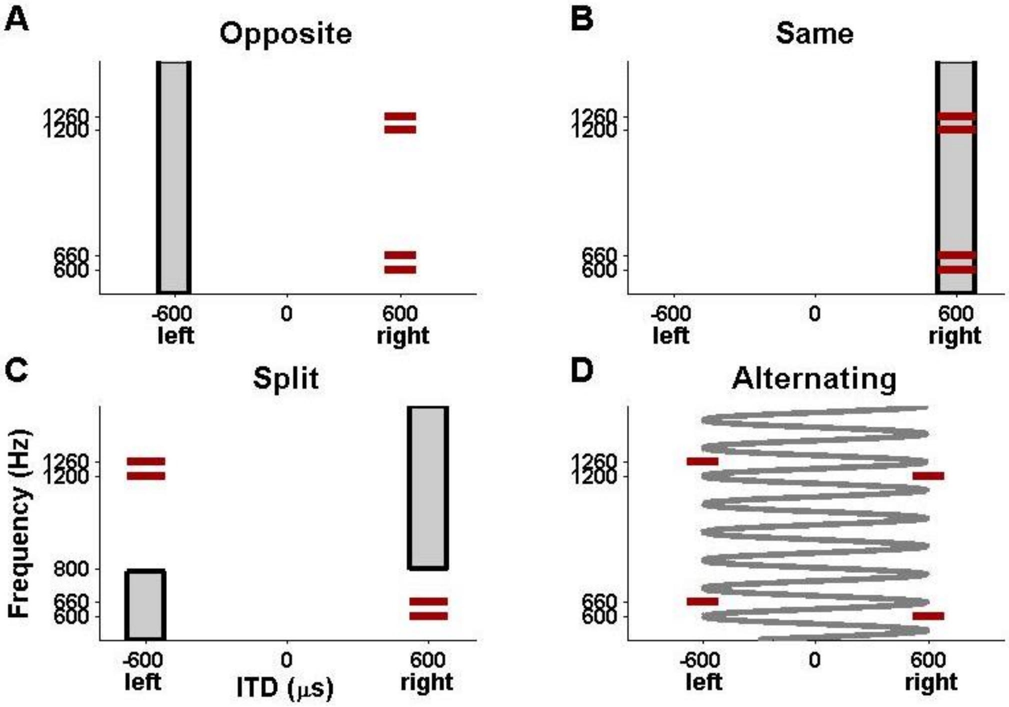

Figure 5: Schematic of four spatial conditions. The 4 harmonics of the Reference vowel are represented as red bars centred on their ITD, the masker is represented by grey rectangles centred on its ITD. The masker has power from 200Hz to 4kHz but is not fully represented. **A.** Opposite condition: the vowel is presented with a positive ITD so on the right side of the head, and the masker with a negative ITD so on the left side of the head. **B.** Same condition: the vowel and maskers are presented on the right side of the head. **C.** Split condition: corresponding frequency bands of the vowels and masker are presented on opposite sides of the head, with the first formant of the vowel on the right side and the second formant on the left side. **D.** Alternating condition: corresponding frequency bands of the vowels and masker are presented on opposite sides of the head, with the first harmonic of each formant presented on the right side and the second harmonic on the left side.

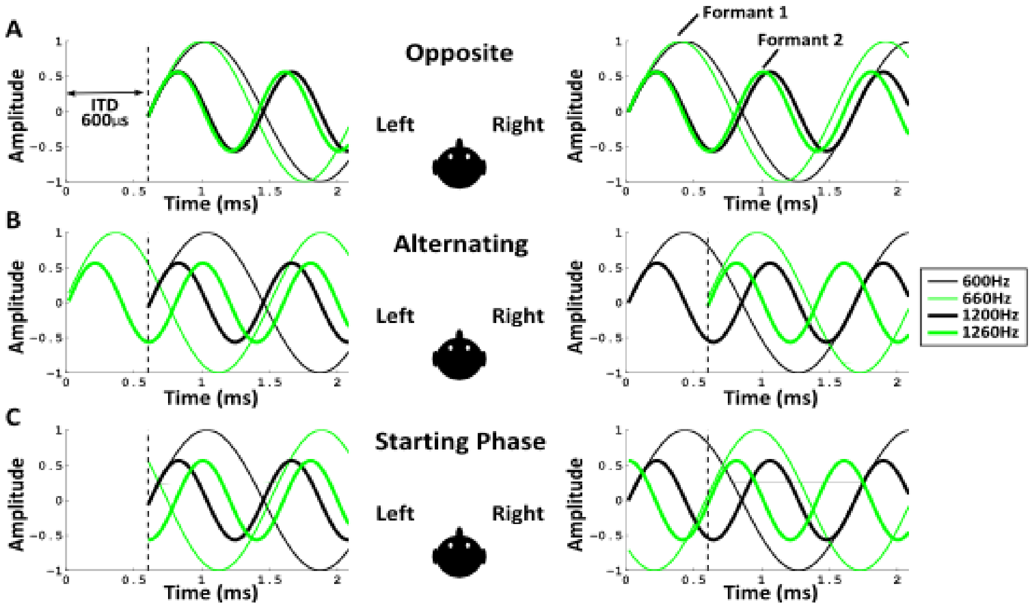

Figure 6: Schematic of the first 2ms of the Reference vowel harmonics represented over time at the left ear (left column) and right ear (right column). The harmonics are represented in sine phase rather than cosine phase for clarity and the 5ms cosine ramp that was applied to the beginning of the vowel is not represented. **A.** Opposite condition: all harmonics start in phase at both ears. They have an ITD of +600µs, which means they reach the left ear 600µs after the right ear giving rise to a perception of the vowel on the right side of the head. **B.** Alternating condition: the 600Hz and 1200Hz harmonics are identical to what they were in the Opposite condition, but the 660Hz and 1260Hz harmonics have a -600µs ITD which means they reach the left ear before the right ear. We notice that the starting phase relationships between the harmonics at each ear are altered by this modification. **C.** Starting Phase condition: all the harmonics have a +600µs ITD, but their starting phase relationships are identical to those in the Alternating condition: the 600Hz and 1200Hz harmonics start in phase with each other but out of phase with the 660Hz and 1260Hz harmonics.

In our stimulus, each vowel's formant is composed of two harmonics distant of 60Hz. It is hence probable that they will fall in the same auditory filter as unresolved harmonics. We can consider the waveform of each formant as the sum of the two harmonics that arrives at each ear. In the Same condition, the signal that arrives at both ears is identical (Figure 7A). As expected, we observed a fine structure of 630Hz for Formant 1 and of 1230Hz for Formant 2 for the Reference vowel, with an envelope of 60Hz no ITD between the two ears. In the Opposite condition, we applied an ITD of 600µs to each harmonic. We computed the ITD between each peak of the waveform for each formant, and we observed that the instantaneous ITD of the first and second formant was constant at a value of 604µs (Figure 7B). The envelope ITD was also constant at 604µs. In this condition, we hence expect spatial unmasking to be possible as the coherent ITDs of the two harmonics gives rise to a constant fine structure and envelope ITD for the full formant. We made the same observations (not shown) for the Split condition with a constant ITD of 604µs for Formant 1 and -604µs for Formant 2. In the Alternating condition we applied an ITD of 600µs to one harmonic and of -600µs to the other for each formant, resulting in a more complex stimulus. For the first formant of the Reference vowel, we observed a fine structure ITD alternating between 0µs and 770µs and a constant 4.0ms envelope ITD (Figure 7C). For the second formant, the fine structure ITD alternated between 0µs and 375µs with an envelope ITD of 8.9ms. These ITDs depend on the frequency of the formants, so they would be different for each vowel of our stimulus set. When we considered the full formants in the Alternating condition, we observed that the fine structure ITDs varied over time and were different from the envelope ITDs. We hence expect spatial unmasking to be difficult or impossible in the Alternating condition.

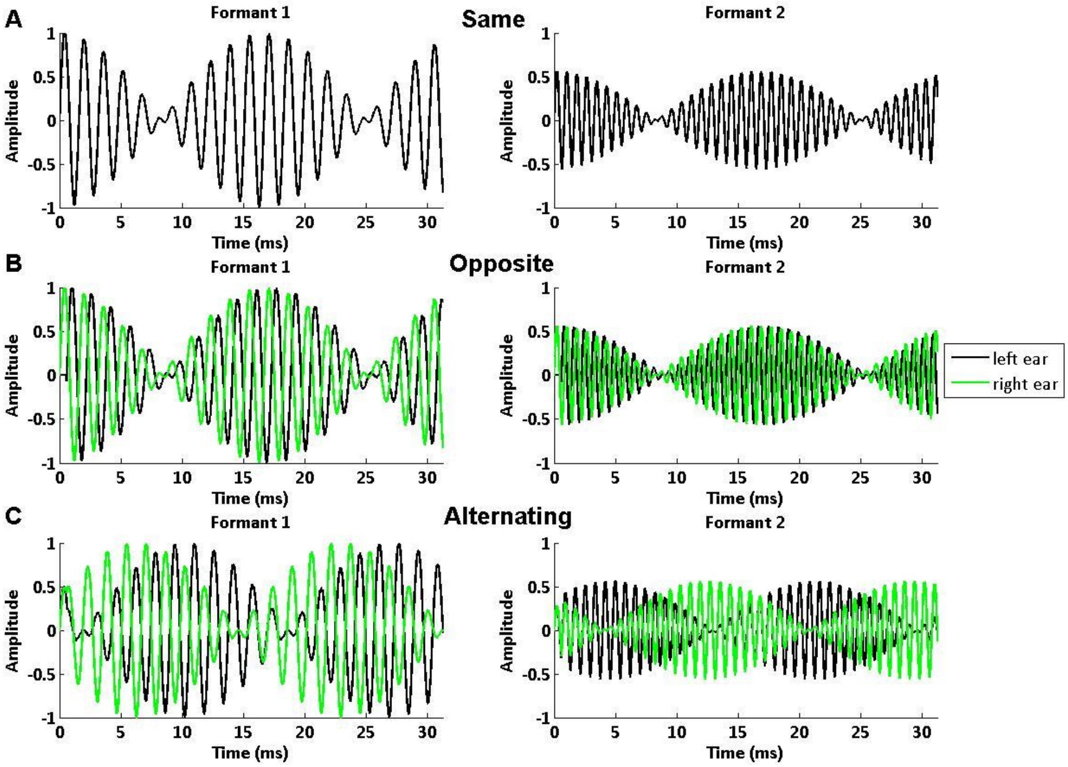

Figure 7: Schematic of the first 30ms of the Reference vowel first (left column) and second (right column) formant represented over time at the left ear (black) and right ear (green). Each formant is represented as the sum of its two harmonics. **A.** Same condition. **B.** Opposite condition. **C.** Alternating condition.

d. Discussion and predictions

We used full waveform delays to apply ITDs to our stimuli, so onset, ongoing and offset ITDs are present in the fine structure and envelope. We did not apply ILDs to our sounds but some intensity differences could still be present due to fluctuations of the sound intensity over time and frequency. However, our masker is a sum of many sentences, so its intensity does not fluctuate a lot over time. Some better ear effects could still influence the performance if for very short times one ear has a better signal to noise ratio than the other due to intensity fluctuations of the masker. The babble noise that we played was identical in all the spatial conditions with a randomized order, so the intensity fluctuations should influence every condition equally. The difference in performance between conditions can hence be attributed to binaural cues.

The babble noise masker can be defined as an energetic masker because it is not intelligible and has energy at the same frequencies as the vowels. We hence expect the performance to depend on acoustical cues with no additional effects of attention as were observed for informational masking. This will allow us to compare gerbil and human behavior and to relate behavioral performance to midbrain activity with more confidence.

If we consider the Same and Opposite conditions, our stimulus differs from the standard BMLD paradigm. We present tones with a single ITD, which gives rise to different IPDs in function of the frequency of the tone. The masker also has an ITD which means that the IPD of a harmonic is not the same as the phase difference between that harmonic and the masker at the same frequency. In usual BMLD paradigms, these two measures are the same because the target or masker is presented diotically (0 ITD and 0 IPD). For binaural release for masking, the important cue is the difference between the IPD of the masker (IPD_M) and the IPD of the target (IPD_T) which is the vowels in our paradigm. This difference can be expressed as:

$\Delta P = IPD_M - IPD_T = 2\pi F * ITD_M - 2\pi F * ITD_T$, for a frequency F

In our case, $ITD_M = -ITD_T$

So, $\Delta P = 4\pi F * ITD_M$

For humans we used $ITD_M = 600\mu s$, so $\Delta P = 4.75\mathrm{rad} \approx 3 * \pi/2$ between the masker and a tone of 630Hz and $\Delta P = 3.0\mathrm{rad} \approx \pi$ between the masker a tone of 1230Hz, so we can expect appreciable BMLDs.

For gerbils, we used $ITD_M = 160\mu s$, so $\Delta P = 1.3\mathrm{rad}$ for a tone of 630Hz and $\Delta P = 2.5\mathrm{rad} \approx \pi$ for a tone of 1230Hz, so we can also expect appreciable BMLDs.

One could argue that the harmonics of the second formant are too close to the upper frequency limit of ITD sensitivity for pure tones to produce BMLDs. However, the harmonics are not presented alone but rather in pairs of two unresolved tones. This creates an envelope equal to the distance between the two tones, in our case 60Hz. We saw that subjects seem sensitive to ITDs applied to SAM tones in this frequency range, so we expect the envelope to give additional binaural cues. From physiology experiments, we also saw that neurons of characteristic frequency around 1kHz can be sensitive to envelope and/or fine structure ITD cues (see Introduction I.5.d).

In light of these arguments, we expect the discrimination performance to be better in the Opposite than in the Same conditions, due to binaural release from masking.

We have considered the phase differences between the masker and the vowels, but the phase differences of the vowels between the two ears are the critical parameter for the perception of the sound location. It is worth noting that applying a 600µs ITD to a 630Hz pure tone gives rise to an IPD smaller than π (in that case, period≈1600µs and IPD=2*π*Freq*ITD=2.4rad) whereas applying a 600µs ITD to a 1230Hz pure tone gives rise to an IPD greater than π and smaller than 2*pi (period≈800µs and IPD=4.6rad). Hence, the two tones should be lateralized to opposite sides of the head. However, the vowels are clearly lateralized in the Opposite condition, in the direction expected for the 630Hz tone. The four harmonics could be grouped together by strong cues such as harmonicity and onset time, which gives rise to this unified perception. Moreover, both formants have the same 60Hz fundamental frequency, so the envelope IPDs and ITDs are consistent for the vowel. As we discussed in previous sections, lateralization does not seem crucial for binaural release from masking and BMLDs so this point is not central to our study.

Our Split condition was inspired by Edmonds and Culling's (2005) study. We expect binaural processing to take place independently in different auditory filters, so we expect the performance in the Split condition to be as good in as in the Opposite condition. We expect that binaural cues will be obtained from the fine structure of all harmonics, and that the lesser fine structure ITD sensitivity for the second formant harmonics will be at least partly compensated by envelope cues.

The Alternating condition was inspired by the fast alternating stimulus of Beutelmann, Brand, and Kollmeier (2009). An important difference is that we vary the ITDs of the sounds instead of their IPDs. As we just discussed, our ITD manipulations give rise to phase differences of π/2 to π if we consider the single harmonics, so a BMLD is still possible. Beutelmann, Brand, and Kollmeier (2009) varied the IPDs on a logarithmic scale to accommodate for larger auditory filter bandwidths at higher frequencies. In our stimulus, we varied the ITDs linearly. This means that the number of ITD reversals of the masker that happen in one filter will increase for high frequency filters of larger bandwidth. Because the harmonics are unresolved over the whole frequency range, and because there is at least one ITD reversal in any auditory filter, we assume that increasing the number of reversals further at high frequency will not have a major effect on the performance.

We expect performance in the Alternating condition to be as bad as in the Same condition even though the masker ITD and vowel ITD are opposite for each frequency because the ITDs are incoherent within one auditory filter. If we consider interaural correlations, the Alternating manipulation decreases the interaural correlation of the masker and the target vowel. This was shown to be detrimental to BMLDs and binaural release from masking, and we expect it to have a strong effect on discrimination performance.

The ITD manipulations in the Alternating condition also change the starting phase of the harmonics, so the performance could depend on monaural starting phase cues as well as binaural ITD and interaural correlation cues. We thus added a control condition where the starting phases are the same as in the Alternating condition, but the ITD configuration is identical to the Opposite condition. Starting phases do not seem to be a very salient cue for sound discrimination (Plomp and Steeneken 1969), and we restore the interaural correlations so we expect the performance to be as good in the Starting Phase condition as in the Opposite condition.

As we will show, this paradigm allowed us to show a binaural release from masking effect, gather more evidence in favor of the independent processing of binaural cues across auditory filters, and observe the effects of incoherent ITD cues within one auditory filter and decreased interaural correlations for spatial unmasking. We collected human psychophysical data and anesthetized gerbil physiological data for this task, and found models of neuronal population analysis that qualitatively match the psychophysical trends.

2. Methods

a. Subjects

12 subjects with normal hearing participated in the experiment. Their age ranged from 24 to 34 years old. 6 had no or little experience in psychophysical experiments while 6 had extended experience performing various types of psychophysical experiments. They were all paid a fixed amount for their participation. The experiment was approved by the UCL research ethics committee under Project ID 5069/001.

b. Stimuli

Vowels of 250ms duration were synthetized by adding sine waves of appropriate phase, frequency and intensity to construct two formants. Each formant was composed of two sine waves (or harmonics) of equal intensity placed at +30Hz and -30Hz from the formant center frequency. The harmonic frequencies were always multiples of 60Hz hence defining a fundamental frequency (F0) of 60Hz. The second formant was attenuated by 5dB relative to the first formant to mimic the structure of natural vowels. All sine waves started in cosine phase in the Opposite and Same conditions and the phase in the other conditions were dictated by ITDs (Figure 6).

The Reference vowel formant frequencies were set at F1r=630Hz for the first formant and F2r=1230Hz for the second formant. The formant frequencies of the Different vowels were chosen by an adaptive procedure explained in the next section, with frequencies ranging from 330Hz to 570Hz for the first formant and 930Hz to 1170Hz for the second formant. For Different 1, the first formant frequency was lower than F1r and the second formant frequency was equal to F2r. For Different 2, the first formant frequency was equal to F1r and the second formant frequency was lower than F2r. For Different 3, both formant frequencies were lower than the Reference vowel formant frequencies (Figure 4A).

The masker was synthetized by adding 25 sentences from the UCL SCRIBE data base spoken by male and female talkers and low pass filtered at 4kHz. 5 snippets of 1.9s duration were chosen and presented in random order.

For the Opposite condition, an ITD of +600µs was applied to the vowels and an ITD of -600µs to the masker (Figure 5A). For the Same condition, an ITD of +600µs was applied to the vowels and the masker (Figure 5B). The ITDs were applied by delaying (or advancing) the left ear signal and advancing (or delaying) the right ear by half the final ITD value when the ITD was positive (or negative).

For the Split condition, an ITD of +600µs was applied to the two harmonics of the first formant, and an ITD of -600µs to the two harmonics of the second formant. The masker was low-pass filtered at 800Hz and an ITD of -600µs was applied to it. It was also high-pass filtered at 800Hz and had an ITD of +600µs applied to it (Figure 5C).

For the Alternating condition, the masker was convolved with an impulse response designed such that the ITD alternated between -600µs and +600µs with a full ITD cycle taking place in 120Hz. This procedure is similar to that used in Beutelmann, Brand, and

Kollmeier (2009) and the code was kindly provided by Rainer Beutelmann. The ITDs were applied to each harmonic of the vowels during the vowel synthesis process such that the ITD of each harmonic was the opposite of the masker ITD at that frequency (Figure 5D).

For the Starting Phase condition, we created the vowels such that the starting phase between their harmonics was the same as in the Alternating condition, and applied an ITD of +600µs to all the vowels (Figure 6C). All the masker was presented at -600µs ITD.

c. Procedure

The subjects were seated in a sound proof booth. Sounds were presented through Phillips SHP2500 headphones using a laptop HD Audio internal sound card. The mean intensity level was 64dB SPL. Responses were recorded through left mouse button presses.

The subjects were informed that they were taking part in a synthetic vowel discrimination task involving sounds at different locations. They were told that they would hear successive sound streams containing two vowels (Figure 4A) and instructed to press the mouse button after the end of each stream if the two vowels were the same and to refrain from pressing if the two vowels were distinct from each other. After each 1.9s long trial, the subjects had 2s of silence during which they could press the button. At the end of this period, feedback was given to indicate whether the decision (pressing or not pressing) was appropriate by displaying ‘right’ or ‘wrong’ on the laptop screen. The next trial then began automatically until the desired number of trials was reached and the procedure stopped.

The subjects began by completing a training block with no masker and the vowels at +600µs ITD. The three Different vowels were presented three times each, and the Reference vowel was presented a total of 9 times so that the probability of hearing the Reference or any Different vowel was equal. The subjects then repeated this procedure with the masker at -600µs ITD presented 8dB SPL more intense than the vowels. They repeated the block with the masker until they reached 80% correct performance and felt comfortable doing the task, which took a maximum of two attempts for all the subjects.

The subjects then completed an adaptive procedure designed to make all the Different vowels equally difficult to discriminate from the Reference vowel. The task was the same as before with a Go/No-go paradigm on pairs of vowels, with the masker 14dB more intense than the vowels in Opposite condition. A 4-down 1-up staircase (Levitt 1971) converging to 84% correct performance was conducted simultaneously on all three Different vowels, while still presenting the Reference on half the trials. The adapted variable was the frequency of formant 1 for D1, the frequency of formant 2 for D2 and the frequency of formant 1 and 2 for D3. The frequencies varied in steps of F0 from their initial values, increasing toward F1r or F2r if the subject responded well hence making the discrimination harder, or decreasing if the subject made mistakes hence making the discrimination easier.

The procedure stopped once 4 reversals happened for each Different vowel or once the performance plateaued at maximal or minimal difficulty for at least 12 trials. If the performance for one or more Different vowels plateaued at minimal (or maximal) difficulty, the masker level was increased (or decreased) by 2dB SPL. If the performance for one Different vowel plateaued at maximal difficulty and the performance for another one plateaued at minimal difficulty, the masker level was not modified. Each Different vowel formant frequency was set as the mean of the 4 reversal values or the plateau value.

The mean across all subjects was F1=500Hz for D1, F2=1026Hz for D2 and F1=505Hz, F2=1105Hz for D3 with the masker 13.9dB SPL higher than the vowels. The Reference vowel always had a fixed F1r=630Hz and F2r=1230Hz.

Blocks of one ITD condition with 10 repetitions of each Different vowel (and 30 repetitions of the Reference vowel) were then presented to the subjects in random order. 9 subjects repeated the Opposite, Same, Split and Alternating conditions three times, leading to 30 repetitions of each Different vowel in each spatial condition. 4 subjects additionally did the Starting Phase condition three times, interleaved with the other conditions.

d. Analysis

The subjects were instructed to press a button when they heard a pair of identical vowels, and to refrain from pressing if they heard a pair of two distinct vowels. The data from all the Different vowels were grouped together as one condition. We call ‘hit’ a button press when the Reference vowel was presented twice during the trial (correct answer), and ‘false alarm’ a button press when any Different vowel was presented as the second vowel in the trial (incorrect answer).

For each condition, we computed the sensitivity d’ as:

$$d' = Z(hit\,rate) - Z(false\,alarm\,rate)$$

Z being the inverse Gaussian transformation. d’ is independent of response bias but it is calculated making the assumption that the distributions of stimulus similarity are normal. In our case, we can define similarity as how similar a vowel is to the Reference vowel. The normal assumption seems justified for the Reference vowel similarity distribution and for each Different vowel similarity distribution. But we analyse the data from all the Different vowels as one condition, so the similarity distribution of all the Different vowels might not be normal.

We presented the sound stimuli to the subjects in blocks of 60 pairs of vowels that were all presented in one spatial condition. Each subject had to complete 3 blocks of each spatial condition which were presented in a random order. They completed all the blocks during one single session on one day, and we observed that their performance was stable over time (Figure 8). We hence considered the data from the three blocks as one data-set for the rest of the analysis.

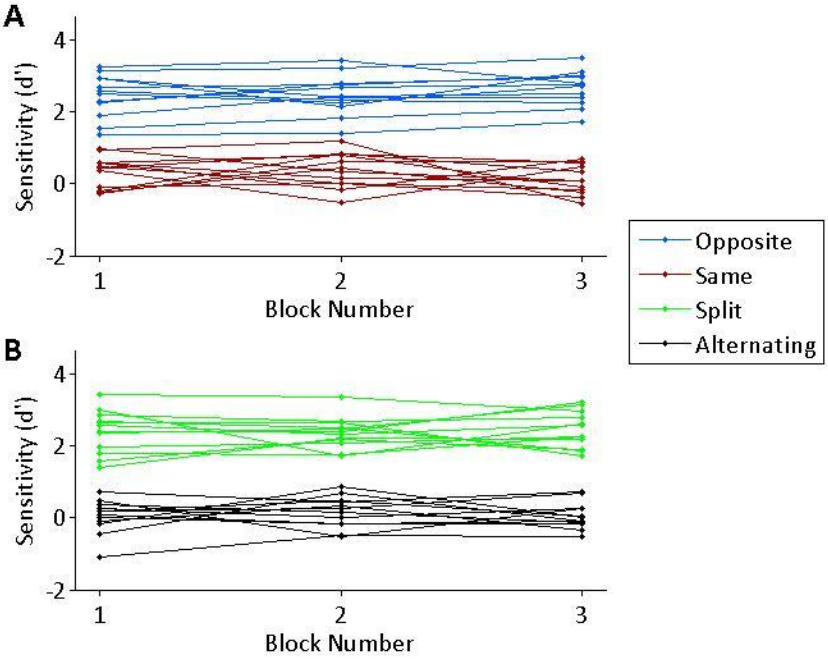

Figure 8: Sensitivity d' for the discrimination between the Reference vowel and all the Different vowels for each subject for the three blocks of each spatial condition. A significance measure of the difference between the responses for each block was computed using a one-way ANOVA for each spatial condition. A p-value larger than 0.05 does not allow us to reject the null-hypothesis that the mean results from the three blocks are equal. **A.** Opposite (p=0.63) and Same (p=0.32) conditions. **B.** Split (p=0.91) and Alternating (p=0.77) conditions.

As a comparison to the sensitivity measure, we computed the percentage of correct responses as the number of correct responses (hits and correct rejections) over the total number of trials. This measure has the advantage of requiring no assumptions on the data, but it is sensitive to response bias.

We also computed the response biases using the criterion c as:

$$c = -\frac{1}{2}[Z(hit\,rate) + Z(false\,alarm\,rate)]$$

If c is positive, it means that the subjects tend to respond ‘no’ more often, which means not pressing the button in our paradigm. If c is negative, it means that subjects tend to respond ‘yes’ more often which means pressing the button more often.

To have a measure of data stability, d’, percentages correct, and response biases were computed a hundred times for each subject using trials randomly drawn with replacement. This bootstrapping method allowed us to define a confidence interval by computing the standard deviation of the hundred measures.

e. Data exclusion

We observed that despite the adaptive procedure aimed at making each Different vowel equally distinct from the Reference vowel, some subjects were not able to discriminate D2 from the Reference vowel in the Opposite condition (Figure 10C, blue dots). We expected the performance to be best in the Opposite condition and to decrease or be constant in the other conditions. It is unclear whether this inability to discriminate our vowels based on the second formant comes from a lack of sensitivity to binaural cues (fine structure or envelope cues) or a lack of sensitivity to frequency changes in that frequency range in presence of a masker.

To avoid floor effects and see whether this phenomenon was important for our conclusions, we classified our subjects in a ‘non-excluded’ and ‘excluded’ group. Subjects were classified as excluded if their performance for discriminating any Different vowel from the Reference vowel was lower than chance.

The performance was computed for each Different vowel as the number of times the subject didn’t press the button in response to the Different vowel (correct response) divided by the total number of times that Different vowel was presented. This is a measure of the percentage of trials involving the Different vowel where the subject gave the correct response. A d’ measure couldn’t be computed for each Different vowel because it was difficult to define a false alarm rate. Indeed if the subject didn’t press the button when the Reference vowel was presented twice, it is an incorrect answer and should be considered a false alarm, but we do not know which Different vowel they mistook the second Reference vowel for.

The threshold for putting subjects in the excluded group was thus set to the percentage of correct responses to a Different vowel being inferior to 50% in the Opposite condition. This happened only for Different vowel 2 for 6 subjects (Figure 10).

We computed the performance of the subjects in the excluded group again, excluding all the data from trials involving Different vowel 2. If we consider that these subjects use only the first formant to do the discrimination task, we can include the data from the Different vowels 1 and 3 in most of the analysis. A problem arises for the Split condition: if we consider only frequencies below 800Hz, the stimulus is the same as in the Opposite condition. If these subjects do not use high frequency information, the performance in the Split and Opposite conditions should be the same but this would not inform the question of whether presenting one sound with different ITDs in distinct auditory filters has an effect on binaural processing. Hence, we chose to exclude these subjects entirely from the Split analysis. For the other conditions, the conclusions can rely only on the low frequencies so we chose to include the data from Different vowels 1 and 3 for the subjects from the excluded group.

3. Results

12 human subjects completed a synthetic vowel discrimination task where they were instructed to press a button if they heard a pair of identical vowel and refrain from pressing the button if they hear a pair of distinct vowels. Our stimulus set included 4 vowels, one we refer to as the Reference vowel because it was always part of the vowel pair, and three other vowels we refer to as Different vowels. The vowels were always presented in presence of a babble noise masker of higher intensity.

a. Influence of ITD cues on vowel discrimination performance

We measured the discrimination performance between the three Different vowels pooled together and the Reference vowel using the sensitivity d’ and the percentage of correct responses. The confidence intervals for each subject were measured by a bootstrapping procedure where trials from each subject were drawn randomly with replacement one hundred times. The confidence intervals for the means over all subjects were computed using the standard deviation between the each subject’s mean performances over all the trials they completed.

This was done for our five spatial conditions where the vowels and the masker were presented with the same or opposite ITDs, and with coherent or incoherent ITDs.

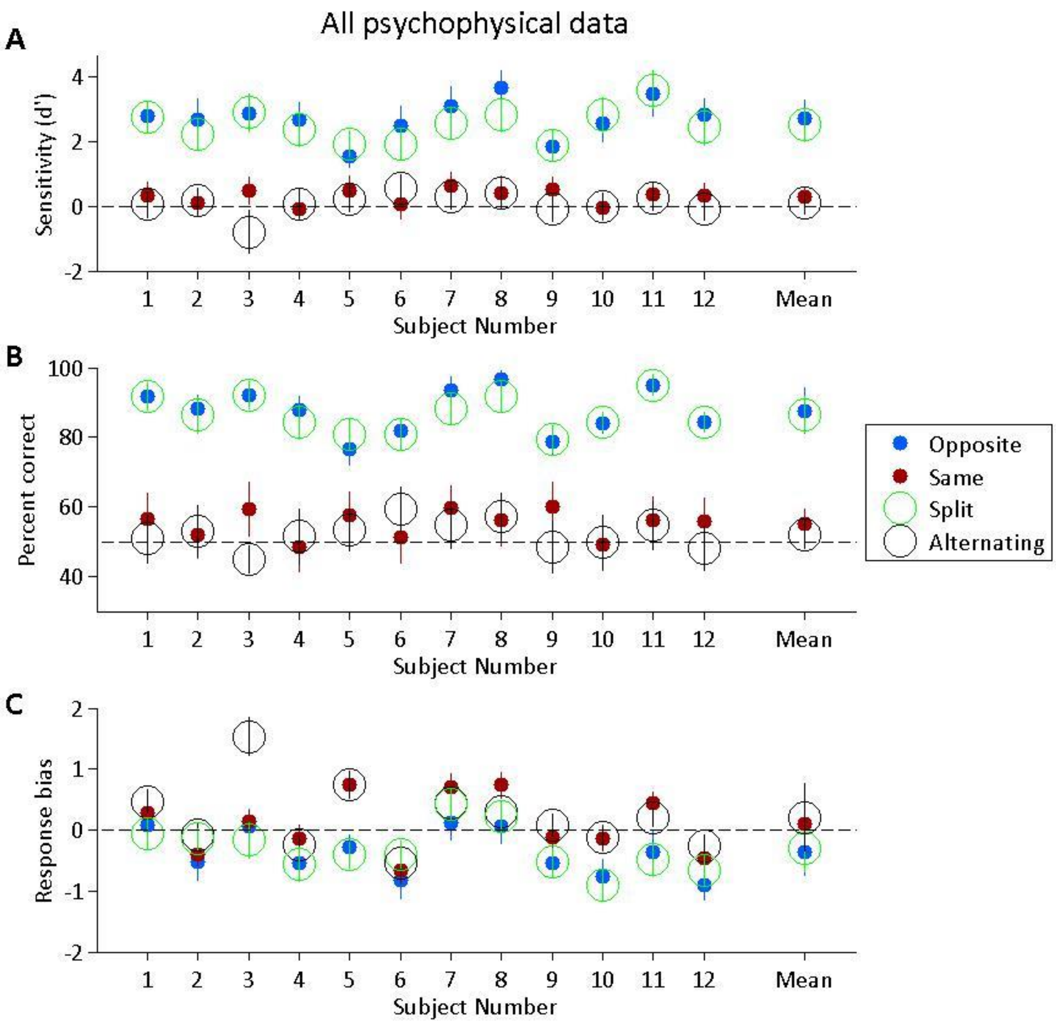

Figure 9: A. Sensitivity d' for the discrimination between the Reference vowel and all the Different vowels for each subject in 4 spatial conditions. The error bars represent 95% confidence intervals measured for each subjects by bootstrapping, and across all subjects by computing the standard deviation using each subject's mean value. The dashed line indicates the d' value where the subjects did not report perceiving a difference between the Reference and the Different vowels (d'=0). **B.** Same representation for the percentage of correct responses (PC). The dashed line indicates the PC value where the subjects chose to press the button or not at the end of each trial by chance (PC=50%). **C.** Same representation for the response bias (c). The dashed line indicates the c value where the subjects did not show a preference for pressing or not pressing the button (c=0).

In the Opposite condition where the vowels came from the right of the head and the masker from the left, we found that all subjects could discriminate very well between the different vowels (Figure 9A blue dots, mean d' over all subjects=2.7 +/-0.6). In the Same condition where both the vowels and the maskers came from the right of the head all the subjects discriminated the vowels very poorly (Figure 9A red dots, mean d'=0.2 +/0.2). The only difference between the two conditions was the ITD of the masker that changed from -600µs to +600µs ITD while the vowels were always presented at +600µs ITD. Hence the vowels and the masker should be perceived on opposite sides of the head in the Opposite condition and on the same side of the head in the Same condition. The fact that the discrimination performance was significantly higher (the confidence intervals did not overlap) in the Opposite condition compared to the Same condition showed that presenting the vowels and the masker at an opposite ITD was beneficial for vowel discrimination. This shows that binaural unmasking indeed took place in our paradigm.

In the Split condition, the vowels were split in a low frequency formant presented on the right side and a high frequency formant presented to the left side of the head, with the corresponding masker frequency bands presented to the opposite side. The subjects could discriminate between the vowels in this condition (Figure 9A green circles, mean d'=2.5 +/-0.5). The fact that the performance in the Split condition was not different from that in the Opposite condition shows that presenting each formant of the vowel at a different ITD did not impair the discrimination performance and hence did not prevent spatial unmasking from being beneficial. This replicated Edmonds and Culling (2005) result: ITDs are processed independently across auditory filters.

In the Alternating condition each frequency band of the vowels and masker were presented on opposite sides of the head but the two harmonics from each formant were presented on opposite sides of the head. The subjects could not discriminate between the different vowels in this condition (Figure 9A black circles, mean d'=0.1 +/-0.3). The fact that the performance in the Alternating condition was significantly worse than in the Opposite condition showed that binaural unmasking could not take place. The two harmonics from each formants had frequencies separated by 60Hz and were hence probably unresolved (i.e. falling in the same auditory filter). An auditory filter sensitive to the formant's frequencies would hence be stimulated by two harmonics of opposite ITDs. This showed that having incoherent ITDs within one auditory filter disrupted spatial unmasking, replicating the results of Beutelmann, Brand, and Kollmeier (2009).

These results are all conserved when using a percent correct measure (Figure 9B), suggesting that the similarity distributions of all the Different vowels and of the Reference vowel were indeed normal, and that response biases are small.

We computed the response biases and indeed observed values close to 0 (Figure 9C). They were between -1 and 1 for all subjects in all conditions except for subject number 3 in the Alternating condition. We observed that overall and for each subject individually the response bias for the Opposite and Split conditions (mean over all subjects in Opposite condition=-0.4+/0-.4; in Split condition=-0.3+/-0.4) was lower than the response bias for the Same and Alternating conditions (mean in Same condition=0.1+/-1.5, in Alternating condition=0.2+/-0.5). A negative response bias means that subjects responded ‘yes’ more often, which means they pressed the button more often in our paradigm. Conversely, a positive response bias means that they refrained from pressing the button more often. It hence seemed that when the subjects had more difficulty discriminating the vowels in the Same and Alternating conditions they pushed the button less often than when they could discriminate the vowels better in the Opposite and Split conditions. The difficulty of the spatial conditions hence seemed to influence the tendency of subjects to press or not press the button, but the response biases remained small overall.

b. Behavior in response to the Different vowels

To understand better the responses we observed, we looked at the subjects’ performance for each Different vowel independently. If we consider only one Different vowel, we defined a hit rate by counting the number of times the subject didn’t press the button when the Different vowel was presented (correct answer). Defining a false alarm rate is difficult because if the subject didn’t press the button when the Reference vowel was presented (incorrect answer), we do not know which one of the Different vowels he confused the Reference vowel for. Since we established in our global analysis that the response biases were small and that the sensitivity and percentage correct measures gave similar results, we chose to compute the performance for each vowel as the number of times the subject didn’t press the button when the Different vowel was presented (correct answer) divided by the total number of times the Different vowel was presented. For the Reference vowel, we used the number of times the subjects pressed the button when the Reference vowel was presented (correct answer) divided by the total number of times it was presented.

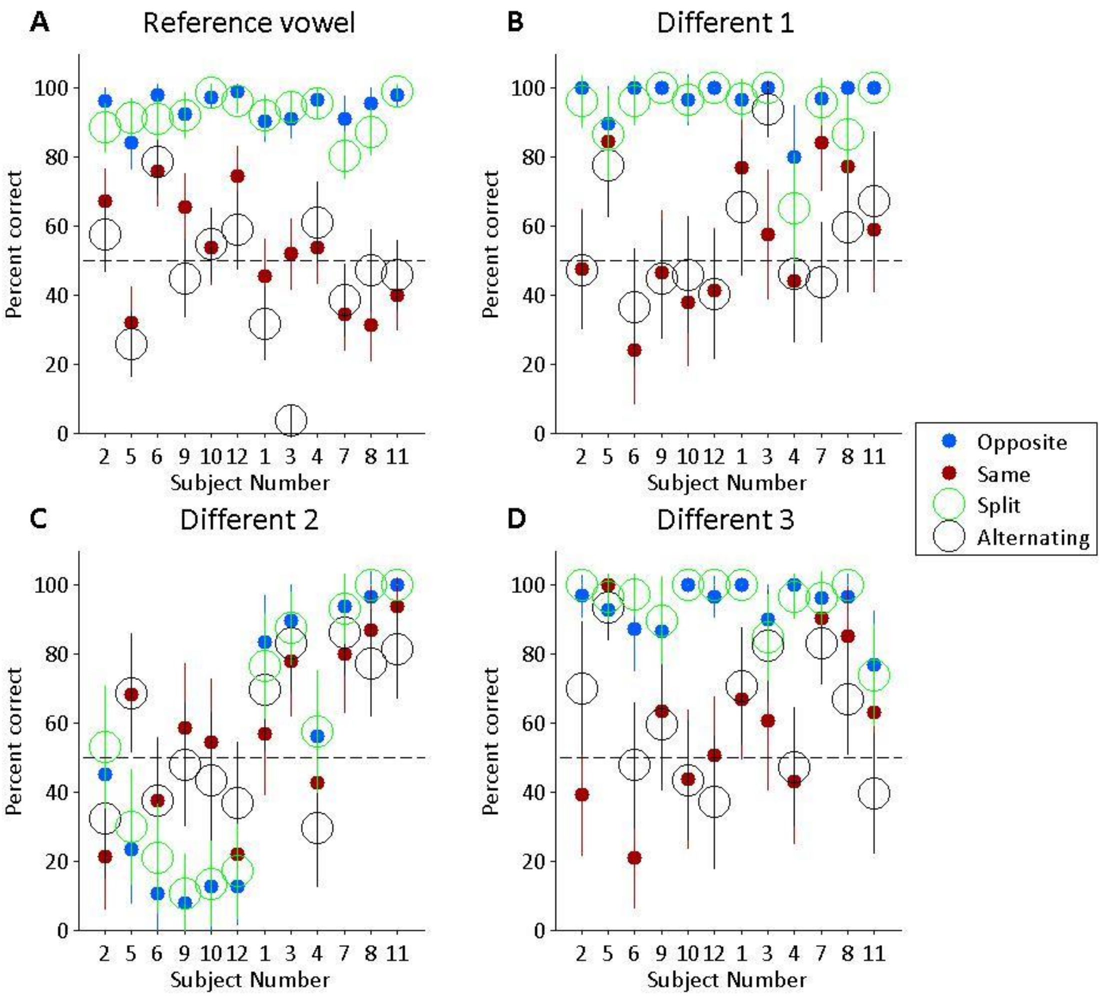

Figure 10: Performance analysis for each vowel independently. Percentage of correct answers of each subject computed as the number of correct answers in response to the vowel divided by the total number of presentations of that vowel. The subjects from the excluded group are presented on the left (subjects 2, 5, 6, 9, 10 and 12) and the subjects from the non-excluded group on the right. For **A.** the Reference vowel, **B.** the Different 1 vowel (frequency of the first formant lower than the frequency or the first formant of the Reference vowel), **C.** the Different 2 vowel (frequency of the second formant lower than the frequency or the second formant of the Reference vowel) and **D.** the Different 3 vowel (frequency of both formants lower than the frequency or both formants of the Reference vowel).

For the Reference vowel, the Different vowel 1 and the Different vowel 3 (Figure 10A, B, D) we observed the same trends as in the global analysis: the performance was good in the Opposite and Split condition and bad in the Same and Alternating conditions. For the Different vowel 2 (Figure 10C), we observed that 6 subjects had a performance under chance level (<50%) for discriminating Different vowel 2 in the Opposite condition. This means that even in the most favorable spatial condition they often misidentified Different vowel 2 as the Reference vowel. We used this criterion to define the excluded subject group: all the subjects that performed under 50% correct for any vowel in the Opposite condition were classified in the excluded group, and the other subjects in the non-excluded group.

For 3 of these subjects (Figure 10C subjects 6, 9 and 10), the percentage of correct answers to Different vowel 2 was lower than 15% in the Opposite condition but closer to chance in the Same condition. It hence seemed that they actively thought that Different vowel 2 was the Reference vowel in the easiest listening condition and pressed the button when it was presented, whereas they just pressed the button by chance in the more difficult listening condition.

Since these results were surprising, we nonetheless attempted to compute the sensitivity d’ for individual vowels. For the Different vowels, we defined the false alarm rate as any trial where the subject did not press the button in response to a presentation the Reference vowel. This is probably an over-estimation of the real false alarm rate for a single Different vowel. We observed that d’ was very seldom negative, even for the excluded subjects in response to the Different vowel 2 (Figure 11). This suggests that the subjects might in fact have performed at chance when the Different vowel 2 was presented and not actively perceived the Different vowel 2 as the Reference vowel.

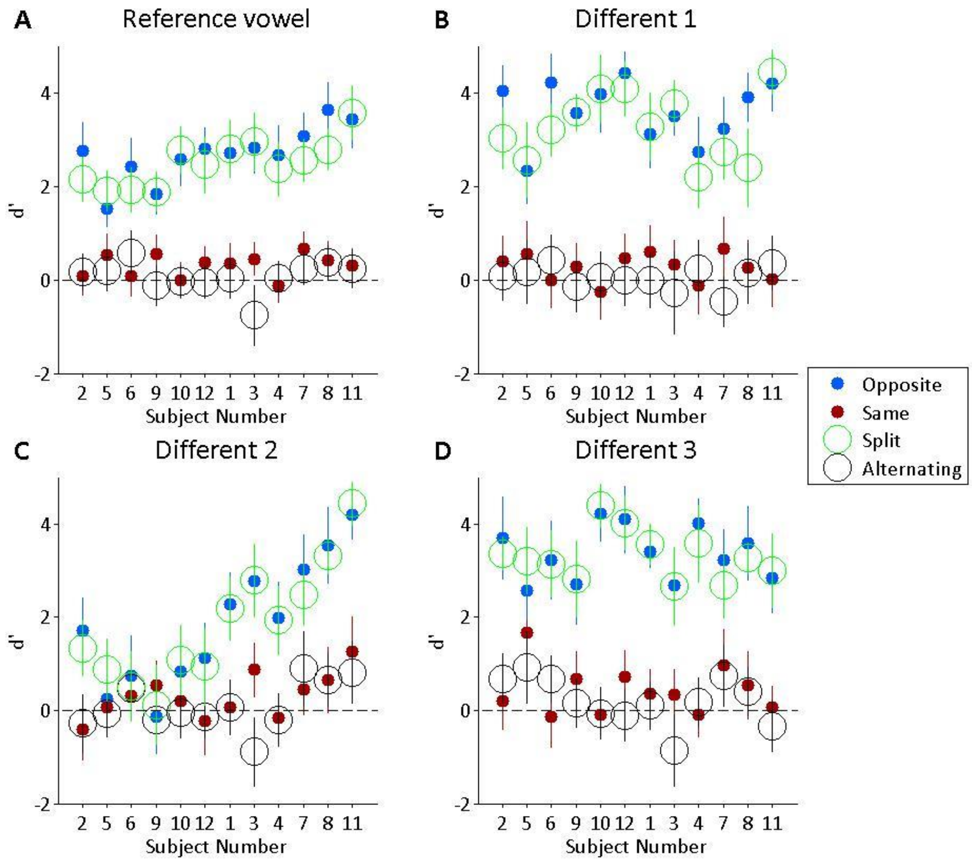

Figure 11: Same representation as Figure 10 using the sensitivity d’ as a measure of performance. **A.** For the Reference vowel, a hit was defined as a button press after the Reference vowel was presented and a false alarm as a button press after any Different vowel was presented. **B to D.** For the Different vowels, a hit was defined as a lack of button press when the Different vowel was presented and a false alarm as a lack of button press when the Reference vowel was presented.

Different vowel 2 had the same first formant as the Reference vowel of a center frequency of 630Hz. The second formants only were different, with a center frequency of 1230Hz for the Reference vowel and of 1110Hz on average for Different vowel 2. These subjects hence seemed unable to do vowel discrimination based only on the second formant. It could be due to a bad acuity to frequency differences in this frequency range, or to a problem perceiving spatial cues in the Opposite and Split conditions. The fact that their performance was different between the Opposite and Same conditions suggested that they were able to perceive the spatial cues to some extent, but could not discriminate between the Different vowel 2 and the Reference vowel.

Before starting the main experiment, the subjects had completed an adaptive procedure aimed at making all the Different vowels equally similar to the Reference vowel. At the end of this procedure, the formant frequencies of the Different vowels were set at frequencies closer or further from that of the Reference vowel depending on how well or poorly the subjects discriminated between the vowels. We observed that for 5 out of 6 subjects of the excluded group, Different vowel 2 was set to be as distinct as possible from the Reference vowel (Figure 12B, F2r – F2(d2) = 300Hz) after the adaptive procedure, whereas it was never set to such an 'easy' frequency difference for non-excluded subjects.

We observed that the subjects from the excluded group also performed worse for discriminating Different vowel 1 from the Reference vowel in the adaptive procedure as the mean difference between these vowels' first formants was set at 160+/-31.0Hz whereas it was set at a mean of 100+/-31.0Hz for the non-excluded group (Figure 12A). The two groups performed in a more similar way for Different vowel 3, with a mean distance of 140+/-49.0Hz for the excluded group and of 110+/-59Hz for the non-excluded group (Figure 12C). In the main experiment, it seemed that the subjects from the excluded group performed slightly worse in the Alternating and Opposite conditions for Different vowel 1 (Figure 10B).

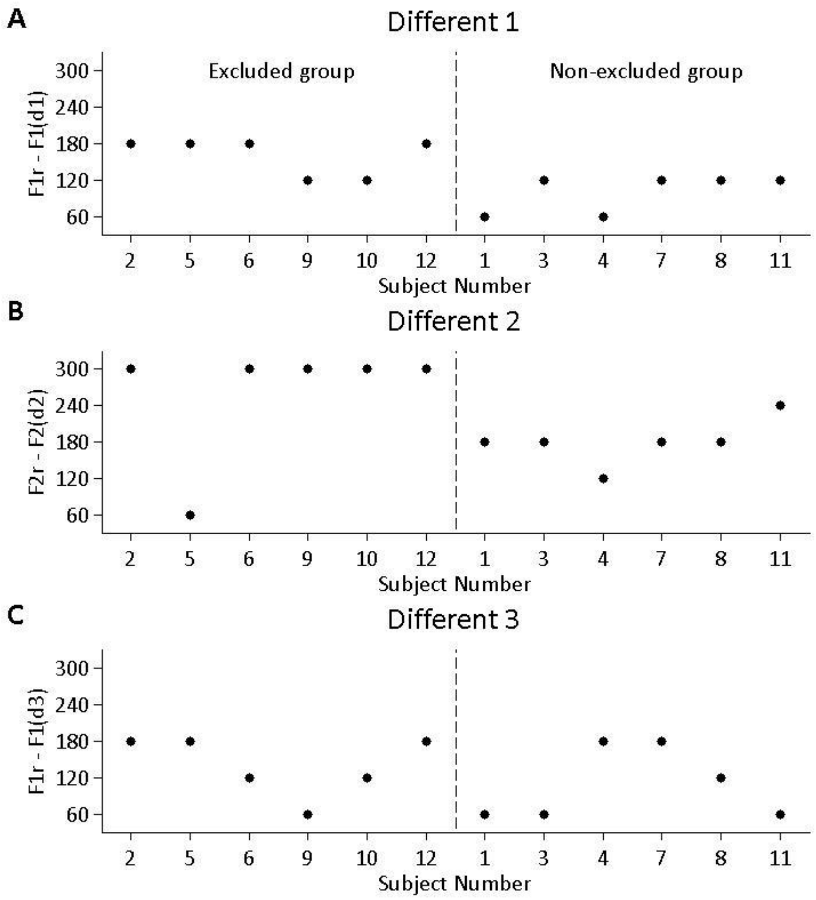

Figure 12: Results of the adaptive procedure that subjects did before the main experiment. The subjects from the excluded group are shown on the left (subjects 2, 5, 6, 9, 10 and 12) and the subjects from the non-excluded group on the right. **A.** Difference in Hz between the frequency of the first formant of the Reference vowel F1r and the frequency of the first formant of Different vowel 1 F1(d1). A smaller difference makes the two vowels more similar to each other and hence more difficult to discriminate. **B.** Same representation for the difference of the second formant frequency of the Reference vowel F2r and of Different vowel 2 F2(d2). **C.** Same representation for the difference of the first formant frequency of the Reference vowel and of Different vowel 3. For Different vowel 3, it was set that F1r – F1(d3) = F2r – F2(d3).

These results seem to indicate that subjects from the excluded group had more difficulty discriminating between vowels using only one formant. We took this problem into account during the experiment and made the main task easier for these subjects by lowering the intensity of the masker by 2dB. In retrospect, we saw that this adjustment was sufficient to equalize the ability of subjects in both the excluded and non-excluded group to discriminate between Different vowel 1 and the Reference vowel in the Opposite and Same condition, and to be very close in the Same and Alternating condition (Figure 10B). However, it was not sufficient to allow subjects in the excluded group to discriminate between Different vowel 2 and the Reference even in the easiest spatial condition (Figure 10C). We do not know whether lowering the intensity of the masker further would have changed their behavior to one that would match the other subjects' responses or whether their performance would have been at ceiling for all the vowels and conditions.

c. Exclusion of high frequency discrimination data for some subjects

Seeing that some subjects could not do the discrimination task in the Opposite condition for one of the Different vowels, we chose to do the analysis again excluding the data from these subjects. As explained in the methods section (II.2.e), we chose to exclude the data from Different vowel 2 in all conditions for these subjects, and to exclude the Split condition entirely from the analysis for these subjects.

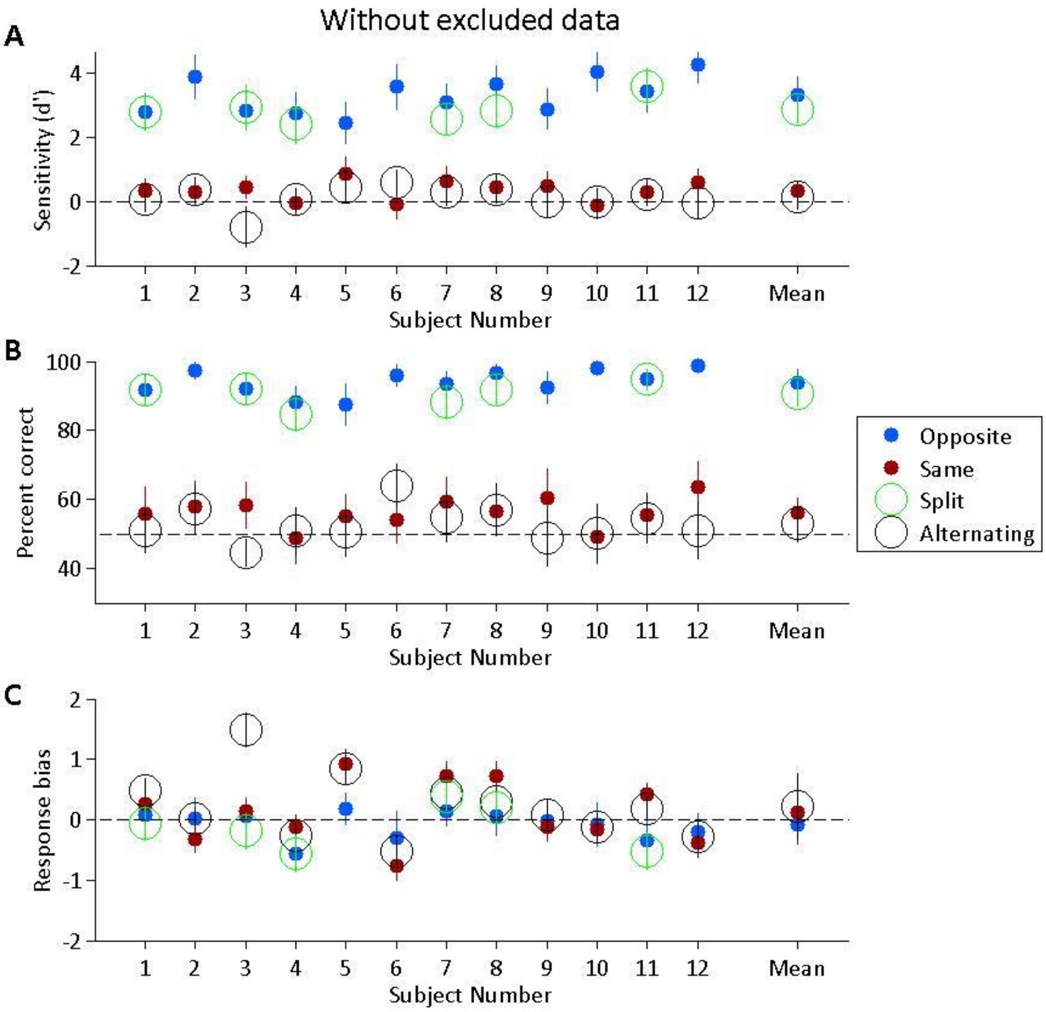

Figure 13: Same as Figure 9 without the data from Distractor 2 in all conditions and without the data from the Split condition for subject number 2, 5, 6, 9, 10 and 12. **A.** Sensitivity d'. **B.** Percentage of correct responses. **C.** Response bias.

Excluding these data (Figure 13) did not affect the results we observed in the previous section (Figure 9), so we concluded that our paradigm indeed showed that ITDs are processed independently across auditory filters and that the spatial unmasking effects are annulled if the ITDs of each sound are incoherent within one auditory filter.

d. Influence of the starting phase of the harmonics

In the Same and Opposite conditions, all the harmonics from each vowel start in phase with each other (Figure 6A). In the Split condition, the two harmonics of each formant start in phase with each other, but the two formants start at different phases. In the Alternating condition, the two harmonics of each formant start with different phases (Figure 6B). The fact that the harmonics begin out of phase in the Alternating condition might participate to making the vowel discrimination harder, so we wanted to test the effects of the starting phase on the discrimination performance. We tested 4 subjects with an additional Starting Phase condition where the vowels were presented from the right side of the head, the masker from the left side, and the harmonics of each formant started out of phase with each other (Figure 6C).

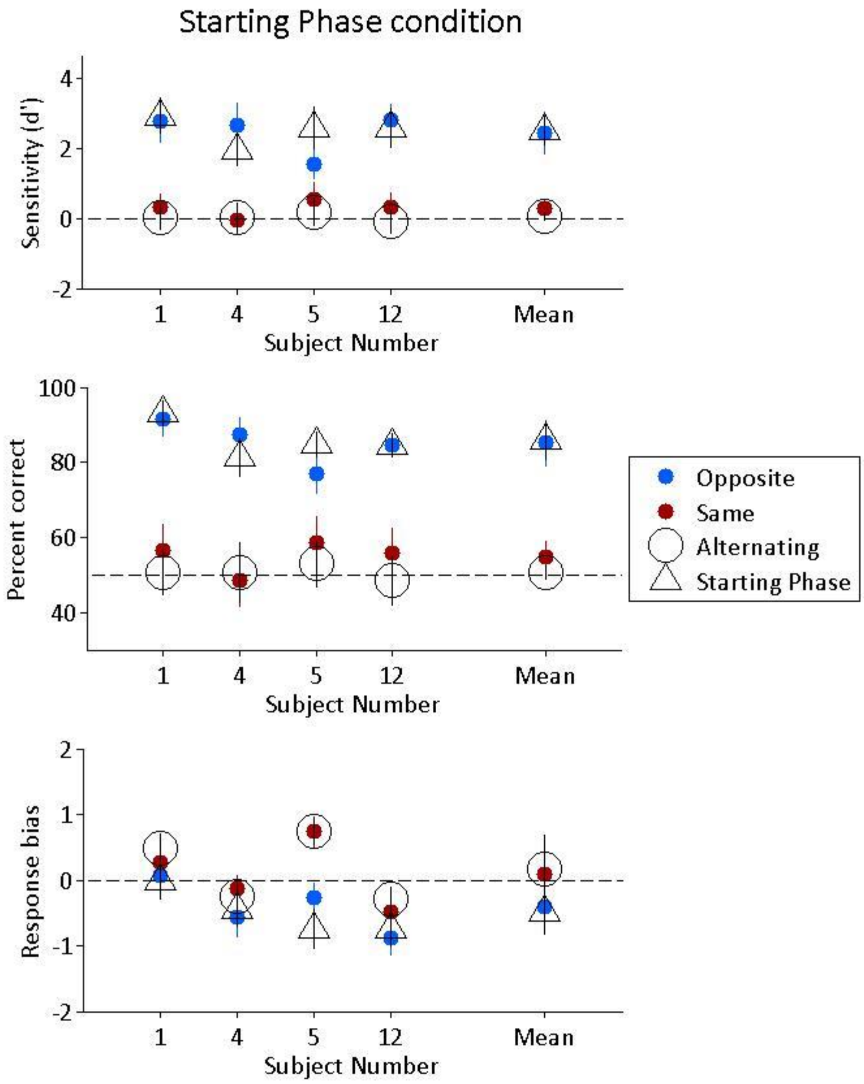

Figure 14: Same as Figure 9 for the Opposite, Same, Alternating and Starting Phase conditions, without the from Distractor 2 in all conditions for subjects 5 and 12 that were part of the excluded subject group. **A.** Sensitivity d'. **B.** Percentage of correct responses. **C.** Response bias.

We observed that the subjects' performance in the Starting Phase condition (mean d' over all subjects=2.5+/-0.4) was the same that in the Opposite condition (mean d'=2.5+/-0.6), suggesting that the phase relationship between the vowels' harmonics did not influence the vowel discrimination performance (Figure 14A). The performance in the Starting Phase condition was significantly higher than in the Alternating condition (mean d'=0.04+/-0.1) showing that applying a coherent ITD to the vowels while conserving the harmonics' phase relationships was sufficient to restore spatial unmasking.

As previously, we observed that the results using the percentage of correct answers (Figure 14B) were very similar to the results using a sensitivity measure (Figure 14A), and that the response biases were small and tended to be negative for conditions where the vowels were easy to discriminate and positive for more unfavorable conditions (Figure 14C).

This showed that disrupting relative starting phases of our vowels' harmonics did not impair binaural unmasking and that the lower performance in the Alternating condition was not due to relative starting phase differences.

4. Summary

We observed that each individual subject showed a better discrimination performance in the Opposite, Split and Starting Phase conditions than in the Same and Alternating conditions. The discrimination performance between the Reference vowel and the Different vowels was measured by the sensitivity d' and the percentage of correct responses which gave similar results. The subjects received minimal training and their performance was stable over time, showing a robust effect of our ITD manipulations. The task was conceptually easy but made difficult by the low signal to noise ratio between the vowels and the masker. The subjects reported detecting the vowels easily in the Opposite, Split and Starting Phase conditions but having great difficulty detecting them in the Same and Alternating conditions.

The presence of only four harmonics per vowel seemed sufficient to do this discrimination task. One subject (not shown here) performed the task with the full harmonic spectrum and their performance was similar. It is probable that given the high level of noise, only the few most intense harmonics are detected even if the full spectrum is present.

These results suggest that we indeed observed binaural level masking differences with our ITD manipulations. One remaining question is whether the second formant was really informative for the task. Indeed, we observed that 6 subjects were unable to discriminate the Different vowel 2 from the Reference vowel in all spatial conditions while their performance for the other Different vowels was the same as for other subjects. We observed that these subjects also performed worse during the adaptive procedure and seemed to have more difficulty discriminating vowels when only one of their formants was different. It is hard to say whether it was due to lack of sensitivity to binaural cues or to frequency differences, especially in the second formant frequency range. These two groups did not correlate with the amount of previous experience in psychophysical experiments of the subjects. Hence, it seems that some subjects were able to use the information contained in the second formant during the task and to discriminate vowels based on only one formant while others were not, and we could not find another explanation than individual differences.

In conclusion, we were able to replicate results showing the independence of ITD processing across frequency bands and the disruption of spatial unmasking when ITD cues are incoherent within one auditory filter using a simple synthetic vowel discrimination paradigm. We will probe the neural mechanisms involved in this discrimination task by recording from neurons in the inferior colliculus of anesthetized gerbils, and set the basis for further awake passive and awake behaving experimentation using this animal model.

III. Physiological experiment: how are ITD cues processed in the inferior colliculus?

1. Methods

All the surgical, recording, sound delivery and spike sorting methods were identical to those used by Garcia-Lazaro, Belliveau, and Lesica (2013) and are presented again here for reference.

a. In vivo recordings

Adult male gerbils (70–90 g, P60–P120) were anesthetized for surgery with an initial injection of a mix of ketamine, xylazine, and saline and the same solution was infused continuously during recording. A small metal rod was mounted on the skull and used to secure the head of the animal in a stereotaxic device in a sound-attenuated chamber. A craniotomy was made over the inferior colliculus, an incision was made in the dura mater, and a multi-tetrode array was inserted into the brain. Only recordings from the central nucleus of the IC were analyzed. Because the array covered a large area, recording sites in the central nucleus of the IC could be distinguished from those in other areas by comparison of their responses to tones (shape of the frequency response areas and best frequencies, Aitkin et al. 1975).

Sounds were delivered to speakers (Etymotic ER2) coupled to tubes that were positioned at the entrance of both ear canals for dichotic sound presentation along with microphones for calibration. The frequency response of the speakers measured at the entrance of the ear canal was flat (+/- 5dB SPL) between 0.2 and 5kHz.

b. Spike sorting

The procedure for isolation of single-unit spikes consisted of (1) band-pass filtering each channel between 500 and 5000Hz; (2) whitening each tetrode, i.e., projecting the signals from the four channels into a space in which they are uncorrelated; (3) identifying potential spikes as snippets with energy (Choi, Jung, and Kim 2006a) that exceeded a threshold (with a minimum of 0.7ms between potential spikes); (4) projecting each of the snippets into the space defined by the first three principal components for each channel; (5) identifying clusters of snippets within this space using KlustaKwik (http://klustakwik.sourceforge.net) and Klusters (Hazan, Zugaro, and Buzsáki 2006a) and (6) quantifying the likelihood that each cluster represented a single unit using isolation distance (Schmitzer-Torbert et al. 2005a). Isolation distance assumes that each cluster forms a multi-dimensional Gaussian cloud in feature space and measures, in terms of the SD of the original cluster, the increase in the size of the cluster required to double the number of snippets within it. The number of snippets in the “noise” cluster (non-isolated multiunit activity) for each tetrode was always at least as large as the number of spikes in any single-unit cluster. Only single-unit clusters with an isolation distance <20 were analyzed.

c. Stimuli

We used the same sounds as in our psychophysical experiment: vowels composed of 2 formants that each contained only 2 harmonics were presented with a masker in five ITD configurations (Figure 5B). Like before, each trial consisted of 750ms of masker alone, 250ms of masker with the Reference vowel, 350ms of masker alone, 250ms of masker with the Reference vowel or one of the three Different vowels and 350ms of masker alone. The response pause between trials was reduced to 1s. To limit recording time, the Reference vowel was presented the same number of time as each Different vowel.

The formant frequencies were the same for every animal: F1r=630Hz and F2r=1230Hz for the Reference vowel, F1=510Hz and F2=F2r for Different 1, F1=F1r and F2=1110Hz for Different 2 and F1=510Hz, F2=1050Hz for Different 3. The experiment was conducted with the vowels played at 70dB SPL and the masker played with 5dB (SNR=-5dB) and 14dB (SNR=-14dB) more intensity than the vowels. The ITDs ranged between -160µs and +160µs so that the sounds would be perceived by the gerbil at 90° left and right from the head’s midline. The ITDs are smaller than for humans because gerbils have a smaller head size and hence less distance between their ears. We will use the same convention that sounds coming from the right of the head have a positive ITD.

The same masker was used as in the psychophysics experiment but more snippets were chosen such that the noise presented during each trial was unique.

We recorded from the left IC which is mainly sensitive to sounds in the contralateral hemisphere and hence to sounds witch come from the right side of the head. We also wanted to obtain data from cells in the right IC, which would be more sensitive to sounds coming from the left side of the head. To be able to collect such data without complicating the surgical procedures, we recorded data from the left IC using the same stimulus paradigm but with the sounds previously presented to the left ear switched to the right ear and vice-versa. This is presumably equivalent to recording from the right IC using the exact same paradigm, and we will refer to this data set as coming from the right IC for more clarity.

d. Spike count decoding

For each trial, we want to predict whether the response of one neuron was elicited by the Reference vowel R or any Different vowel D1, D2, or D3. The response was represented by the number of spikes (see Figure 16) recorded from the onset to 50ms after the offset of the vowel presentation. We observed that the responses to the first and second presentation of the Reference vowel during each trial were the same (Figure 15), so we used the data from both indistinctively.

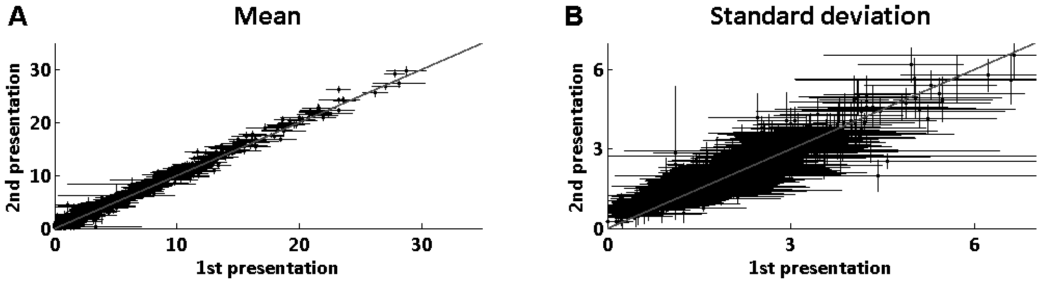

Figure 15: A. Mean and **B.** standard deviation of gaussian distributions truncated at 0 fitted to the spike count histograms of each cell in each ITD condition for the first and second presentation of the Reference vowel. Only cells that showed a significantly different spike count distribution in response to the target vowel than in response to 300ms of masker alone (Student t-test with unequal variance) were included.

The spike count histograms for each vowel were computed separately and fitted to Gaussian distributions truncated at 0 using a maximum likelihood estimator, giving the probability $p(r|s)$ of having a spike count *r* given that any vowel *s* was presented. The three distributions for the distractor vowels were then summed and normalized to have a total probability of one. This gives a measure of the probability $p(r|s \in \{D1, D2, D3\})$ of having a spike count *r* given that any Different vowel was presented.

A spike count *r* can be predicted as being elicited by the presentation of the Reference vowel if $p(s = R|r) > p(s \in \{D1, D2, D3\}|r)$. Bayes rules states that

$$p(s|r) = \frac{p(r|s)\,p(s)}{p(r)}$$

For our responses, $p(s|r) \propto p(r|s)$ because the total number of Reference and Different vowel presentations were equal. Each response was thus predicted as coming from the condition yielding the highest probability $p(r|s)$. In order to avoid over fitting, the probabilities were computed on all the trials excluding the one being decoded.

To compare the results of this decoding procedure to those that would have been obtained by chance, we repeated the decoding procedure 20 times while shuffling the pairings between the stimuli and the responses. This means that each spike count that was measured was randomly assigned to one of the 4 vowels, and a decoding performance was computed using these pairings. We considered a cell as having a significant decoding performance when its real decoding performance was equal or superior to the average of the 20 shuffled decoding performances + 3 * their standard deviation.

To decode the activity of a cell population, we compared $p(r_1, r_2 \dots r_n|s = R)$ and $p(r_1, r_2 \dots r_n| s \in \{D1, D2, D3\})$, with $r_1, r_2 \dots r_n$ the spike counts for n cells. We assume the cells are conditionally independent because there are no noise correlations in the IC (Garcia-Lazaro, Belliveau, and Lesica 2013a), so $p(r_1, r_2 \dots r_n| s) = p(r_1| s) * p(r_2| s) * \dots * p(r_n| s)$. The population response was thus predicted to come from the condition with the highest log likelihood $L_s = \log(p(r_1| s)) + \log(p(r_2| s)) + \cdots + \log(p(r_n| s))$.

e. Tuning curve measurement and significance

Frequency tuning of the cells was assessed by measuring a frequency response area (FRA) by playing a stream of 75ms long pure tones with a 5ms cosine ramp separated by 75ms of silence, of frequencies ranging from 300Hz to 8.2kHz and intensity ranging from 16 to 80dB SPL. The ITD tuning of cells was assessed by measuring a noise ITD tuning curve by playing a stream of 75ms long frozen noise snippets with a 5ms cosine ramp separated by 75ms of silence, with ITDs ranging from -1000µs to 1000µs.

The significance of frequency and ITD tuning curves was assessed by a spike count decoding method similar to that used for vowel identity decoding: spike count histograms were computed for each frequency (including all the intensity levels) or ITD; Gaussian distributions truncated at 0 were fitted to the histogram using a maximum likelihood estimate and spike counts were predicted to come from the frequency or ITD yielding the highest $p(r|s)$. The percentage of correct predictions was computed, and its variability was estimated by repeating the procedure 100 times with spike counts chosen randomly with replacement. The shuffled percentage of correct predictions was also computed using the same trials shuffled across frequencies or ITDs. The tuning curve was deemed significant if the 95% confidence intervals for the normal and shuffled percentages correct did not overlap. This inclusive criterion selected cells that had frequency or ITD tuning without rejecting cells with weak tuning.

2. Results

a. Paradigm and hypothesis

We recorded from single units in the anesthetized gerbil inferior colliculus (IC). We were interested in the neuronal coding of low frequency speech-like sounds and hence concentrated on the dorsal low frequency part of the central IC. Like in the previous chapter, we played vowel-like sounds and masking sounds with different spatial configurations defined only by the ITDs applied to the sounds. We will use the same convention that sounds that come from the right side of the head have a positive ITD, and sounds coming from the left side a negative ITD.

We first recorded from the left IC with the same stimulus we used for the psychophysical experiment, with the noise on the left side and the vowels on the right side of the head. This creates an asymmetry since the majority of cells in the left IC have a best ITD corresponding to locations on the right of the head (Figure 18B). We then recorded from the left IC again with the same stimulus except that the sounds coming from the right now came from the left and vice-versa. We know that there are no noise correlations in the IC (Garcia-Lazaro, Belliveau, and Lesica 2013a) which means that the activity from each cell is independent from that of the other cells. The responses of the cells might be correlated because they are presented with the same stimulus (signal correlations) but not for other reasons (no noise correlations). It is hence equivalent to record from different cells with the same stimulus at the same point in time or sequentially. We can thus analyse cells that we recorded in sequential recordings using the same stimulus as being part of the same population. This allowed us to use the second recording with the stimulus reversed between the ears as a model for the activity in the right IC that could have been observed at the same time as the initial recording in the left IC. We will refer to this second recording as being from the right IC for simplicity.

We always presented the vowels at a fixed intensity, and the masker was presented at two different intensities. The first condition was SNR=-14dB where the masker was 14dB more intense than the vowels, like in the psychophysical experiment. The second conditions was SNR=-5dB where the masker was 5dB more intense than the vowels. With this second easier condition, we hoped to see more clearly the neuronal responses to the vowels. We hoped to observe clearer single cell activity, and might be able to correlate better the population activity to the psychophysical results. Indeed, we recorded from only a few hundreds of cells from the IC whereas an animal would have access to the hundreds of thousands of cells composing both ICs to resolve the task. Moreover, we didn’t measure the gerbils’ behavioral thresholds for these behavioral tasks and they might not be able to perform the task at the same low SNR as humans.

Our goal was to compare how well we could recognize which vowels were presented to the gerbils from analysing their neuronal activity (neuronal decoding performance) to how well the human subjects in our psychophysical experiment recognized the vowels (psychophysical performance) in different spatial conditions. To be able to recognize a vowel from the neural activity, we hypothesized that the neurons would need a minimal firing rate. If the neurons did not respond strongly to our stimulus, there might not have been enough information contained in the spike trains to recognize the different vowels. Since our study is focused on the differences in performance between several spatial conditions, the neurons which performances are closest to the human psychophysics are likely to be ITD tuned. Indeed, if they are not ITD tuned and respond in the same way to sounds at all ITDs, it is likely that they will have the same decoding performance in all the spatial conditions. Our stimulus is asymmetrical, with the vowels presented on the right side of the head in the Opposite and Same conditions. We hence hypothesized that neurons that preferred sounds coming from the right side (i.e. tuned to positive ITDs) would have an overall better decoding performance and might show more contrast between spatial conditions.

Finally, we hypothesized that cells should be frequency tuned to be able to discriminate between the vowels. Indeed, the vowels vary only by the center frequency of their formants, so it is likely that if the formants from some of the vowels but not others are within one cell’s receptive field this cell will be able to discriminate between them. This argument might be especially relevant if the cell’s best frequency is close to the first or second formant frequency such that there might be more contrast between the responses to different vowels.

In conclusion, we hypothesized that cells with a minimal firing rate, ITD tuned and maybe more specifically ITD tuned to the right side of the head and frequency tuned and maybe frequency tuned to one or both formant frequencies would have the best decoding performance in the Opposite condition and follow the psychophysical trends.

b. Vowel identity is encoded by spike rate

We used the Victor metric (Jonathan D. Victor and Purpura 1997) to assess whether vowel identity was encoded by spike rate or spike timing, and what was the relevant time scale for vowel discrimination. This metric allows us to determine the similarity between two spike trains by computing the cost of adding, deleting or shifting spikes to transform one spike train into the other. The parameter τ represents the distance (in seconds) between two spikes over which it costs less to delete and add the spike than to shift it. The distances between one spike train and all the others are computed and we predict that the spike train comes from the condition where the distances are minimal. Here, we defined one spike train as the spikes elicited during the presentation of one vowel, and compared each spike train to the spike trains elicited by the presentations of all the vowels. Hence the value of τ where the best decoding performance is achieved represents the spike timing precision that yields the most information about vowel identity.

We applied this method to predict whether each vowel was the Reference vowel or any Different vowel for each cell. Figure 16A shows the decoding performance for one example cell. For this cell, we observed that performance increased with increasing τ until it reached a plateau for all spatial conditions, which suggests that the decoding relied on spike rate. For all the cell population, we computed the decoding performance for τ=512 which corresponds to a purely spike rate based metric because the spikes can be shifted around the whole spike train at no additional cost. We also computed the τ leading to the best decoding performance which corresponds to the optimal decoding time scale for each cell.

For the SNR=-5dB condition, we observed a mean of 3.5+/-14.3% improvement in decoding performance across all the spatial conditions when we used the best τ compared to τ=512 (Figure 16B). 92.7% of the cells showed an improvement in decoding performance less or equal to 10% when considering the best τ instead of τ=512 (Figure 16C). For the SNR=-14dB condition, we observed a mean of 4.3+/-17.9% improvement over all conditions, and 90.8% of cells showed less than 10% improvement in decoding performance when considering the best τ instead of τ=512 .

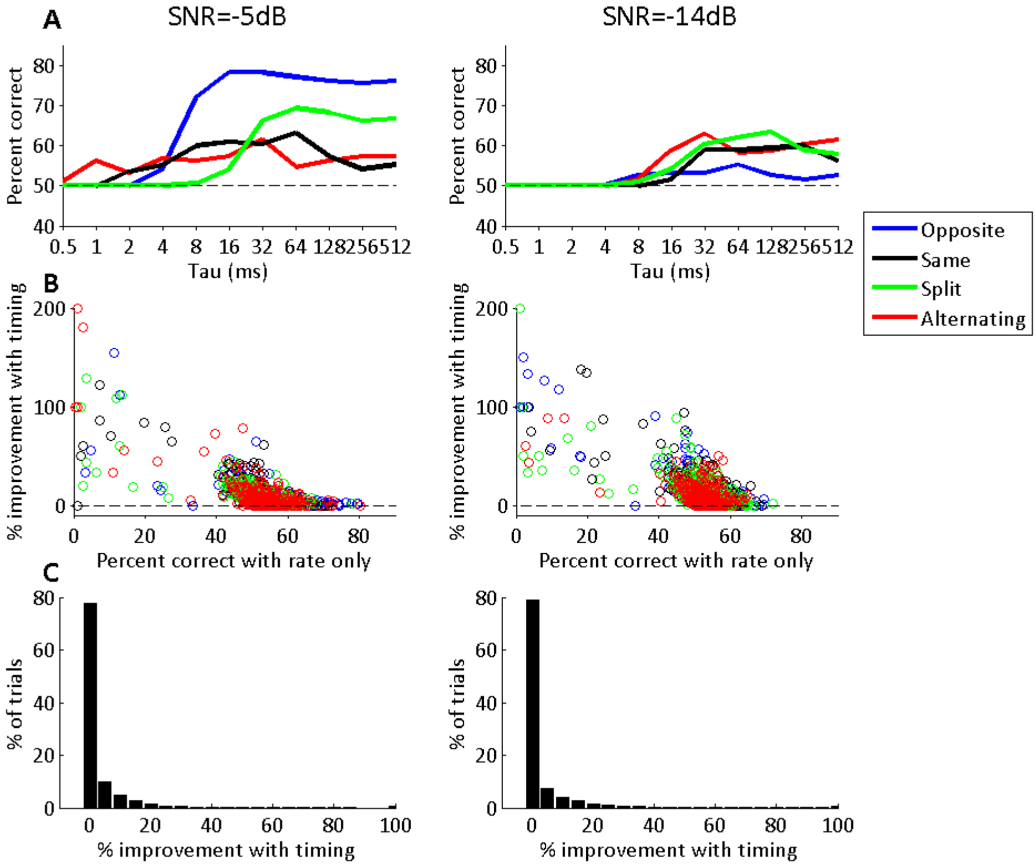

Figure 16: Decoding with the Victor metric for SNR=-5dB (left column) and SNR=-14dB (right column). **A.** Decoding performance in the 4 spatial conditions for one example cell. **B.** Percentage improvement of decoding performance when using the best τ compared to τ=512ms. Each dot represents one cell in one spatial condition. **C.** Histogram of the percentage of improvement of decoding performance when using spike timing instead of spike rate only. Same data as in B pooled across spatial conditions, each trial is the performance of one cell in one spatial condition.

It hence seemed that considering spike timing additionally to spike rate did not yield much improvement in the decoding performance, and we chose to neglect the information contained in spike timing. For the rest of the analysis, we represented each trial by the number of spikes elicited during each vowel presentation. We will call this number of spikes the spike count measured for each vowel.

c. Single cell performance in the Opposite condition

We wanted to assess the performance of our cells in a vowel discrimination task and tried to recognize which vowel was presented to the animal by looking at its neuronal responses. For each cell, we observed that each vowel could elicit a range of spike counts that could be modelled by a normal distribution truncated at 0. We computed the spike count distributions for the Reference vowels and for the three Distractor vowels. We predicted whether each spike count was elicited by the Reference or a Distractor vowel by looking at the distribution where it appeared the most times. Each cell's performance was defined as the percentage of correct predictions of the vowel that elicited each spike count (see III.1.d. for full method).

We hypothesized that every cell that could take part in the vowel decoding task would have a decoding performance significantly better than chance in the easiest spatial condition (the Opposite condition where the vowels are presented from the right side of the head and the masker on the left side). To define chance performance, we measured each cell's decoding performance when the pairings between the vowels we presented and spike counts we recorded were randomly defined. We repeated this measure 20 times with different random pairings to have a measure of variability. We defined better than chance performance as the real decoding performance being 3 standard deviations higher than the mean chance performance.

We considered cells recorded in both the right and left IC and observed that 35.7% of cells recorded in the SNR=-5dB condition and 18.0% of cells recorded in the SNR=-14dB condition decoded the vowels in the Opposite condition significantly better than chance (Figure 17). We will call the population of cells that perform significantly better than chance the decoding population.

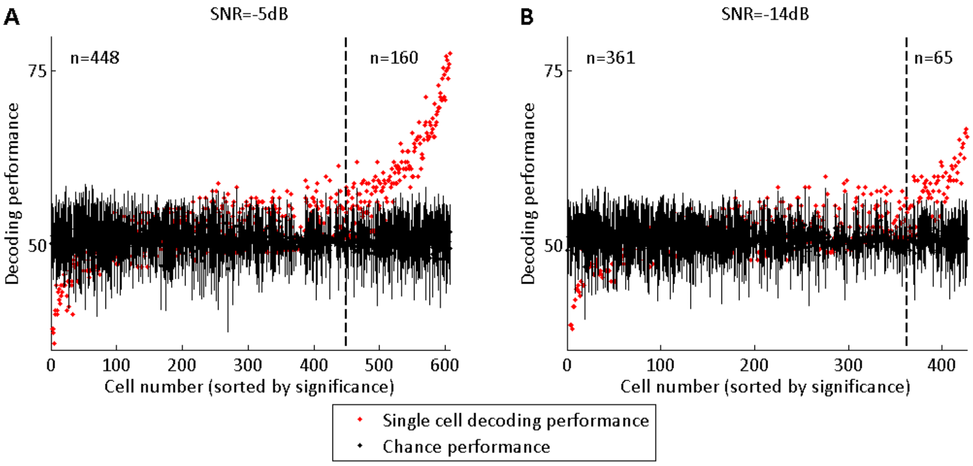

Figure 17: Single cell vowel decoding performance in the Opposite condition for cells recorded in the right and left IC. **A.** Real decoding performance in the SNR=-5dB condition (red) and mean +/-2SD of the chance performance measured by randomizing 20 times the pairings between the vowels and the observed spike counts (black). **B.** Same representation for the SNR=-14dB condition.

d. Cell population

We characterized the basic properties of our cell population by measuring frequency response areas (FRAs) and ITD tuning curves. The FRAs were obtained by measuring the cells' spike rates in response to pure tones at a range of frequencies and intensities. The ITD tuning curves were obtained by measuring their spike rate for frozen noise snippets at different ITDs.

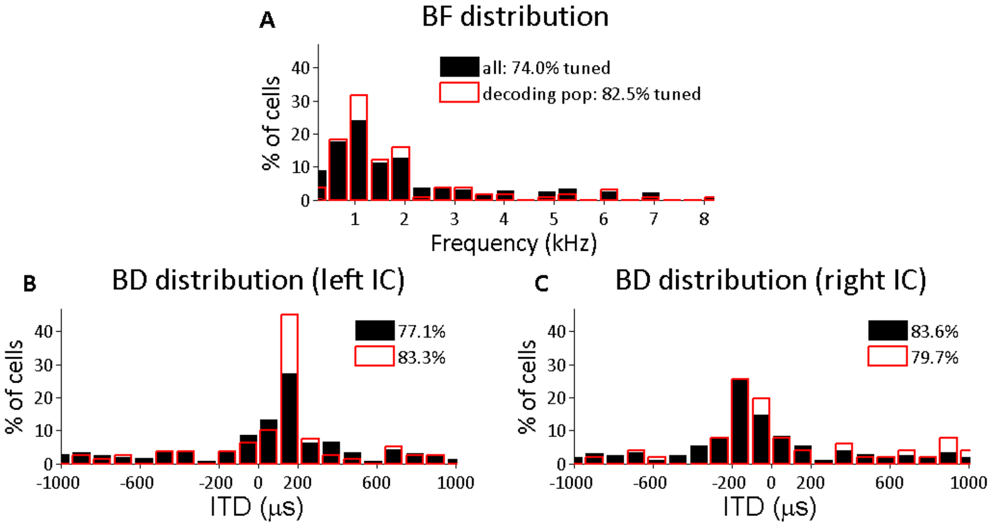

Figure 18: A. Distribution of best frequencies (BFs) for single units with significant frequency tuning in the whole population (full black bars) or in the population that showed a significant vowel decoding performance in the Opposite condition (empty red bars). The percentages in the legend indicate the percentage of singnificantly tuned cell in the population. BF was defined as the frequency eliciting the maximum spike rate over all sound intensities. **B.** Same representation for the best ITDs (BDs) recorded from the left IC. BD was defined as the ITD eliciting the maximum spike rate. **C.** Same representation for the best ITDs recorded from the right IC.

We observed that 74.6% of the whole cell population and 81.8% of the decoding population had best frequencies below 2kHz (Figure 18A), which is typical of the dorsal ICC (Aitkin et al. 1975a). The vowels from our stimulus are in this frequency range, so these cells were probably well suited to the vowel decoding task. 68.0% of the whole population and 68.5% of the decoding population had a contralateral ITD tuning (Figure 18B and C), which is also typical for the IC. There were between 25 and 45% of cells with a preferred ITD close to 160µs or -160µs, which corresponds to a sound location at 90° to the right or left of the head's midline (Figure 18B and C). These cells should hence respond in a different way to stimuli presented on the right and on the left of the head and might be able to benefit from spatial unmasking.

For all further analysis we chose to include only the cells from the decoding population. Indeed, if the non-significant cells did not show a better than chance decoding performance in the fully unmasked spatial condition, it is very unlikely that they participate in the decoding performance of the population for other more complex spatial conditions.

To assess how cells responded to our quite complex sound stimulus, we observed the responses from a few individual cells before going on to a systematic analysis of the whole population. We observed that cells with a positive best ITD (for example Figure 19Ab) were strongly driven by the vowels in the Opposite condition (Figure 19Ac) and by the vowels and the masker in the Same condition (Figure 19Ae). For this cell, we observed that the spike rates elicited by the four different vowels were more diverse in the Opposite condition (Figure 19Ad) in which the vowels and the masker were presented from opposite sides of the head than in the Same condition (Figure 19Af) in which the vowels and the masker were both presented from the right side. We observed that the single cell decoding performance was higher in the Opposite condition than in the Same condition (Figure 19Ag). For this example cell, we could hence observe spatial unmasking on a single cell level, which might be mediated by a higher diversity in spike rates in response to different vowels in the Opposite condition.

We observed large diversity of cell responses and properties in our population. For example, some cells had weak ITD tuning (Figure 19Bb), more similar responses to the vowels and the masker (Figure 19Be) and less diversity in the spike rates elicited by the different vowels (Figure 19Bd and f). Other cells preferred sounds coming from the left side of the head (i.e. negative best ITD or ipsilateral ITD tuning, Figure 19Cb) and very strong responses to the masker in the Opposite and Same conditions (Figure 19Cc and e). Interestingly, the vowels in the Opposite condition inhibited the responses to the masker for some cells (Figure 19Cc). The responses to the different vowels also tended to be less diverse for this type of cell (Figure 19Cd and f), and the decoding performance was lower.

From these examples, we observed that the cell with a high firing rate, strong ITD tuning to the right side and a low best frequency (Figure 19A) had the most diverse responses to the different vowels in the Opposite condition, the best vowel decoding performance in the Opposite condition and showed spatial unmasking between the Same and Opposite conditions. These results obtained for 3 example cells recorded at SNR=-5dB were coherent with our hypotheses and we will go on to analyse the whole cell population systematically.

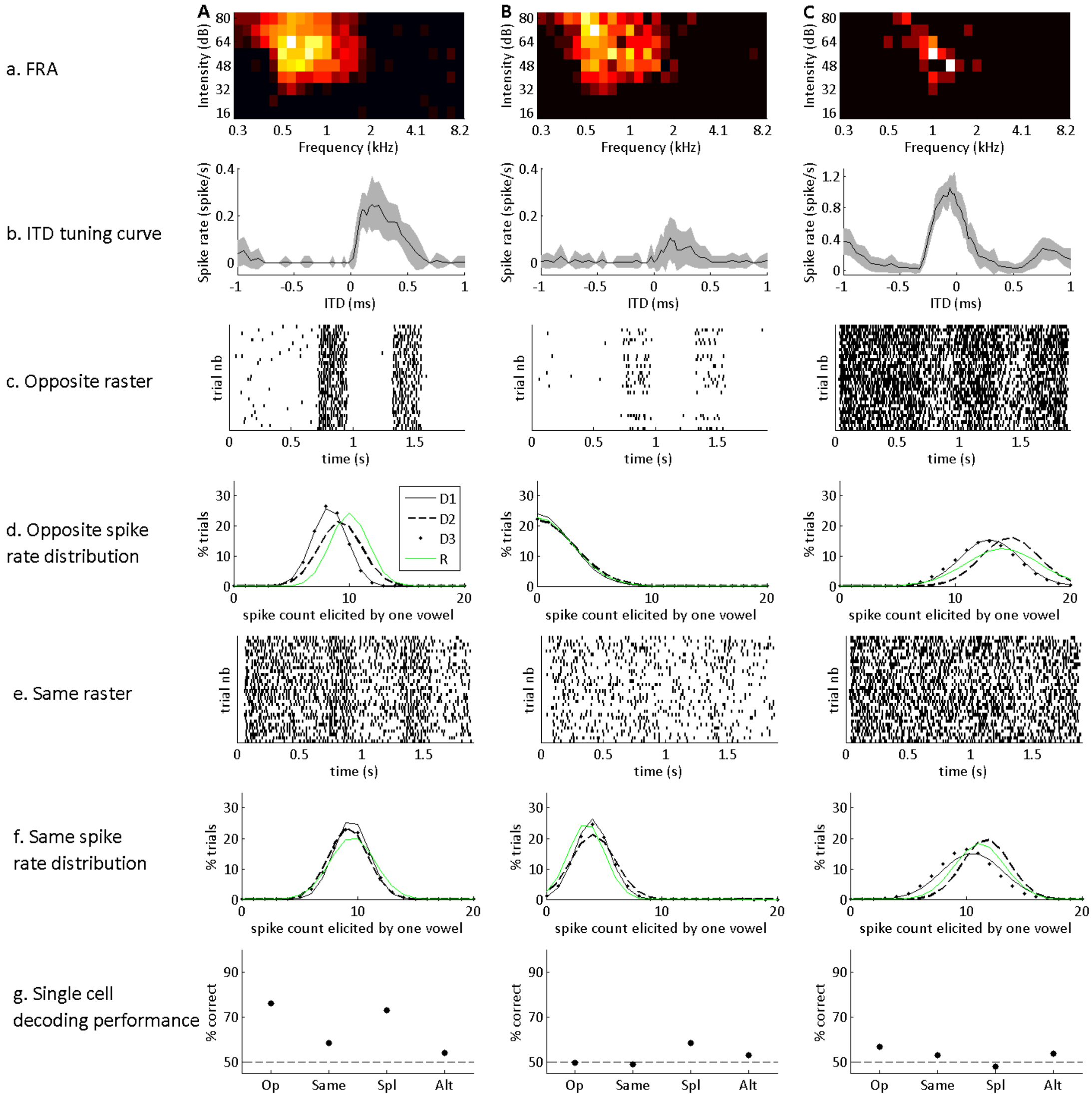

Figure 19: Receptive fields, responses and decoding performances of 3 example cells from the left IC in the SNR=-5dB condition. **A.** Typical dorsal ICC cell with low best frequency and contralateral best ITD. **B.** Example cell with less ITD sensitivity. **C.** Example cell with ipsilateral best ITD. **a.** Pure tone FRAs. **b.** Noise ITD tuning curves. **c.** Rasters of the full stimulus (masker only – Reference vowel – masker only – Reference or Different vowel – masker only) in the Opposite condition. **d.** Gaussian distributions fitted to the spike count histograms for each vowel separately in the Opposite condition. **e.** Raster (same as c.) in the Same condition. **f.** Spike count distributions (same as d.) in the Same condition. **g.** Single cell vowel decoding performance in the 4 spatial conditions.

In the SNR=-14dB condition, the vowels were presented at the same intensity and the masker 9dB more intense than for SNR=-5dB. We observed less difference in firing rate during the vowel presentation compared to when the masker only was presented (Figure 20c and e). In the Opposite condition (Figure 20c) the cells might be inhibited by the intense masker coming from the left side of the head at non-preferred ITDs. Indeed, these cells had a firing rate very close to 0 at negative ITDs, and sounds coming from the left side of the head might inhibit their activity. In the Same condition (Figure 20e) the cells might already have reached the peak of their rate level function with the masker only, and fire the same amount or less spikes when the vowels are added.

For these three example cells, the spike rate distributions were very similar for all the vowels in the Opposite and Same conditions (Figure 20d and f) and this was correlated with a poor vowel decoding performance (Figure 20g). From these examples, it seems that the SNR=-14dB condition does not allow us to see single cell responses to the different vowels or to decode the vowels based on single cell activity. However, we will also study systematically this condition for all the population as analysing the activity of several cells at the same time might allow us to obtain different results.

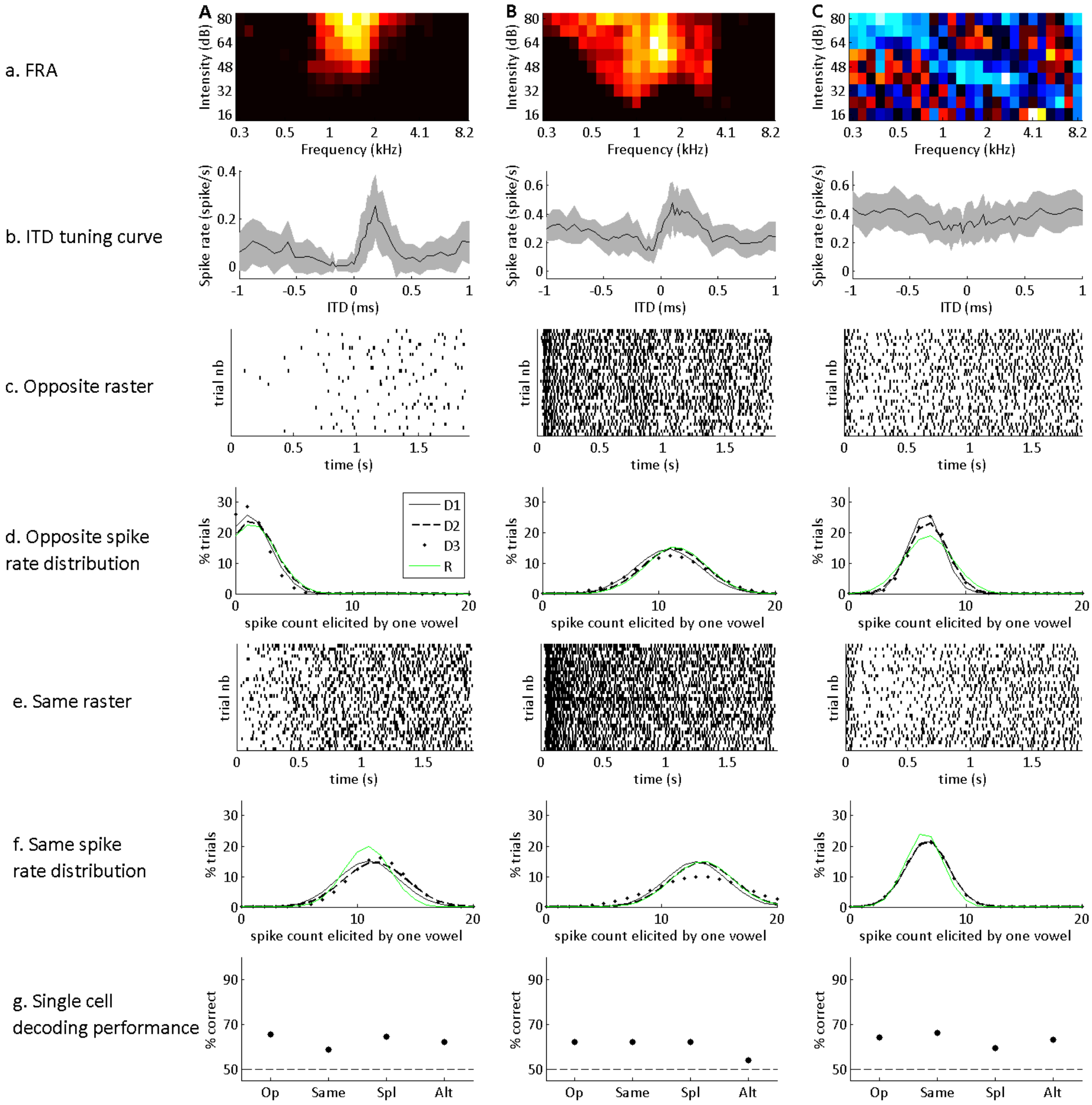

Figure 20: Receptive fields, responses and decoding performances of 3 example cells from the left IC in the SNR=-14dB condition, represented the same way as in Figure 19. The cell presented in Figure 19A is the same as Figure 20A, and the other cells are different.

e. Influence of firing rate on the decoding performance

We observed a large diversity of firing rates in our decoding population, with half the population firing less than 11.1 spikes per trial over all conditions (Figure 21), which corresponds to a firing rate of 5.7spikes/s. We always considered the firing rate during the whole trial, which included the responses to the masker and the responses to the vowels. The question of whether this firing rate represented the response rate to the vowels or the spontaneous firing rate of the cells arose. We did not collect data on the spontaneous firing rate of the cells because the masker was presented during the whole trial with no interruption. We cannot consider the responses to the masker only as spontaneous activity as the masker was presented at different ITDs depending on the spatial condition so the responses of the cells to the masker depended on the spatial condition and on their ITD tuning. We hence chose to work with the firing rates in response to the whole trial as the most comprehensive measure of each cell's evoked firing rate to the whole stimulus. This measure most likely varies with the spontaneous firing rate of the cells, but we were not able to differentiate between these two factors.

We hypothesised that a spike rate about a minimal value would be required for decoding as it would allow the cells to have different responses to the different vowels and convey information about vowel identity. This hypothesis seemed to be verified for some single cells at SNR=-5dB (Figure 19) but not at SNR=-14dB (Figure 20).

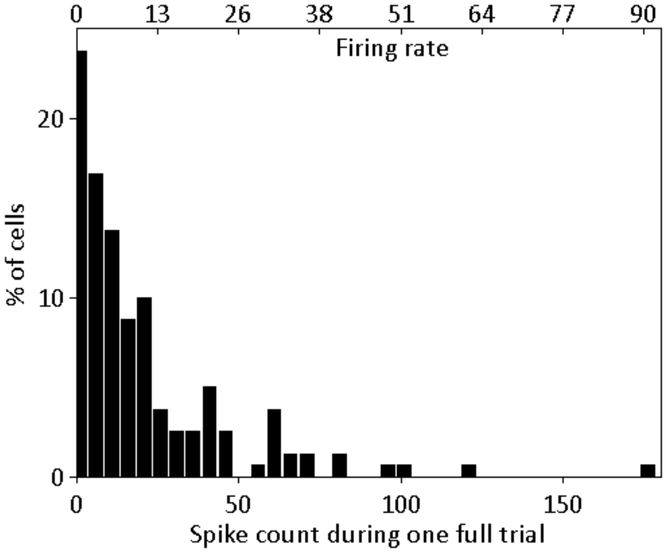

Figure 21: Distribution of the firing rates computed over one full trial (presentation of 2 vowels and of the masker lasting 1.95s). The bottom axis is the number of spikes each group of cells fired during one trial, the top axis is the corresponding firing rate value in spikes per second.

We divided our decoding population in 4 groups containing an equal number of cells sorted by firing rate and assessed their vowel decoding performance. To assess the decoding performance of a group of cells, we used two methods:

- Single cell average decoding: we computed the decoding performance of each single cell, and expressed the population decoding performance as the average of the single cell decoding performances (Figure 22a).
- Population decoding: we computed the decoding performance of the whole population at once, the response of the population being the group of spike counts elicited by each cell for one vowel (Figure 22b,

see III.1.d for full methods). To have a measure of variability, we repeated the measure 20 times with cells chosen randomly with replacement within the population of interest.

We observed that the single cell decoding performances for all spatial conditions increased with the firing rate for both SNR=-5dB (Figure 22a,b) and SNR=-14dB (Figure 23a,b). The cells with higher firing rates showed a larger difference between their chance decoding performance and real decoding performance in the Opposite condition than the cells with low firing rates (Figure 22c and Figure 23c). We hence concluded that the cells not only needed a minimal firing rate to perform vowel decoding but that a higher firing rate was correlated with a better overall decoding performance.

We tested whether the decoding performances in different spatial conditions were different using an ANOVA corrected for repeated measures followed by pairwise comparisons if there was a significant difference overall. We showed here only the pairwise results between the Opposite and Same, Opposite and Split and Opposite and Alternating conditions. From the psychophysical experiment, we expected the performance in the Opposite and Same and Opposite and Alternating conditions to be significantly different, while the performance in the Opposite and Split condition should be the same. We will say that a group of cells ‘follows the psychophysical trends’ if these three pairwise comparisons give the same results as in the psychophysical experiment.

For SNR=-5dB, we observed that all the groups that fired below 25 spikes per trial responded differently to each stimulus, and hence showed spatial unmasking between the Same and the Opposite condition, but did not follow the psychophysical trends (Figure 22A,B,C,a,b). The cells that fired more than 25 spikes per trials followed the psychophysical trends (Figure 22Da) when their performance was computed as the average single cell performance. This confirmed our hypothesis that a higher firing rate conveys more information about vowel identity, and we observed that only the cells that fired the most followed the psychophysical trends.

We noted that decoding each group of cells using population decoding yielded an overall higher decoding performance than using single cell average decoding but both methods showed the same differences between spatial conditions.

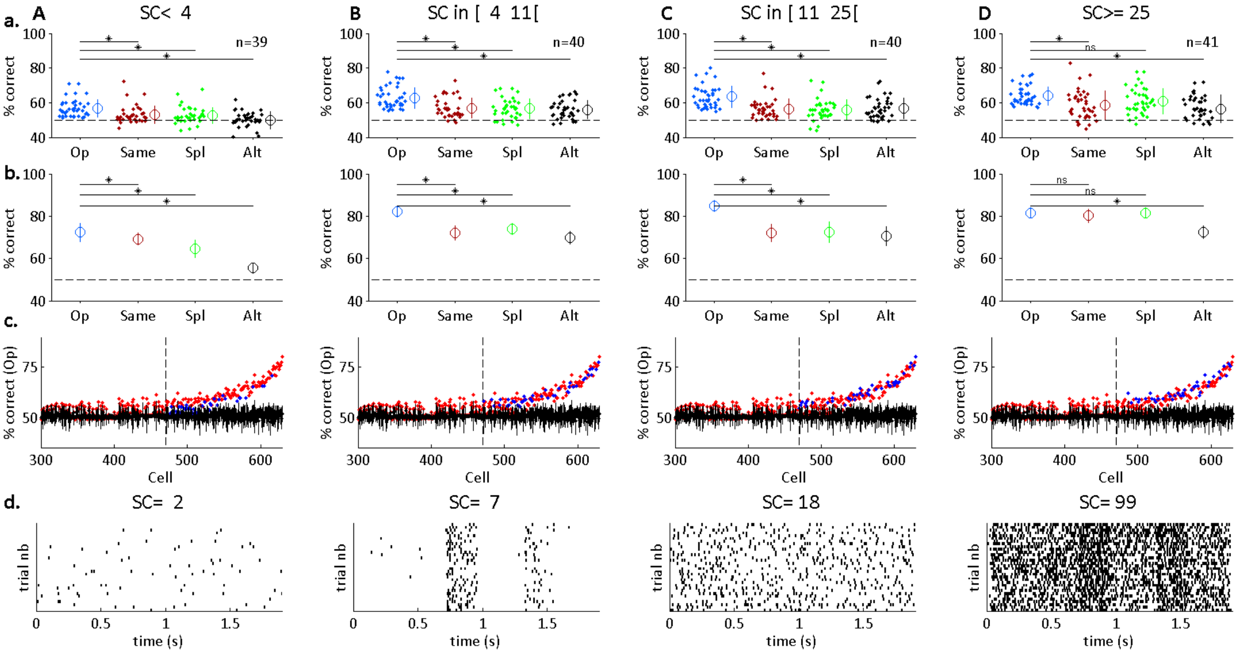

Figure 22: Decoding performance for groups of cells recorded in the SNR=-5dB condition that spike **A.** less than 7 spikes per trial, **B.** between 7 and 18 spikes per trial, **C.** between 18 and 37 spikes per trial and **D.** more than 37 spikes per trial. **a.** Single cell decoding performance for each cell of the group (dots) and their average +/-sd (circles with error bars) for each spatial condition. '*' represents a significant pairwise difference as measured by at-test after an ANOVA corrected for repeated measures; 'ns' represents a non-significant difference for that same test. **b.** Population decoding performance (average of 20 measures with populations chosen randomly with replacement within the group +/-sd) for each spatial condition. **c.** Same representation as Figure 17, with the cells of each group shown in blue: single cell decoding performance in the Opposite condition for the whole population (red dots) and for the cells from the group (blue dots). Chance performance in the Opposite condition for the whole population (black). **d.** Rasters of one full trial for one example cell of the group.

In the SNR=-14dB condition, we also observed that a higher firing rate was correlated to an overall higher decoding performance when doing single cell average (Figure 23a) or population decoding (Figure 23b). It was also correlated to single cell performance in the Opposite condition being more different than chance performance (Figure 23c). We observed spatial unmasking between the Opposite and Same conditions for all the groups except one, but none of them followed the psychophysical trends. This once again confirmed our hypothesis that a high firing rate conveys more information about vowel identity, but indicated that the masker in the SNR=-14dB condition might be too intense to be able to observe psychophysical differences between the spatial conditions.

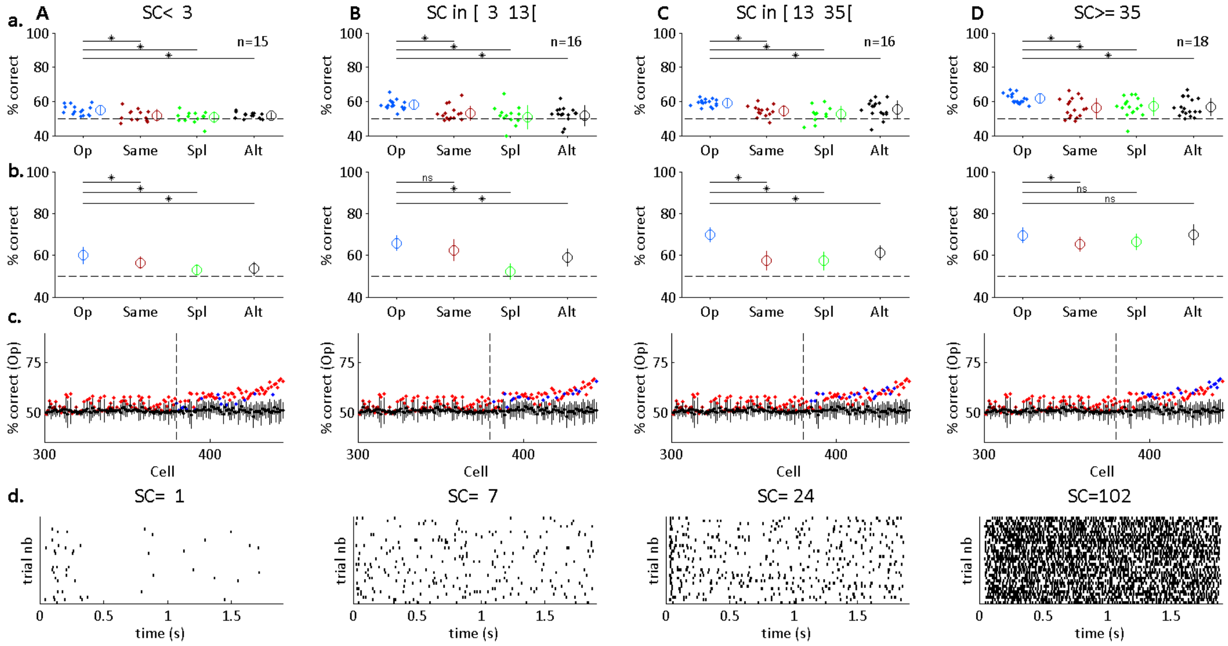

Figure 23: Same representation as Figure 22 for groups of cells recorded in the SNR=-14dB condition.

Overall, we observed that firing rate was an important factor for decoding performance. A higher firing rate over the whole trial correlated with increased decoding performance in all spatial conditions, both when measured as a single cell average and as a population performance. It also correlated to a decoding performance more different than chance in the Opposite condition. This allowed us to refine our hypothesis about firing rate: rather than needing a minimal firing rate to do the task, it seems that groups of cells with any firing rate can perform vowel decoding but that cells with a higher firing rate have a better performance. We observed spatial unmasking between the Same and Opposite conditions in all the groups at both SNR, indicating that spatial unmasking was a strong feature of this cell population. The only group that followed the psychophysical trends was the cells spiking more than 25 spikes per trial at SNR=-5dB.

f. Influence of ITD tuning on the decoding performance

In our stimulus, the only difference between spatial conditions was the ITD of the vowels and the noise. Hence, cells with different ITD tunings might respond in different ways to the stimulus. We hypothesised that ITD tuned cells would show more differences in their decoding performance between spatial conditions and that cells tuned to the right side of the head would follow the psychophysical trends best. We will first explain intuitively the importance of ITD tuning starting from an example, and then go on to a systematic analysis of our decoding population.

A typical cell from the left IC prefers positive ITDs (i.e. sounds coming from the right of the head), as was the case for our example cell (Figure 24B). This cell is sensitive to low frequencies and our vowels hence fell in its receptive field (Figure 24A). If we first consider the SNR=-5dB condition (Figure 24D), we observed that the cell spikes almost only in response to the vowels in the Opposite condition. Indeed, in this condition the vowels were presented at +160µs ITD where the ITD tuning curve shows a high firing rate while the masker was presented at -160µs ITD where the ITD tuning curve shows an absence of spiking. In the Same condition, both vowels and maskers were presented at +160µs ITD and we indeed observe a high firing rate throughout the trial, with a bit more firing during the vowel presentation. For the Split and Alternating conditions, the vowels and masker were distributed between + and -160µs ITD and we indeed observe an intermediate firing rate during the whole trial, with additional firing during the vowel presentation. We observed that the decoding performance was high in the Opposite and Split conditions and low in the Same and Alternating conditions (Figure 24C), which means the performance was better when at least one full formant of the vowels was at +160µs ITD. From this example, we can see that the way each cell responds to the vowel and noise may be predicted by its ITD tuning curve, and this gives us an intuition on the cell's vowel discrimination performance. This confirmed our hypothesis that a cell with a positive preferred ITD would show contrasted responses to the different spatial conditions and follow the psychophysical trends.

In the SNR=-14dB condition (Figure 24E), we observed a lower firing rate in the Opposite condition. This might be due to the more intense masker at -160µs ITD inhibiting the cell's response. In the Same, Split and Alternating conditions, the cell is very strongly driven by the masker. We can see that the cell seems to fire uniformly during the whole trial whether the vowels are present or not. It seems that these rasters do not contain a lot on information on the vowels, and the single cell decoding performance for this cell was indeed at chance for all spatial conditions (Figure 24C).

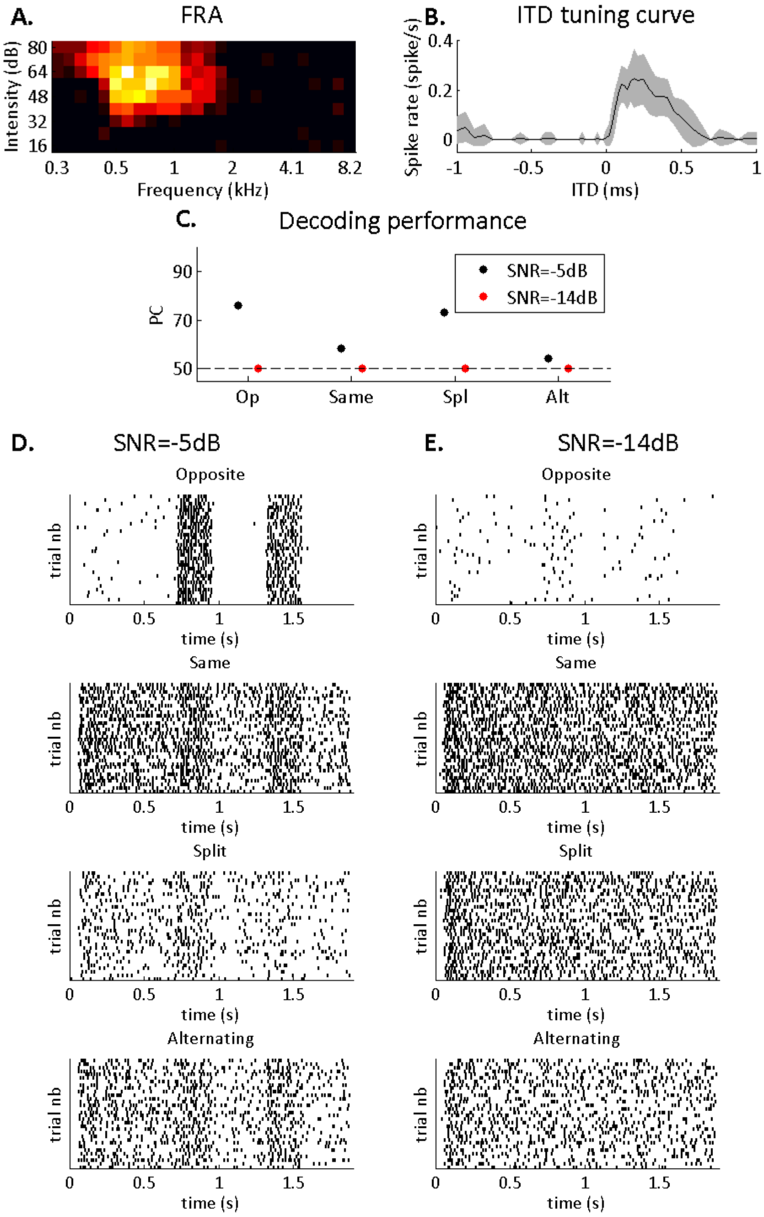

Figure 24: Example of a cell recorded in the left IC. **A.** Frequency response area. **B.** ITD tuning curve. **C.** Single cell decoding performance for SNR=-5dB (black dots) and -14dB (red dots). **D.** Rasters of the cell's activity at SNR=-5dB during 32 repetitions of one full trial where the stimulus was masker only – masker + Reference vowel – masker only – masker + Reference vowel – masker only for the 4 spatial conditions. **E.** Same rasters as in D. recorded at SNR=-14dB.

We analysed our decoding population by splitting the cells into groups defined by their ITD tuning. We observed that the performance of significantly ITD tuned cells was globally better than that of the non-ITD tuned cells, both for single cell average and population decoding and both for SNR=-5dB (Figure 25A,B) and SNR=-14dB (Figure 26A,B). This indicates that being ITD tuned might allow the cell to have more information on vowel identity in all the spatial conditions. This was surprising because while we expected the performance of ITD tuned cells to be better in the Opposite condition, we did not expect it to be better in the Same condition where the vowels and the masker were presented at the same ITD. This suggests that the vowels might not be completely masked in the Same condition.

We observed significant spatial unmasking between the Opposite and Same conditions for ITD tuned cells at SNR=-5dB but these cells did not follow the psychophysical trends (Figure 25A). This confirmed the hypothesis that ITD tuned cells perform vowel decoding better in some spatial conditions but we might have to split them further in groups defined by their preferred ITD values to follow the psychophysical trends. We observed that non-ITD tuned cells showed spatial unmasking when using population decoding but not single cell average decoding (Figure 25B). It is surprising that non-ITD tuned cells also showed spatial unmasking as they should not be able to take advantage of the difference in ITD between the noise and masker. However, when looking at the properties of the non-ITD tuned cells individually we observed that they often had a low firing rate or very variable responses to the frozen noise snippets we used to measure the ITD tuning curve. This prevented us from accurately measuring their ITD tuning curve and was problematic for defining the significance of their ITD tuning. These cells might respond differently to a more complex stimulus such as our full task and have significant ITD tuning during the task.

For SNR=-14dB, we observed the same results: ITD tuned cells showed spatial unmasking for both decoding methods but did not follow the psychophysical trends (Figure 26A) and non-ITD tuned cells showed unmasking only when measured by population decoding (Figure 26B).

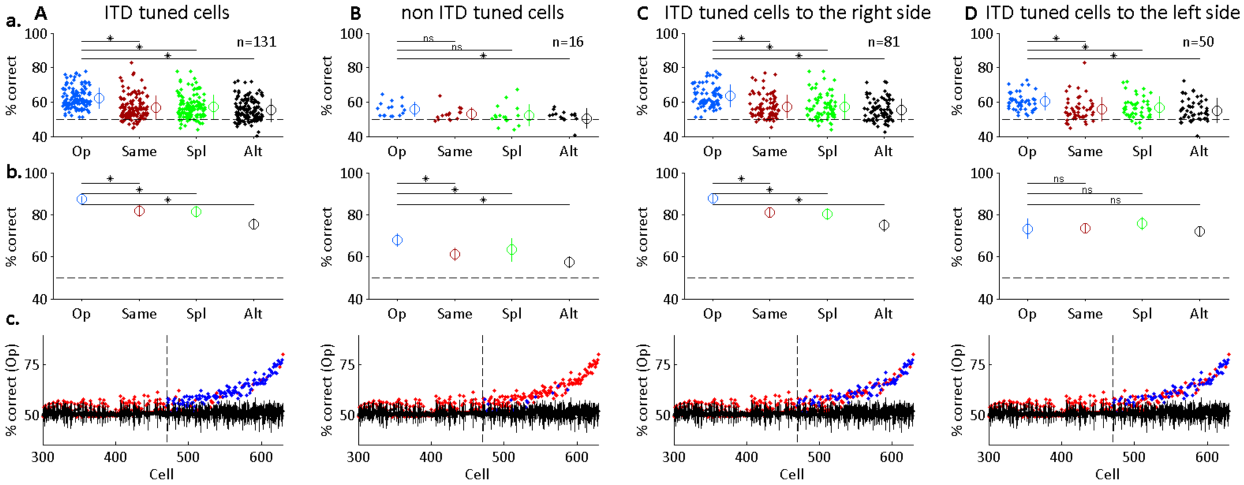

Figure 25: Decoding performance of groups of cells defined by their ITD tuning at SNR=-5dB. **A.** Significantly ITD tuned cells: the decoding performance for the ITDs of the noise snippets played for the ITD tuning curve measurements is at least 2 standard deviations better than the chance decoding performance (see III.1.e). **B.** Non-significantly tuned cells. **C.** Significantly ITD tuned cells with a positive best ITD, i.e. tuned to the right side. **D.** Significantly ITD tuned cells with a negative best ITD, i.e. tuned to the left side. **a.** Single cell decoding performance. **b.** Population decoding performance. **c.** Significance of the decoding performance in the Opposite condition with the cells from the group plotted in blue.

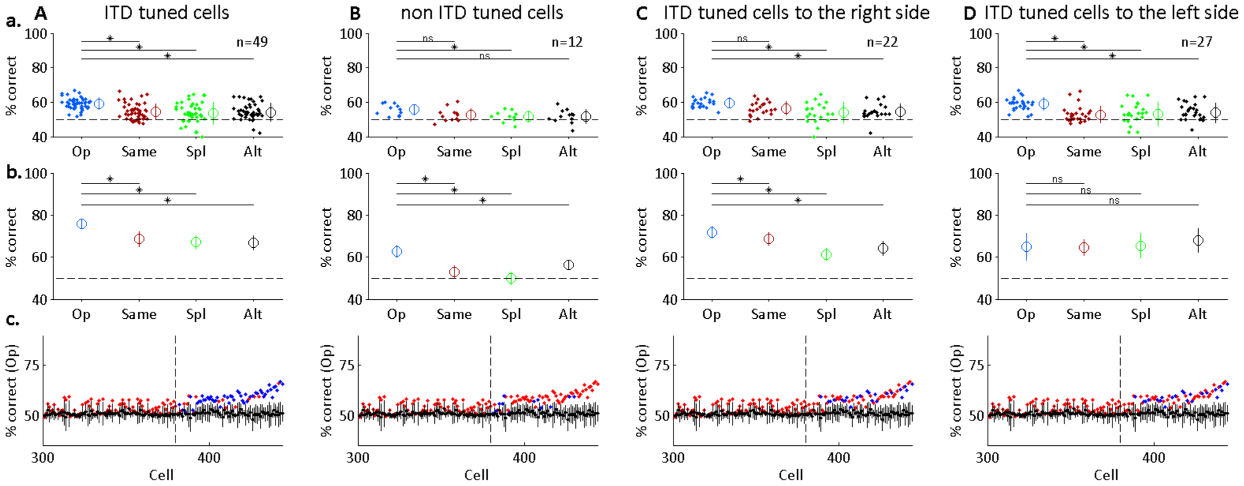

Figure 26: Same representation as Figure 25 at SNR=-14dB.

The side to which the cells are ITD tuned might also play a role in spatial unmasking. For simplicity, let us call ‘preferred side’ the side to which the cells are ITD tuned and ‘non-preferred side’ the side to which the cells responds less or not at all to the sounds. If a cell is ITD tuned to the right side, the vowels are on its preferred side and the masker on its non-preferred side in the Opposite condition, and both the vowels and the masker are on its preferred side in the Same condition. Conversely, if a cell is ITD tuned to the left side, the vowels are on its non-preferred side in the Opposite condition, and both the vowels and the noise are on its non-preferred side in the Same condition. Cells ITD tuned to the left side might hence have a poorer decoding performance.

For SNR=-5dB, we observed spatial unmasking between the Same and the Opposite condition for cells tuned to the right side (Figure 25C). Cells tuned to the left side showed spatial unmasking when using single cell average decoding (Figure 25Da) but not population decoding (Figure 25Db). In this latter case, the performance in the Same condition was the same as in the Opposite condition. We can understand this because in the Opposite condition the masker is at preferred ITDs for these cells and might dominate all the neural activity while in the Same condition both maskers and vowels are at non-preferred ITDs which might give a chance to the cells to respond to the vowels.

For SNR=-14dB, we observed spatial unmasking for cells tuned to the right side using population decoding (Figure 26C) and for cells tuned to the left side using single cell average decoding (Figure 26D) and no group followed the psychophysical trends. This confirmed that spatial unmasking was a strong feature in the population. The fact that the two decoding methods gave different results indicated that the task at SNR=-14dB might be too hard to reliably analyse with our cell population.

To conclude, we observed that the effects observed for ITD tuned cells seemed to be mostly accounted for by cells that preferred sounds coming from the right side of the head. This result confirmed our hypothesis and was intuitive given that the choice of ITDs for the vowels and the masker that seemed to favor cells tuned to the right side. We also observed that some non-ITD tuned cells also showed spatial unmasking which might be due to these cells having different responses to frozen noise snippets and the vowel in masker stimulus. We observed spatial unmasking at SNR=-14dB only for some decoding methods and concluded that the vowel decoding task might be too difficult to resolve in these smaller groups of cell with our analysis.

g. Influence of frequency tuning on the decoding performance

Frequency tuning is also a very important parameter in our paradigm since we are trying to differentiate between vowels that differ only by their formants center frequencies. We hypothesised that cells that are frequency tuned, and in particular frequency tuned to the formants frequencies would have a better vowel decoding performance.

We will first consider one example and then go on to the systematic analysis. A cell with a best frequency of 588Hz (Figure 27A) is likely to have all our vowels in its receptive field. It will probably respond more strongly to the first formant of R and D2 (F1=630Hz) than to the first formant of D1 and D3 (F1=510Hz) because it is closer to its best frequency. The second formant of all the vowels are on the edges of this cell's receptive field, but the second formant of D2 and D3 (F2=1110Hz for D2 and F2=1050Hz for D3) is closer to its best frequency than the second formant of R and D1 (F2=1230Hz). Based on its FRA, we hence expect this cell to fire more to D2, less for R and D3 and even less for D1. These differences in firing rates might then allow vowel discrimination.

In the SNR=-5dB condition (Figure 27D), we observed that our example cell fired more spikes in response to D2 and R in the Opposite condition and in the Split condition where the first formant was on the right side of the head (preferred ITD for this cell, Figure 27B), which corresponds to what we expected from the FRA. In the Same and Alternating conditions the spike count distributions for the different vowels seemed to be very similar. It hence seemed that we could qualitatively predict the cell's responses to the vowels from its FRA when at least one formant was at this cell's preferred ITD. We also observed that the decoding performance was best in the Opposite and Split conditions where the spike count distributions in response to the different vowels were more diverse (Figure 27C).

At SNR=-14dB, the spike count distributions to the different vowels were very overlapping in all the spatial conditions, and the vowel decoding performance was very close to chance. This confirms that the decoding performance seems to be correlated with the diversity of spike count distributions and that the intense masker prevents us from doing successful vowel decoding at SNR=-14dB.

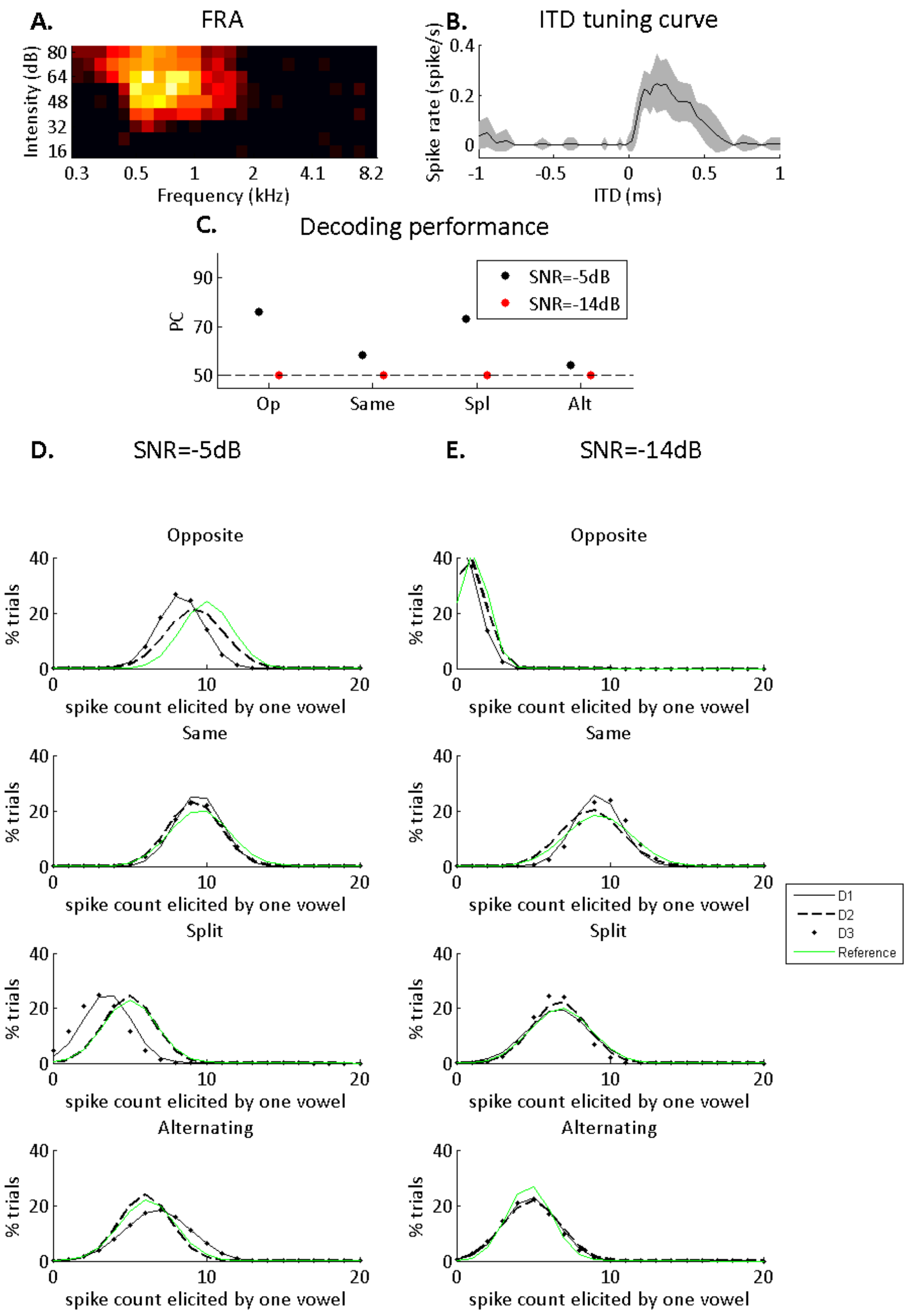

Figure 27: Same example cell as in Figure 24. **A.** Frequency response area. **B.** ITD tuning curve. **C.** Single cell decoding performance. **D.** Distributions of the number of times each spike count was elicited by each vowel fitted to normal distributions truncated at 0, measured at SNR=-5dB in the 4 spatial conditions. **E.** Same representation as D. at SNR=-14dB.

To study the effects of frequency tuning systematically, we split our decoding cell population into groups defined by their frequency tuning significance and range. We observed that most of our cells were significantly frequency tuned (82.5% for SNR=-5dB, Figure 28; 76.9% for SNR=-14dB, Figure 29). The significantly tuned cells performed overall better than the non-significantly tuned cells (Figure 28 and Figure 29 a,b) which confirmed our hypothesis that frequency tuning is an important factor for vowel decoding. However, this did not seem to correlate with the significance of the decoding in the Opposite condition (Figure 28c and Figure 29c).

We observed spatial unmasking between the Opposite and Same condition for both significantly tuned and non-significantly tuned cells at SNR=-5dB (Figure 28a,b). We also observed spatial unmasking for both groups of cells at SNR=-14dB when using both decoding methods (Figure 29a,b). It is surprising that we observed spatial unmasking and a better than chance decoding performance for non-significantly tuned cells as they would respond in the same way to all the vowels if they were not frequency tuned. However, we measured frequency tuning using single pure tones which might not be sufficient to predict the responses to more complex sounds like our vowels in noise for some cells. Once again, this showed that spatial unmasking is a strong feature in our population and suggested that evaluating the cells receptive fields using simple stimuli might not allow us to predict their responses to complex stimuli.

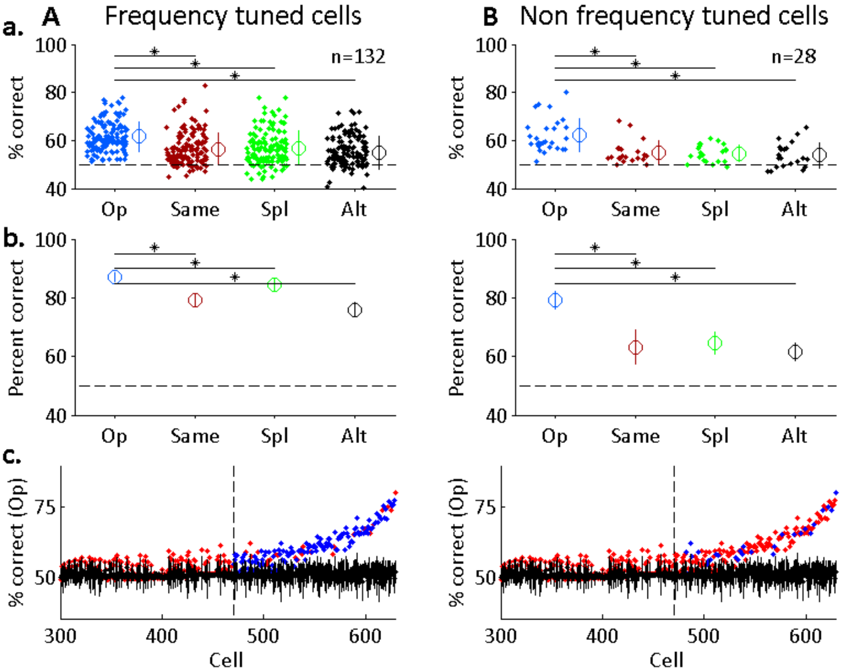

Figure 28: Decoding performance of groups of cells defined by the significance of their frequency tuning at SNR=-5dB. **A.** Significantly frequency tuned cells: the decoding performance for the tone frequencies of the pure tones played for the FRA measurement is at least 2 standard deviations better than the chance decoding performance (see III.1.e). **B.** Non-significantly tuned cells. **a.** Single cell decoding performance. **b.** Population decoding performance. **c.** Significance of the decoding performance in the Opposite condition with the cells from the group plotted in blue.

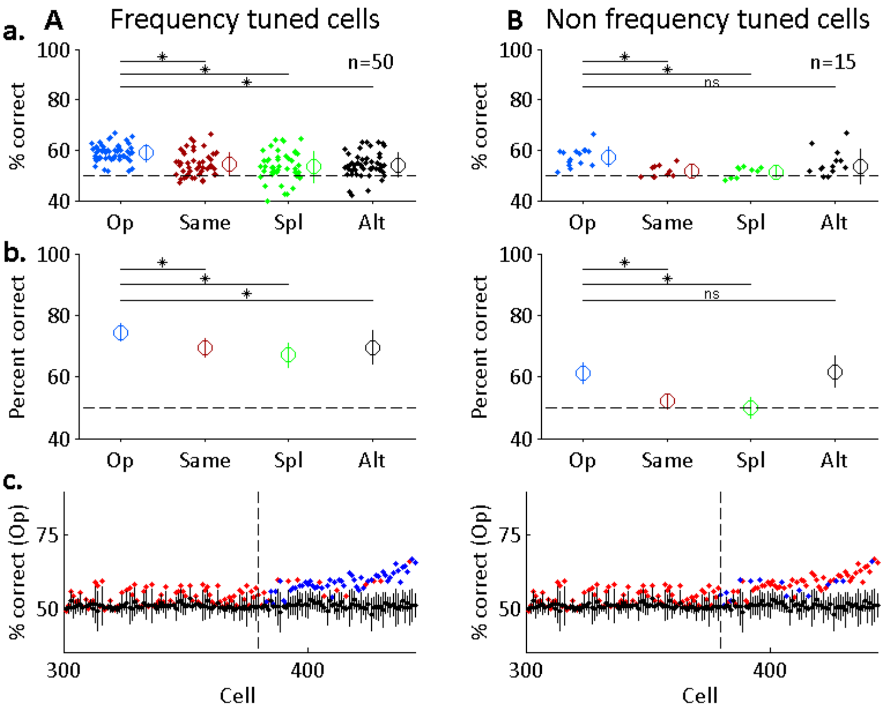

Figure 29: Same representation as Figure 28 at SNR=-14dB.

We observed that significantly tuned cells to any frequency did not follow the psychophysical trends. This group of cells might include cells that have very distant best frequencies from the formants of our vowels and hence have less contrasted responses to the different vowels. To test the effect of best frequency on the vowel decoding performance, we split these cells in groups that had their best frequency in the range of the vowel’s first formants, second formants or away from the formants.

At SNR=-5dB, we observed that all the groups showed spatial unmasking when measured as the single cell average performance (Figure 30a). When measured as a population, only cells tuned to frequencies in the range of the first formant and tuned to non-formant frequencies showed spatial unmasking (Figure 30b). Only the cells with a best frequency in the range of the first formant followed the psychophysical trends when measured as a single cell average and as a population (Figure 30Aa,b). As in the previous section, we observed no correlation between the range of frequency tuning and the significance of the decoding in the Opposite condition (Figure 30c). This confirmed our hypothesis that frequency tuning is important for vowel decoding, and that cells tuned in the range of the formants perform more similarly to the psychophysical trends. The fact that only cells tuned to the first formant and not to the second formant followed the psychophysical trends was surprising. In the Split condition, the first formant was presented on the right side of the head and the second formant on the left side. We noticed that and that slightly more cells in our decoding population were ITD tuned to the right side (n=66) than to the left side (n=40), which might lead to a higher decoding performance for the formant presented on the right side of the head.

At SNR=-14dB, we observed spatial unmasking in cells tuned to non-formant frequencies (Figure 31C) but no spatial unmasking in cells tuned in the range of either formants (Figure 31A,B). We had very few cells tuned to the range the formants so it was difficult to make a statement about these groups.

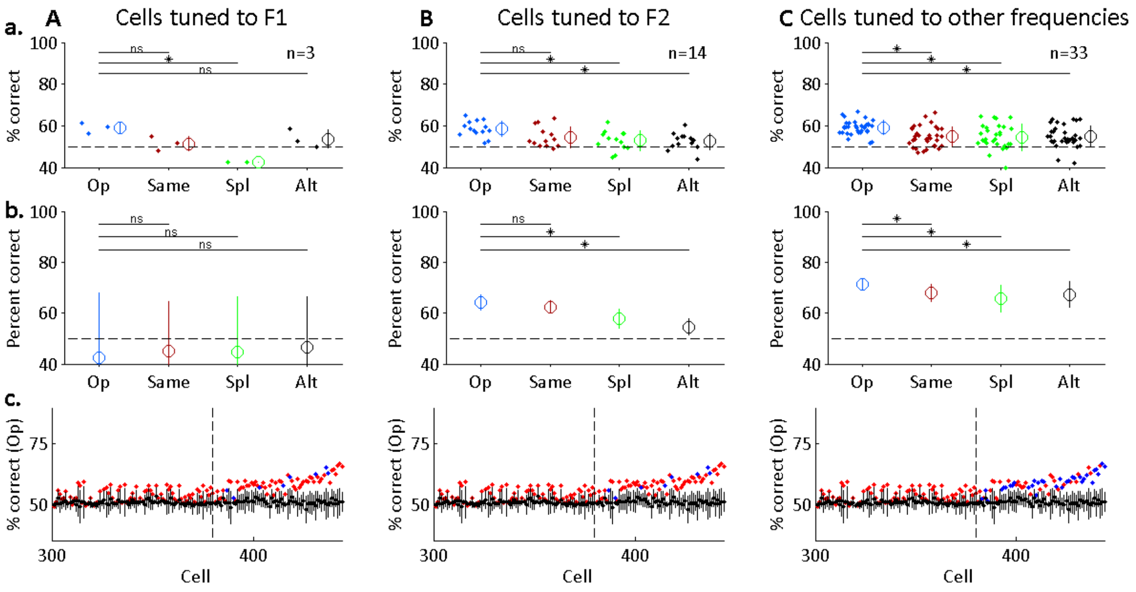

Figure 30: Decoding performance of groups of cells defined by their frequency tuning at SNR=-5dB. **A.** Significantly frequency tuned cells with a best frequency between 450Hz and 690Hz. **B.** Significantly tuned cells with a best frequency between 990Hz and 1290Hz. **C.** Significantly tuned cells to other frequencies. **a.** Single cell decoding performance. **b.** Population decoding performance. **c.** Significance of the decoding performance in the Opposite condition with the cells from the group plotted in blue.

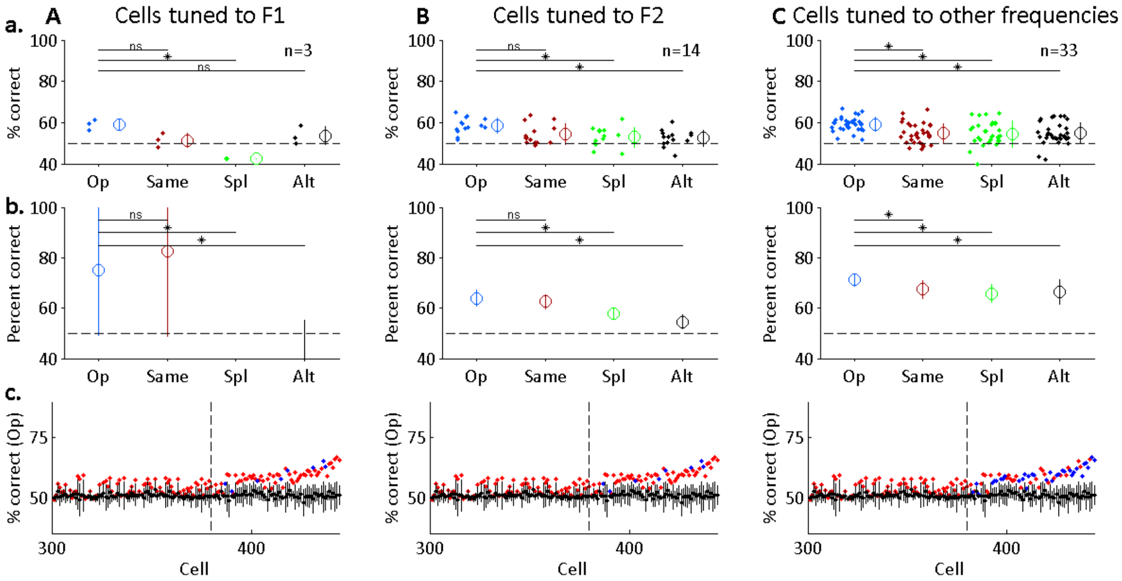

Figure 31: Same representation as Figure 30 at SNR=-14dB.

We used three Different vowels in our stimulus which differ from the Reference vowel by the frequency of one or both formants:

- Reference vowel: $F1_R$=630Hz, $F2_R$=1230Hz;
- Different vowel 1: $F1_{D1}$=510Hz, $F2_{D1}$= $F2_R$;
- Different vowel 2: $F1_{D2}$= $F1_R$, $F2_{D2}$= 1110Hz;
- Different vowel 3: $F1_{D3}$=510Hz, $F2_{D3}$=1050Hz.

We hypothesized that cells tuned to the frequency range of the vowels' first formant might discriminate better between Different vowel 1 and the Reference vowel than between Different vowel 2 and the Reference vowel. Conversely, cells tuned to the second formant frequency range might discriminate better between Different vowel 2 and the Reference vowel than between Different vowel 1 and the Reference vowel. Cells tuned to the frequency range of either formant might have a good discrimination performance for Different vowel 3.

To test these hypotheses, we computed the discrimination performance of groups of cells defined by their frequency tuning for each Different vowel independently. At SNR=-5dB, we indeed observed that cells tuned to the first formant frequency range followed the psychophysical trends for Different vowel 1 (Figure 32A,a) and that cells tuned to the second formant frequency range followed the psychophysical trends for Different vowel 2 (Figure 32B,b). This confirmed our hypothesis that the discrimination performance was better when the Different vowel is different from the Reference vowel in the cells' best frequency range. We noticed that the discrimination performance in the Opposite condition was similar for all groups of cells and all Different vowels and that this effect was most prominent in the Split condition.

We observed that the discrimination performance between Different vowel 3 and the Reference vowel was very high in all spatial conditions including the Same and Split conditions, which sometimes caused the spatial unmasking effect between the Same and Opposite conditions to disappear (Figure 32c). Both formants of Different vowel 3 are different from those of the Reference vowel which might make the task too easy to observe differences between spatial conditions. When considering all the significantly frequency tuned cells (Figure 32C), we observed that they followed the psychophysical trends for Different vowel 1 and 2 but not Different vowel 3.

Overall, we observed that the relation between the best frequency range of cells and the frequency region where the Different vowels differ from the Reference vowel was crucial for the discrimination performance, especially in the Split condition. It seemed that cells which best frequency matched the region where the discrimination had to take place followed the psychophysical trends, and that this effect could still be seen in the whole frequency tuned population (Figure 32a,b). It is possible that the discrimination between Different vowel 3 and the Reference vowel was too easy especially in the Same condition, which made spatial unmasking disappear for some cell populations. It would be interesting to do a similar study using only one Different vowel that differs from the Reference vowel by only one formant to be able to study the effects of our spatial conditions with less potentially confounding factors coming from the frequency receptive fields.

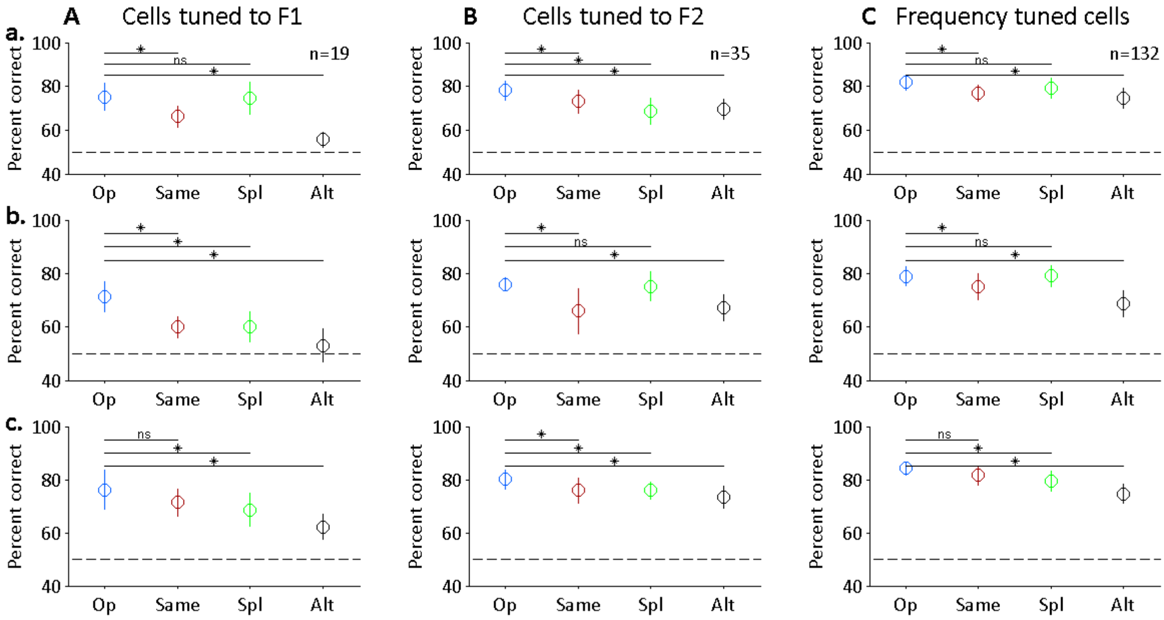

Figure 32: Population decoding performance of groups of cells defined by their frequency tuning at SNR=-5dB for each Different vowel. **A.** Significantly frequency tuned cells with a best frequency between 450Hz and 690Hz. **B.** Significantly tuned cells with a best frequency between 990Hz and 1290Hz. **C.** Significantly tuned cells to any frequency. **a.** Decoding performance between the Different vowel 1 (F1=510Hz, F2=F2$_R$) and the Reference vowel (F1$_R$=630Hz, F2$_R$=1230Hz). **b.** Decoding performance between the Different vowel 2 (F1=F1$_R$, F2=1110Hz) and the Reference vowel. **c.** Decoding performance between the Different vowel 3 (F1=510Hz, F2=1050Hz) and the Reference vowel.

At SNR=-14dB, the low number of cells tuned to the each formant frequency range made it difficult to analyse the data, and no group of cells followed the psychophysical trends (Figure 33). We did observe spatial unmasking between the Same and the Opposite condition for all frequency tuned cells when considering the discrimination between Different vowel 3 and the Reference vowel (Figure 33C,c), suggesting that lowering the SNR might make this particular discrimination task of a more appropriate difficulty to study spatial unmasking.

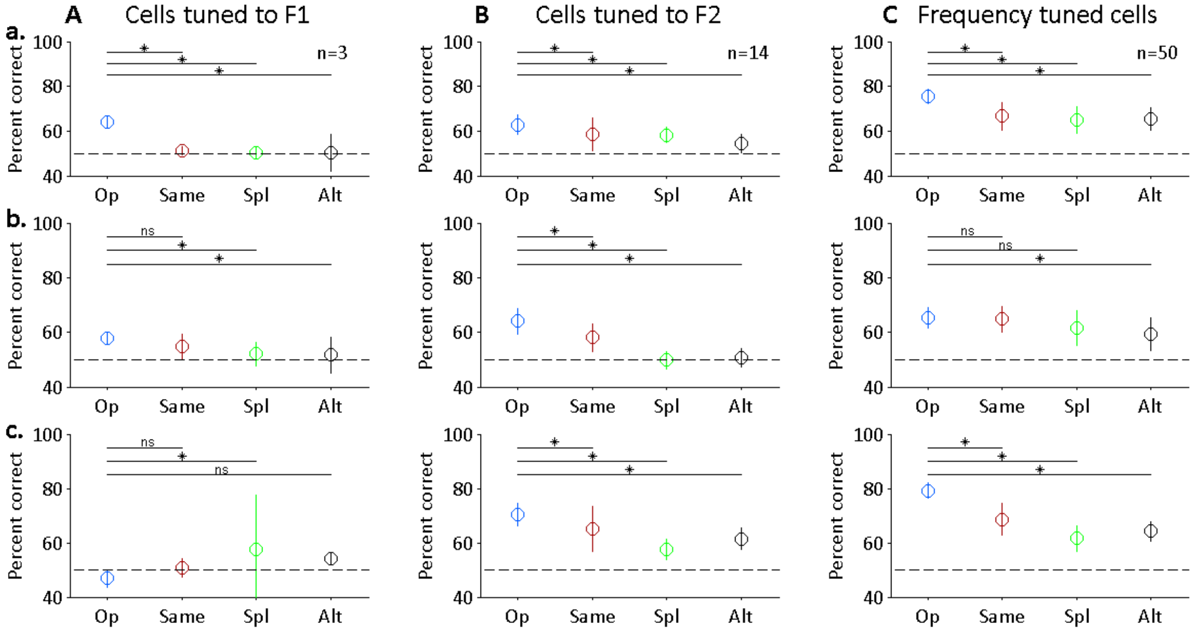

Figure 33: Same representation as Figure 32 at SNR=-14dB.

To conclude, we found that both significantly and non-significantly frequency tuned cells showed spatial unmasking using at least one of the decoding methods. This confirmed that spatial unmasking was a strong feature in our population and suggested that our method to evaluate the frequency receptive field might be insufficient to predict responses to more complex stimuli. However, we observed that the relation between the best frequencies of groups of cells and the discrimination performance for each Different vowel was important for spatial unmasking. When considering all the Different vowels together, only cells tuned to the first formant frequency range followed the psychophysical trends at SNR=-5dB. When considering the Different vowels one by one, we observed that cells tuned to the first formant frequency range followed the psychophysical trends for Different vowel 1, cells tuned to the second formant frequency range followed the psychophysical trends for Different vowel 2, and that the task seemed too easy at SNR=-5dB for Different vowel 3. It hence seemed that the frequency tuning of the cells and our choice of stimuli had a large influence on our results, and that it might be useful to remove this influence in order to study more directly the influence of spatial cues on the discrimination performance.

At SNR=-14dB, we observed that even with our large number of cells, it was difficult to have enough statistical power to study groups of cells tuned to each formant's frequency range. Recording from a larger population of cells could allow us to clarify some of the dependencies on frequency tuning further at that SNR.

h. Characteristics of cells that follow the psychophysical trends

For our analysis, we started by selecting only the cells that decoded the vowels in the Opposite condition better than chance. Then, we split them into groups depending on their firing rate, ITD tuning and frequency tuning and observed each group's decoding performance. At SNR=-5dB, the following groups had a decoding performance that followed the psychophysical trends:

- Cells that spiked more than 25 spikes per trial,
- Cells that were ITD tuned to the right side,
- Cells that were frequency tuned to the frequency range of the first formant.

At SNR=-14dB we never observed a group of cells that followed the psychophysical trends.

These results are quite intuitive as these cells had a high firing rate, which is susceptible to convey information about the vowel identity, were tuned to the right side of the head where at least part of the vowels were presented in all the spatial condition (the full vowels in the Same and Opposite condition, one formant in the Split condition and one harmonic from each formant in the Alternating condition) and have at least the first formant in their frequency receptive field.

In our decoding population, only 4 cells were part of these three categories. Because it is so few cells we were not able to analyse them as a population but we could look at their single cell properties (Figure 34). As expected from the categories they were chosen from, they all had a low frequency FRA and a clear ITD tuning curve with a positive best ITD. However, we could still observe some diversity in their responses with different shapes of the FRAs and ITD tuning curves, and different rasters to the stimulus in the Opposite and Same conditions.

We observed that only one of these cells followed the psychophysical trends in its single cell vowel decoding performance (Figure 34C). We also noticed that other cells we had studied individually before followed the psychophysical trends but were not part of this subset of cells (for example Figure 19A). It hence seemed that there was a correlation between the properties of single cells and their decoding performance as a group, there was no systematic equivalence between the properties of single cells and their following of the psychophysical trends.

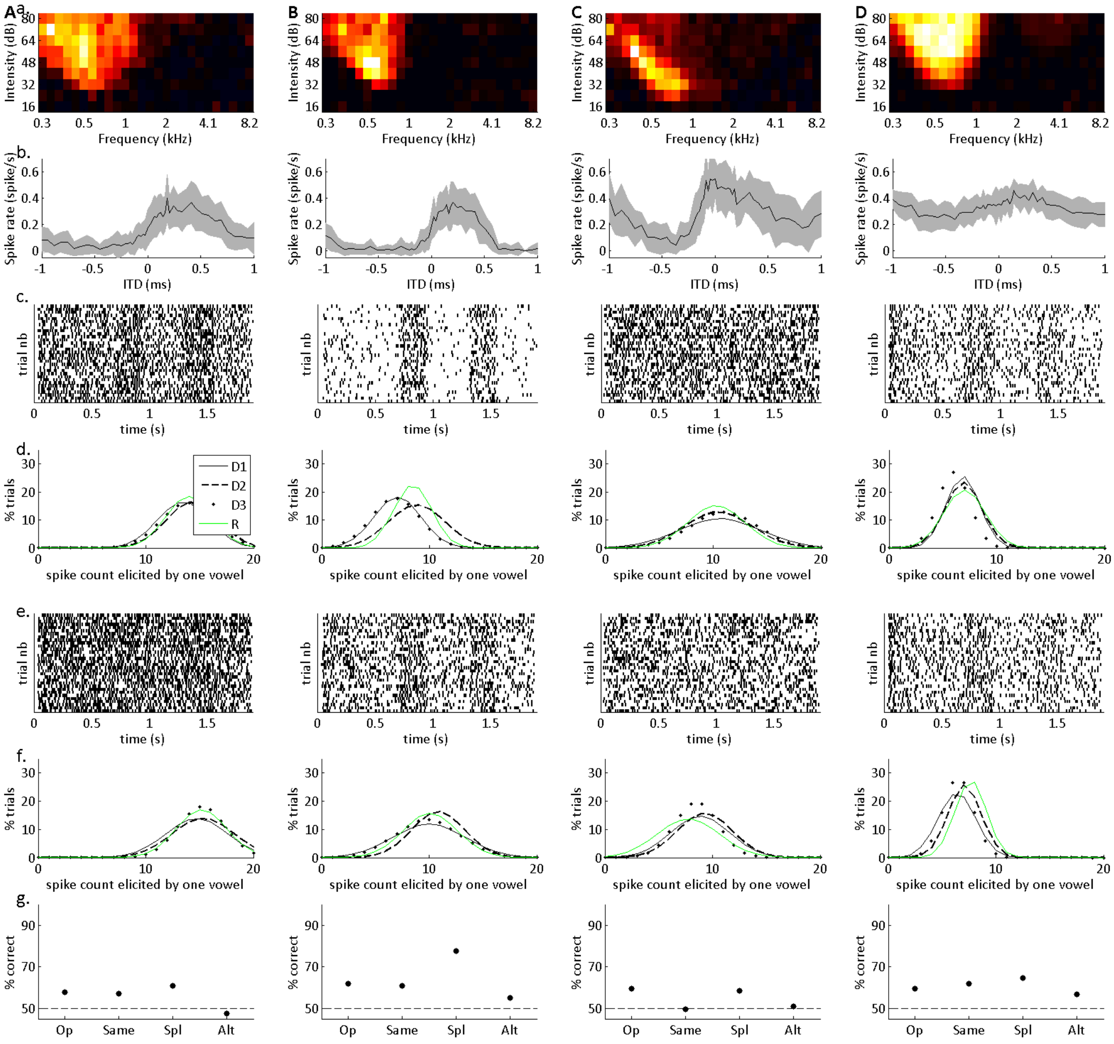

Figure 34: Receptive fields, responses and decoding performances of 4 cells that were part of all the groups following the psychophysical trends in the SNR=-5dB condition. **a.** Pure tone FRAs. **b.** Noise ITD tuning curves. **c.** Rasters of the full stimulus (masker only – Reference vowel – masker only – Reference or Different vowel – masker only) in the Opposite condition. **d.** Gaussian distributions fitted to the spike count histograms for each vowel separately in the Opposite condition. **e.** Raster (same as c.) in the Same condition. **f.** Spike count distributions (same as d.) in the Same condition. **g.** Single cell vowel decoding performance in the 4 spatial conditions.

In hindsight, we could have modified the experimental design to maximise the number of recorded cells that showed all three criteria from the groups that followed the psychophysical trends. We could have recorded only from the left IC so that most cells would be tuned to the right side of the head and made the frequency of first formant higher so that more cells would be tuned to the first formant frequency of our vowels. We could have tried to play the sounds a little louder to promote higher firing rates in the population.

i. Influence of the starting phase

Like in the psychophysical experiment, we recorded neural responses to the Starting Phase condition where the vowels were presented from the right side of the head, the masker from the left side, and the harmonics of each formant started out of phase with each other. We will show here the results for this condition only for the groups of cells that followed the psychophysical trends.

We observed that for all the groups the performance in the Starting Phase condition was the same as in the Opposite condition (Figure 35). It hence seemed that the starting phases of the harmonics did not influence the decoding performance, as we had observed in the psychophysical experiment. We sometimes observed a significant difference between the Starting Phase and Same condition, but it depended on the groups and decoding method we considered. We might have needed a larger population of cells here to achieve more statistical power.

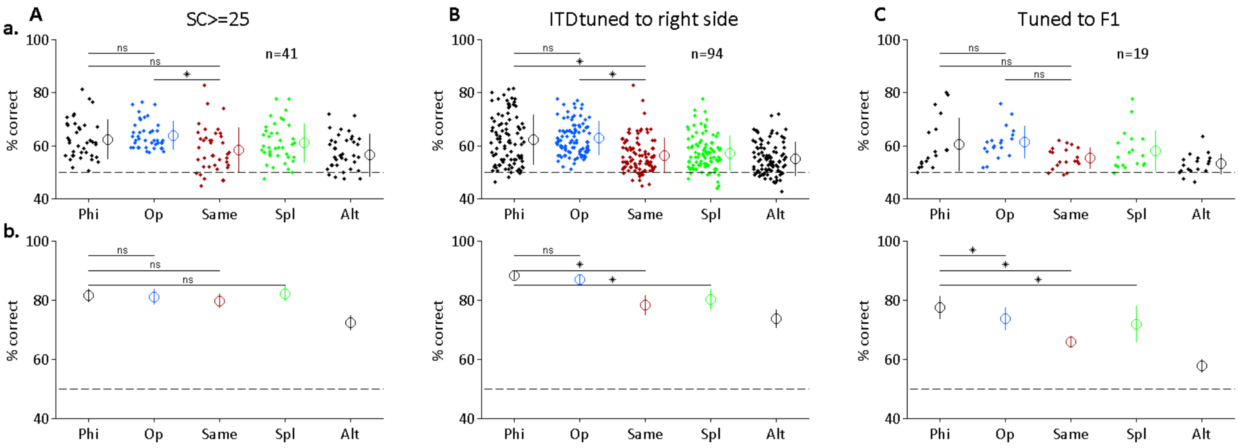

Figure 35: Decoding performance of groups of cells that followed psychophysical trends at SNR=-5dB. **A.** Cells that fire more than 25 spikes per trial. **B.** Significantly ITD tuned cells to the right side of the head. **C.** Significantly frequency tuned cell in the range of the first formant. **a.** Single cell decoding performance. **b.** Population decoding performance.

Overall, the results indicated that the formant's starting phases was not a critical parameter for the vowel discrimination task, as we had shown in our psychophysical experiment.

3. Discussion

a. Population coding in the IC

We showed that vowel identity in the IC is encoded by the spike rate of large populations of neurons. Cells could phase lock to the fundamental frequencies of the vowels but that would not be informative about vowel identity in our stimulus, which is consistent with previous studies showing that vowel identity is encoded by spike rate while consonant are encoded by spike timing (Perez et al. 2012).

We analysed the activity of population of cells grouped by their responses to the stimulus or their frequency or ITD receptive fields. We observed that almost all of these subgroups showed a better performance in the Opposite condition where the vowels and the masker were presented at opposite ITDs than in the Same condition where they were presented at the same ITD. While we expected some of the groups to show this spatial unmasking (for example for ITD tuned cells or frequency tuned cells), it was more surprising for other groups like non-ITD tuned or non-frequency tuned cells. We also observed that even though the decoding performance was worse at SNR=-14dB condition than at SNR=-5dB and we sometimes lacked statistical power, we still observed this spatial unmasking in most groups at SNR=-14dB. We hence concluded that spatial unmasking was a strong feature of our population.

We made several hypotheses about which cells would have a better performance and follow the psychophysical trends that were confirmed and refined by our results. We found that cells with higher firing rates had a better decoding performance and followed the psychophysical trends. This seems to indicate that a higher firing rate could convey more information about vowel identity. We found that cells that were ITD tuned to the right side followed the psychophysical trends, which was expected given the properties of our stimulus. But we also found that cells that weren't ITD tuned showed spatial unmasking. Finally, we found that cells that were frequency tuned to the first formant followed the psychophysical trends, but that non frequency tuned cells could show spatial unmasking. We hypothesised that non-ITD tuned and non-frequency tuned cells might be non-significantly tuned when measured using simple stimuli (frozen noise bursts or pure tones) but show a significant amount of tuning when using more complex stimuli (here vowels in the masker). Overall, we observed that only some specific categories of cells followed the psychophysical trends, while most of the others showed spatial unmasking.

We observed that there was a correlation but no direct correspondence between single cell properties and the decoding performances in the subgroups. It seems that the behavior of a cell population could not be directly predicted from the behavior its single cells. This might be due to an advantage of analysing several cells at the same time as a population or to the large diversity of cells. We noticed that analysing the cell populations as an average of single cells or as a population sometimes yielded different results, but we lacked statistical power to understand the difference.

This population code is reminiscent of the most recent models of ITD encoding in the IC, whether the hemispheric 2-channel model (David McAlpine, Jiang, and Palmer 2001; N. A. Lesica, Lingner, and Grothe 2010), a pattern decoder (Day and Delgutte 2013a; Belliveau, Lyamzin, and Lesica 2014). It seems that information about spatial location and simple sound identity are contained in large diverse populations of IC neurons. This could allow robustness to different sound types and reverberation as any subset of cells would be able to encode sound identity and perform spatial unmasking.

The main ascending projection of the ICC is to the ipsilateral medial geniculate body (MGB) and it is not yet fully understood (Jeffery A. Winer and Schreiner 2005a). Some IC regions project to single MGB regions and others to many distinct MGB regions. Taking advantage of spatial unmasking using a large and diverse population of ICC neurons could allow the effect to propagate to the MGB through this rather diffuse connection.

b. SNR-5dB VS -14dB

We played our sound stimuli in two intensity conditions. The intensity of the vowels was always the same and the intensity of the masker was 5dB or 14dB more intense than the vowels. We observed that most subgroups of cells showed spatial unmasking in both conditions, which again showed the robustness of this effect. However, no subgroup of cells followed the psychophysical trends at SNR=-14dB.

Even with our large number of cells, only very few had a performance in the Opposite condition at SNR=-14dB significantly better than chance. This small number of cells caused the subgroups to be quite small and we lacked statistical power to analyse these properties. It is likely that a behaving gerbil could not perform the discrimination task at SNR=-14dB, but some information about the vowel identity might still be present in some IC neurons. Indeed, we observed that 65 out of 361 cells (18.0%) showed a better than chance performance in the Opposite condition. If we consider the whole IC composed of about a million cells and assume that this ratio is conserved, about 180,000 cells would show some spatial unmasking in the Opposite condition. This might not allow an animal to resolve the task, but it is nonetheless interesting to study the brain responses to such a difficult SNR condition.

We might also run into different effects due to the cells rate level functions. Indeed, some cells increase monotonously their spike rate with sound intensity while other cells plateau or decrease their spike rate with intensity after it has crossed a given threshold. This could make some cells of our population spike less in the SNR=-14dB condition which would hinder their performance.

c. Importance of both hemispheres

We found that both ispsilateral and contralateral ICs participated in spatial unmasking with target sounds in the contralateral hemi-field or split between the two sides. Other studies indeed suggest to consider the two ICs as one entity based on the importance of commissural connections for sound frequency and intensity representation (Orton and Rees 2014). This may help to explain the range of deficits observed after IC lesions (William M. Jenkins and Bruce 1982): the lesions mostly affected sound localization performance on the contralateral side but it remained above chance while the performance in the ipsilateral side was closer to the control values but still affected.

This is similar to observations in auditory cortex, where unilateral lesions mostly impact sound localization in the contralateral hemi-field, but some performance impairments are observed in the ispsilateral one (William M. Jenkins and Bruce 1982; Malhotra, Hall, and Lomber 2004b). It hence seems that both hemispheres of the auditory pathway have a functional cooperation.

We found that most cells had excitatory responses to the vowels in the contralateral hemi-field which were disrupted by noise in the same hemi-field, providing an intuitive explanation spatial unmasking at the single cell level. However, a wide variety of cells contributed to the unmasking including cells for which vowels were in the ipsilateral hemi-field. These cells had several response properties such as ‘inverted’ ITD tuning curves with ispsilateral preferred ITDs or preferred contralateral ITDs with inhibitory responses to the vowels which modulated the response to the noise in the contralateral hemi-field. We speculated that this inhibition might come from earlier nuclei such as the DNLL (Siveke, Leibold, and Grothe 2007) or from a non-monotonic rate-level function. This illustrates how both hemispheres might cooperate through a wide variety of mechanisms which could participate to the robustness of the spatial unmasking phenomenon.

d. Influence of anesthesia and attention

Neural activity in the IC seems to be robust to anesthesia: temporal coding in awake passive and anesthetized gerbils is similar (Ter-Mikaelian, Sanes, and Semple 2007a) and interaural correlation (Coffey et al. 2006) and ITD encoding (Day and Delgutte 2013a) in the unanesthetized rabbits is similar to what was observed under anesthesia. However the animals were not attending to the sounds in these studies which might mask essential differences. In the primary auditory cortex (A1) spatial tuning was first described as very broad in the anesthetized cat, especially for low-frequency neurons (Middlebrooks and Pettigrew 1981). This was confirmed by experiments on awake passive cats (Mickey and Middlebrooks 2003) but when the animals were trained to use sound location spatial tuning considerably sharpened (Lee and Middlebrooks 2011; Middlebrooks and Bremen 2013).

Studies in behaving primates suggest that some IC cells act as reward predictors (Metzger et al. 2006) but it is difficult to compare the primate to the rodent data and assess the influence of attention on neuronal properties (Metzger et al. 2006; Ryan and Miller 1978). Unpublished observation of neuronal activity in the awake passive gerbil IC in our laboratory showed very different neuronal spontaneous activity, so we believe that the animal’s state is instrumental in shaping neuronal responses. Gerbils can be trained to perform simple auditory tasks (Schebesch et al. 2010) so awake behaving IC recordings could bring interesting insights in the future.

e. Importance of the IC

We showed that the decoding performance of some neuronal populations in the IC followed the same qualitative trends as human psychophysical behavior in several complex spatial conditions. This suggests that discriminating simple sounds and performing spatial unmasking could be already done in the midbrain without requiring attention. Other studies indeed suggest that spatial segregation of two streams could be done before cortical processing (Middlebrooks and Bremen 2013) and the IC seems to have a more prominent role for sound processing in humans than previously thought as its activity is dependent on context, language and musical training (Chandrasekaran, Skoe, and Kraus 2013).

Maddox et al. (2012) used a similar paradigm to ours in the zebra finch where they studied the encoding of bird-song in the A1 analog using different spatial configurations of the bird-song and masking noise source. Interestingly, they also observed a wide variety of individual neuronal behavior and could find small neuronal populations that encoded bird-song accurately in any spatial configuration, but the behavioural performance was best represented by averaging over the whole population. Although it is hard to directly compare this study to our data, we could think that the effects they observed were already present in the IC analog.

If at least coarse spatial segregation and sound encoding already takes place in the IC independently of attention, we can conjecture that precise spatial information doesn't need to be conveyed to the cortex in a resting state as it has already been processed in the midbrain.

f. Speculations about developing cochlear implants

Even though the use of bilateral cochlear implants is becoming more common, patients still cannot take advantage of natural ITD cues (Hoesel and Tyler 2003). Previous research has shown that normal subjects could take advantage of ITD cues for spatial unmasking of speech in noise even when a sound source had several different ITDs as long as the ITDs were coherent within one auditory filter (Beutelmann, Brand, and Kollmeier 2009), and we have confirmed these findings with a simpler stimulus. This suggests that the emphasis should be put on conveying ITD cues through each frequency channel but that the ITDs do not have to be synchronized across channels. We could consider optimizing only one electrode of cochlear implants for conveying ITD cues, and matching the 'ITD-optimized' electrodes in frequency for bilateral implants. Making only one electrode per implant more reliable for conveying spatial information could facilitate the design of the implant and might be sufficient to allow patients to take advantage of natural ITD cues. In that spirit, it would be interesting to test whether presenting a sound source with coherent ITDs on a few neighbouring auditory filters and incoherent ITDs in all the other filters is sufficient to observe a spatial unmasking effect.

IV. The neural representation of interaural time differences in gerbils is transformed from midbrain to cortex

This chapter is a reproduction of the journal article ‘The neural representation of interaural time differences in gerbils is transformed from midbrain to cortex’ by Lucile A.C. Belliveau, Dmitry R. Lyamzin, and Nicholas A. Lesica published in The Journal of Neuroscience, 10 December 2014, 34(50). This project was led by Nicholas Lesica and I contributed to performing the data collection in the IC and writing the manuscript.

1. Abstract

Interaural time differences (ITDs) are the dominant cue for the localization of low-frequency sounds. While much is known about the processing of ITDs in the auditory brainstem and midbrain, there have been relatively few studies of ITD processing in auditory cortex. In this study, we compared the neural representation of ITDs in the inferior colliculus (IC) and primary auditory cortex (A1) of gerbils. Our IC results were largely consistent with previous studies, with most cells responding maximally to ITDs that correspond to the contralateral edge of the physiological range. In A1, however, we found that preferred ITDs were distributed evenly throughout the physiological range without any contralateral bias. This difference in the distribution of preferred ITDs in IC and A1 had a major impact on the coding of ITDs at the population level: while a labeled-line decoder that considered the tuning of individual cells performed well on both IC and A1 responses, a two-channel decoder based on the overall activity in each hemisphere performed poorly on A1 responses relative to either labeled-line decoding of A1 responses or two-channel decoding of IC responses. These results suggest that the neural representation of ITDs in gerbils is transformed from IC to A1 and have important implications for how spatial location may be combined with other acoustic features for the analysis of complex auditory scenes.

2. Introduction

The ability to accurately localize sounds is critical for directing behavior, as well as for identifying and segregating individual sources within complex acoustic scenes (Cherry 1953; A. Bronkhorst 2000; C J Darwin 2008). The dominant cue for the localization of a low-frequency sound such as speech is the difference in its arrival time at the two ears, referred to as the interaural time difference or ITD (Wightman and Kistler 1992). ITD sensitivity in the mammalian brain arises in the medial superior olive (MSO) in the auditory brainstem where cells are sensitive to microsecond differences in the arrival time of inputs from the two ears. The spike rates of cells in the MSO and subsequent sub-cortical processing stages are modulated by ITD, with most cells responding preferentially to sounds with ITDs corresponding to locations in the contralateral hemifield (Grothe, Pecka, and McAlpine 2010).

While ITD processing in sub-cortical areas has been extensively studied, there have been relatively few studies of ITD processing in auditory cortex. It is clear that auditory cortex is necessary for ITD processing in both animals and humans, though lesions in either hemisphere cause a contralateral deficit in spatial processing in animals (Malhotra, Hall, and Lomber 2004a; W M Jenkins and Masterton 1982; W M Jenkins and Merzenich 1984), while right auditory cortex appears both necessary and sufficient for ITD processing in humans (Yamada et al. 1996; Tanaka, Hachisuka, and Ogata 1999). ITD tuning in primary auditory cortex (A1) was first reported several decades ago (J. F. Brugge et al. 1969; J. F. Brugge and Merzenich 1973), but the few studies in A1 with large samples that have been performed since have produced inconsistent results: a study in cats reported results similar to those in sub-cortical areas, with nearly all cells responding preferentially to ITDs corresponding to locations in the contralateral hemifield (Reale and Brugge 1990), while studies in chinchillas, rabbits and monkeys reported a weaker contralateral bias with preferred ITDs distributed more evenly across the physiological range (Fitzpatrick, Kuwada, and Batra 2000; Scott, Malone, and Semple 2009; Benson and Teas 1976). There have been no direct studies of single cell ITD sensitivity in human cortex, but recent EEG and MEG studies suggest a strong contralateral bias (Salminen et al. 2010; Magezi and Krumbholz 2010).

In this study, we characterize the neural representation of ITD in A1 of gerbils, one of the most widely used model species for studies of ITD processing. In gerbils, the vast majority of cells in sub-cortical structures have preferred ITDs corresponding to locations in the contralateral hemifield (Pecka et al. 2008; Siveke et al. 2006; Nicholas A Lesica, Lingner, and Grothe 2010; Spitzer and Semple 1995), consistent with a two-channel representation in which the ITD of a sound is encoded by the difference in the overall activity of the two brain hemispheres (D McAlpine, Jiang, and Palmer 2001). Here we show that the neural representation of ITDs is transformed between IC and A1 such that the preferred ITDs of A1 cells are distributed evenly throughout the physiological range without any contralateral bias. We examine the impact of this transformation on the population coding of ITDs and assess the ability of two-channel and labeled-line codes to account for gerbil behavioral acuity.

3. Methods

a. In vivo recordings

All procedures were approved under the UK Animals (Scientific Procedures) Act of 1986. 19 adult male gerbils (70-90 g, P60-120) were anesthetized for surgery with an initial injection of a mix of either fentanyl, medetomidine, and midazolam or ketamine and xylazine, and the same solution was infused continuously during recording. A small metal rod was mounted on the skull and used to secure the head of the animal in a stereotaxic device in a sound-attenuated chamber. A craniotomy was made over the inferior colliculus or the primary auditory cortex, an incision was made in the dura mater, and a multi-tetrode array (Neuronexus) was inserted into the brain. The array had four shanks spaced 0.2 mm apart, and each shank had two tetrodes spaced 0.15 mm apart. Recordings were made with a sampling rate of 25 kHz. Only recordings from the central nucleus of the IC and A1 were analyzed. Because the array covered a large area, recording sites in the central nucleus of the IC could be distinguished from those in other areas by comparison of their responses to tones (Aitkin et al. 1975b; Syka et al. 2000), and A1 could be distinguished from other fields based on the direction of the tonotopic gradient (Thomas et al. 1993). A1 recordings were made between 1 and 1.5 mm below the cortical surface (most likely layer V; see (Happel, Jeschke, and Ohl 2010)). We choose to record in layer V because we found the single-unit yield to be higher there than in layer IV in pilot experiments (we did not try other layers). Though it is difficult to say exactly why this would be the case, the fact that, relative to layer IV, layer V cells are large and sparsely packed, and spike with lower rates and less synchronously, may allow for single units to be more easily separated from multi-unit background. In both IC and A1, recordings were targeted to areas with low preferred frequencies.

b. Spike sorting

The procedure for the isolation of single-unit spikes consisted of 1) bandpass filtering each channel and the tetrode array between 500 and 5000 Hz, 2) whitening each tetrode, i.e. projecting the signals from the 4 channels into a space in which they are uncorrelated, 3) identifying potential spikes as snippets with energy (Choi, Jung, and Kim 2006b) that exceeded a threshold (with a minimum of 0.7 ms between potential spikes), 4) projecting each of the snippets into the space defined by the first three principal components for each channel, 5) identifying clusters of snippets within this space using KlustaKwik (http://klustakwik.sourceforge.net) and Klusters (Hazan, Zugaro, and Buzsáki 2006b), and 6) quantifying the likelihood that each cluster represented a single unit using isolation distance (Schmitzer-Torbert et al. 2005b). Isolation distance assumes that each cluster forms a multi-dimensional Gaussian cloud in feature space and measures, in terms of the standard deviation of the original cluster, the increase in the size of the cluster required to double the number of snippets within it. The number of snippets in the 'noise' cluster (multi-unit activity) for each tetrode was always at least as large as the number of spikes in any single-unit cluster. Only clusters with an isolation distance greater than 20 were classified as single units and included in our analysis.

c. Sound delivery

Sounds were generated with a 48 kHz sampling rate, attenuated, and delivered to speakers. Speakers (Etymotic ER2) coupled to tubes were inserted into both ear canals for sound presentation along with microphones for calibration. The frequency response of these speakers measured at the entrance of the ear canal was flat (± 5 dB) between 0.2 and 5 kHz. At each recording site, a sequence of tones with different frequencies and intensities with 5 ms cosine on and off ramps were presented to characterize frequency tuning. Speech and broadband noise were then presented at 60 dB SPL with 9 different ITDs spanning the physiological range for gerbils (± 160 μs in 40 μs steps) to characterize ITD tuning (with positive values of ITD denoting sounds leading at the ear contralateral to the recording site). These sounds were 500 ms in duration and were presented 32 or 64 times each in random order with a 500 ms pause between sounds and 2 ms cosine on and off ramps. Two different tokens of speech were used. Token 1 was presented to all cells in IC and A1 (n = 188 and 906, respectively). Token 2 was presented to all cells recorded in the left A1 under fentanyl, medetomidine, and midazolam (n = 517). Broadband noise was presented to a subset of cells recorded in the left A1 (n = 492) and all cells recorded in the right A1 (n = 100) and all IC cells.

d. Decoding ITD from spike rates

To decode responses based on spike rate alone, we used maximum likelihood decoding. The probability that a spike rate r was evoked by an ITD s is given by Bayes' rule as $p(s\,|\,r) = (\,p(r\,|\,s)\,p(s))\,/p(r)$. Because all ITDs were presented with equal probability, $p(s|r) \propto p(r|s)$. Thus, the ITD that is most likely to have caused a given response is simply $\arg\max_s p(r|s)$. We assumed that the distribution of spike rates evoked by a given ITD was Gaussian (with truncation at zero if necessary). This assumption improved performance in cross-validated testing. We did not place any constraints on the mean spike rates at each ITD, i.e. the shape of the ITD tuning curve. The significance of ITD tuning was assessed by decoding responses after randomizing the pairing of responses and ITDs (Monte Carlo resampling). ITD tuning was considered significant if decoding performance was more than 4 standard deviations above the mean performance for 100 different sets of shuffled responses.

To decode population responses with a labeled-line decoder, the joint probability of a set of spike rates from N cells, $p(r_1, r_2, \cdots, r_N|s)$, was computed as the product of the probabilities of the spike rate of each cell, $\prod_{n=1}^{N} p(r_n|s)$. Conditional independence was assumed because noise correlations were extremely weak (see Results). To decode population responses with a two-channel decoder, populations were split into two groups, and the responses for the second group were flipped with respect to ITD (as the majority of responses were recorded in the left A1 and our sample of responses from the right A1 was relatively small, we used only the responses from the left A1 for population decoding). The joint probability of the set of total spike rates in each hemisphere, $p(r_{left}, r_{right}|s)$, was then computed as the product of the probabilities of the total spike rate in each hemisphere, $p\left(r_{left}|s\right)p(r_{right}|s)$. For each population size tested, 100 different random subpopulations were drawn, with replacement, from the full sample of cells (bootstrap resampling). For the comparison with behavior, decoding performance was measured for responses to some ITDs that were not actually presented during the experiment. To measure decoding performance for these ITDs, we simulated responses (spike counts) by sampling from a Gaussian distribution with mean and variance taken as a linear interpolation between the values for the closest ITDs that were actually presented (i.e. it was assumed that the mean and variance of the spike rates for each cell were smooth functions of ITD). This approach was chosen to obtain a robust estimate of decoding performance for smaller ΔITDs than were actually presented. An alternative approach would have been to extrapolate to smaller ΔITDs from the performance at ΔITDs that were actually presented, but this would require an assumption about the shape of the performance curve for small ΔITDs, which is difficult to predict directly from our data.

e. Decoding ITD from spike times

To decode responses based on spike timing, we used the metric introduced by Victor and colleagues (J D Victor and Purpura 1996), which measures the distance between two spike trains as the overall cost of the set of operations required to transform one spike train into the other, with possible operations including the insertion of a spike, the deletion of a spike, and the time-shift of a spike (D. H. Goldberg et al. 2009). By changing the cost of time-shifting a spike relative to deleting the spike at one time and inserting it at another, the metric can be used to evaluate the distance between spike trains at different timescales. Decoding using this metric was performed as follows: 1) A single spike train was removed from the full set of all spike trains. 2) The distance between the removed spike train and each of the remaining spike trains in the set was computed across a range of timescales spaced logarithmically between 1 ms and 1 s. 3) For each timescale, the removed spike train was assigned to the sound for which its average distance to the remaining spike trains evoked by that sound was smallest. This process was repeated for all spike trains in the set to obtain a percent correct for each timescale, and the overall percent correct was taken as the maximum value across timescales. Decoding based on spike timing was considered significantly better than decoding based on spike rate alone if decoding performance based on spike timing was more than 4 standard deviations above the mean performance for decoding based on spike rate computed via bootstrap resampling.

4. Results

We made multi-tetrode recordings (Figure 1A) from populations of single units in the inferior colliculus (IC) and primary auditory cortex (A1) of anesthetized gerbils. Our methods for IC recordings have been described in detail previously (Garcia-Lazaro, Belliveau, and Lesica 2013b). For A1 recordings (Figure 1B), we aligned the shanks of the tetrode array along the rostrocaudal axis (approximately parallel to the tonotopic gradient in A1) and recorded from depths between 1 mm and 1.5 mm below the cortical surface (most likely layer V; see (Happel, Jeschke, and Ohl 2010)). The direction of the tonotopic gradient in A1 was evident in the multi-unit activity across tetrodes, as illustrated by the frequency responses areas (FRAs) for an example recording site shown in Figure 1C. We used a semi-automated clustering algorithm (see Methods) to isolate single units based on the first three principal components of their spike waveforms across each of the four channels of a tetrode. Clusters corresponding to single units (colors) and multi-unit ‘noise’ (gray) are shown for an example tetrode in Figure 1D (note that is a 2-D projection of a 12-D space). We quantified the quality of each cluster based on its isolation distance (Schmitzer-Torbert et al. 2005b) and set a threshold value of 20 for a cluster to be classified as a single unit (this value corresponded to detection of approximately 90% of spikes from a target neuron with a false alarm rate of approximately 1% in paired intracellular and tetrode recordings in hippocampus). The spike waveforms for two example single units with relatively low (24.5) and high (69.9) isolation distances are shown in Figure 1E. The median isolation distance across our sample of single units in A1 was 32.5 (Figure 1F).

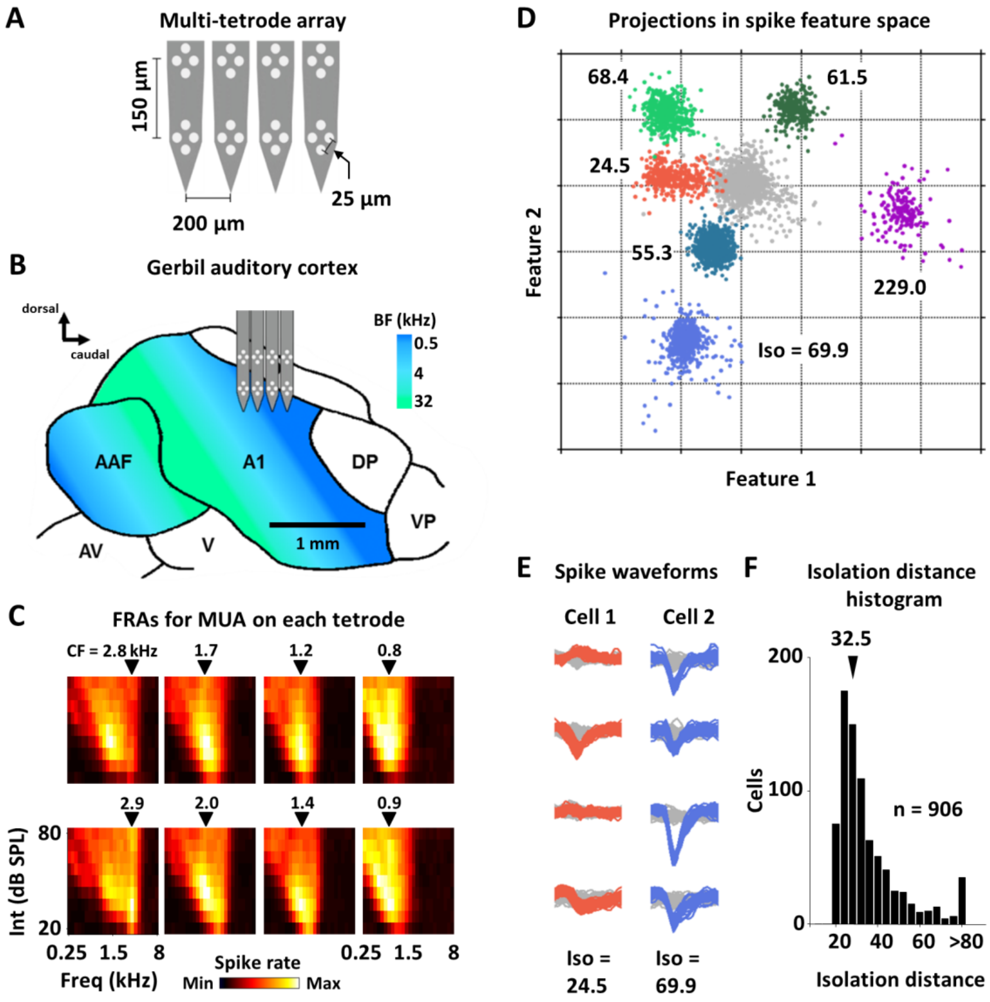

Figure 1: Multi-tetrode recordings in gerbil A1. **A.** A schematic diagram of the electrode arrangement on the multi-tetrode array. 32 electrodes were grouped into 8 tetrodes. **B.** A schematic diagram of the gerbil auditory cortex illustrating the alignment of the tetrode array with respect to the tonotopic gradient in A1 (modified from Thomas et al. (1993)). **C.** The frequency response areas (FRAs) for the multi-unit activity on each tetrode from a typical recording site. The first column shows the FRAs for the tetrodes on the most rostral shank, while the last column shows the FRAs for the tetrodes on the most caudal shank. Multi-unit activity was summed across the four electrodes on each tetrode. For each tetrode, the center frequency (CF) estimated from the FRA is indicated. **D.** Spikes from single -units were identified by projecting spike waveforms into principal component space (12 dimensions corresponding to 3 principal components for each electrode). An example two-dimensional projection that illustrates the isolation of different single-unit clusters is shown, along with the isolation distance of each cluster. Single-unit clusters are shown in color, undifferentiated multi-unit 'noise' is shown in gray. **E.** Spike waveforms for two single -units (overlaid on a sample of multi-unit 'noise' waveforms). **F.** A histogram of the isolation distances for all of the single-units in our A1 sample.

We targeted our recordings to areas with low preferred frequencies. The distributions of best frequencies (BFs) for our samples of IC and A1 cells are shown in Figure 2A. As our main goal was to compare the neural representation of ITDs with existing measures of gerbil behavioral acuity in the localization of a single broadband low-frequency sound source (Nicholas A Lesica, Lingner, and Grothe 2010), we restricted our analysis to responses to broadband sounds (speech and noise) with ITDs spanning only the physiological range (±160 μs in 40 μs steps; with positive values of ITD denoting sounds leading at the ear contralateral to the recording site). Note that we have chosen to use a range of ITDs that is slightly larger than that measured for gerbils by Maki and Furukawa (Katuhiro Maki and Furukawa 2005), as their measurements were made for frequencies above 1.5 kHz and the physiological range of ITDs tends to increase for lower frequencies (Rébillat et al. 2014).

The responses to speech with different ITDs from example cells with significant ITD tuning are shown in Figure 2B. Each column shows the FRA for one cell, along with raster plots for the responses to speech at 5 different ITDs spanning the physiological range and the tuning curve showing the mean spike rate as a function of ITD. To assess the strength of each cell's ITD tuning, we used a decoder to measure the accuracy with which the spike rate on a single trial could be used to infer which of 9 possible ITDs evoked it. We considered ITD tuning to be significant if decoding accuracy was more than four standard deviations above the mean performance for shuffled responses.

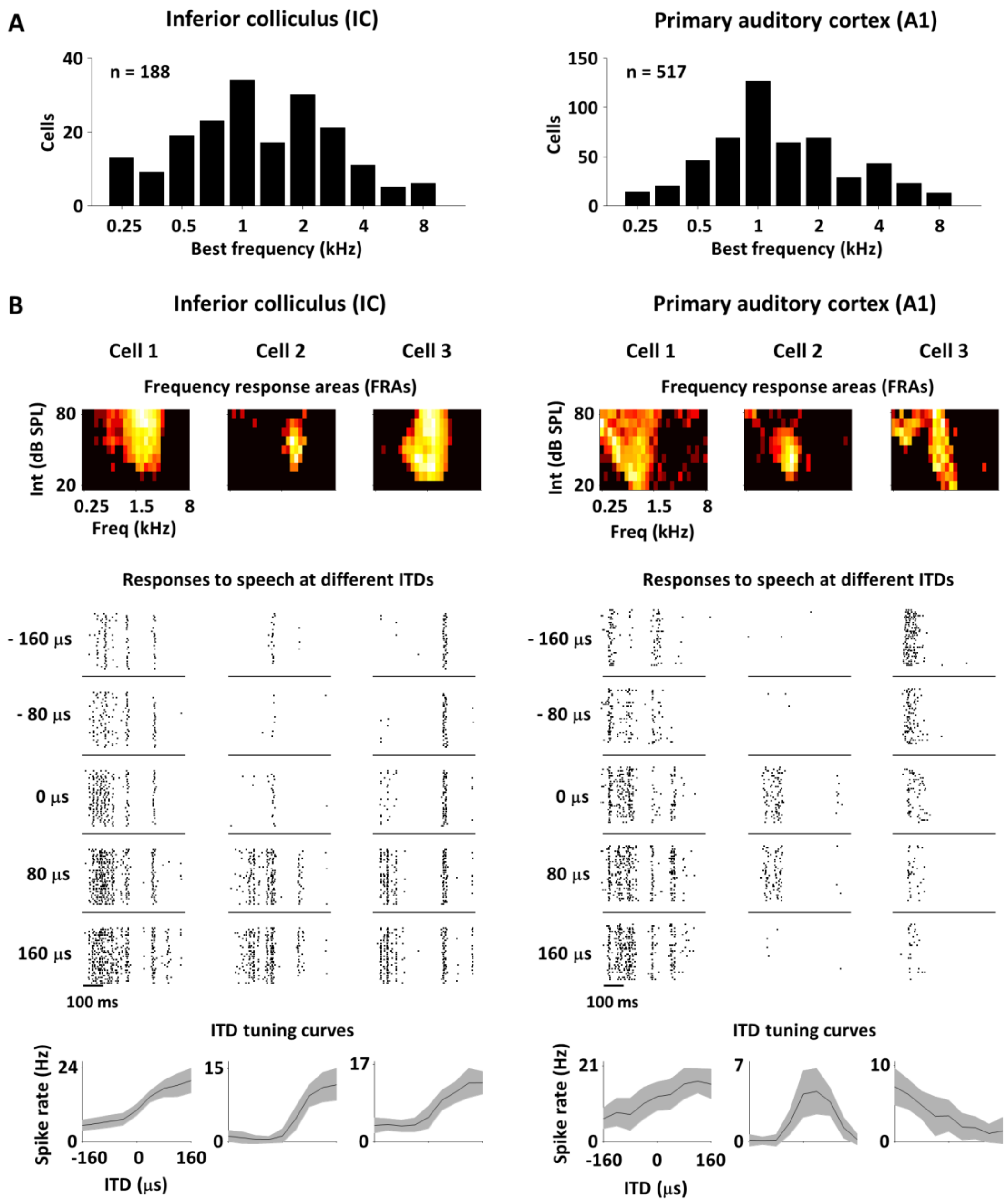

Figure 2: Responses to speech at different ITDs in gerbil IC and A1. **A.** The distribution of best frequencies in our samples of IC and A1 cells. **B.** The responses of example cells with significant ITD tuning in IC and A1. Each column shows the frequency response area for one cell, along with raster plots for the responses to speech at 5 different ITDs spanning the physiological range and the tuning curve showing the mean spike rate in response to speech as a function of ITD. The black line and gray bands on the tuning curve plots indicate the mean ±1 standard deviation.

The fraction of cells with significant ITD tuning for speech was higher in IC than in A1 (IC: 117/188 cells, 62%; A1: 239/517, 46%). This difference was not due to differences in the distribution of BFs in the two populations; the fraction of cells with significant tuning in our entire A1 sample was the same as that of random subsamples of A1 cells with BFs matched to those of our sample of IC cells (45 ± 3%). The fraction of cells with significant ITD tuning for broadband noise was also higher in IC than in A1 (IC: 134/188 cells, 71%; A1:

133/203, 65%). For all subsequent analyses in this study, only cells with significant ITD tuning were included.

a. Best ITDs in A1 are distributed evenly across the physiological range

To compare the representation of ITDs in IC and A1, we began by measuring the spike rate tuning curve for each cell’s response to speech at different ITDs. All of the example IC cells shown in Figure 2B responded most strongly to the ITD corresponding to the contralateral edge of the physiological range (+160 μs), while each example A1 cell had a different preferred ITD. These examples were representative of the IC and A1 in general; each row in the images in Figure 3A shows the ITD tuning curve for one cell (with cells sorted by best ITD), and the histograms in Figure 3B show the distributions of best ITDs across all cells.

The majority of IC cells in our sample had a best ITD of +160 μs, while the best ITDs in A1 were evenly distributed across the physiological range. To quantify the degree to which the distribution of best ITDs in each area was biased toward ipsilateral or contralateral values, we measured the percentage of cells with best ITDs in the contralateral hemifield (cells with best ITD = 0 were ignored). The distribution of best ITDs in the IC was strongly biased toward the contralateral side, with 83% of cells having best ITDs in the contralateral hemifield. In contrast, the distribution of best ITDs in A1 was unbiased, with only 53% of cells having best ITDs in the contralateral hemifield.

The best ITDs of A1 cells were unrelated to their BFs, as illustrated in left panel of Figure 3C, which shows the same ITD tuning curves for all A1 cells as in Figure 3A, but with cells sorted by BF. For any particular BF, there were cells with a range of different best ITDs, and, across the entire population, best ITD and BF were uncorrelated ($r = 0.03$, $p = 0.61$). The right panel of Figure 3C also shows the same tuning curves, but with the cells sorted by ITD decoding performance. For any particular level of performance, there were cells with a range of different best ITDs, though there was a weak, but significant, correlation between best ITD and decoding performance across the entire population ($r = 0.18$, $p = 0.003$), indicating that ITD tuning was slightly stronger for cells with best ITDs corresponding to contralateral locations than for cells with best ITDs corresponding to ipsilateral locations.

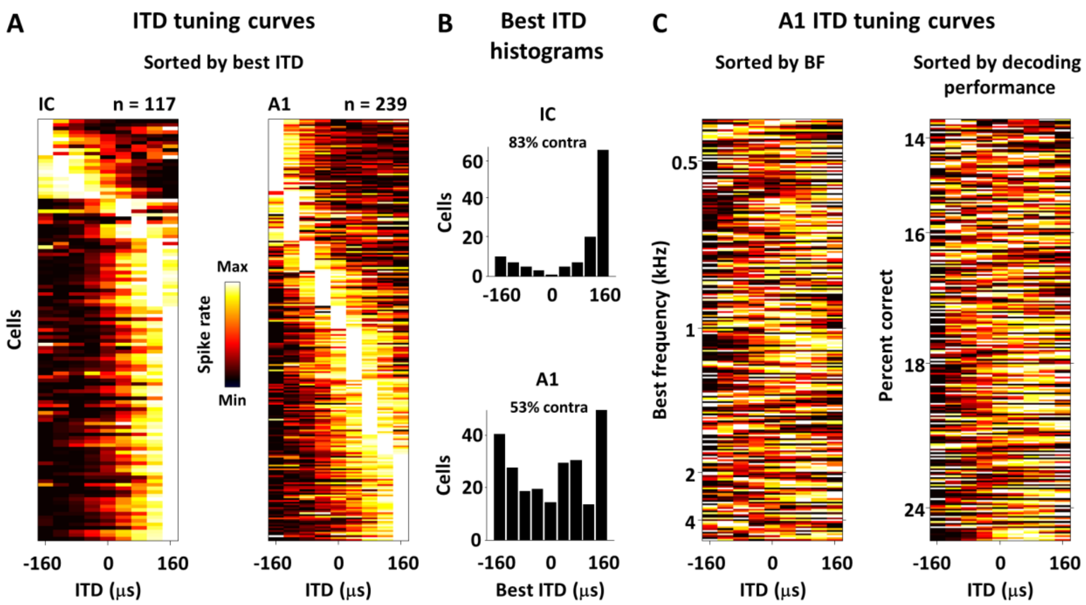

Figure 3: Best ITDs in A1 are distributed evenly across the physiological range. **A.** The ITD tuning curves for speech for all significantly tuned cells in our samples of IC and A1 cells. Each row shows the ITD tuning curve for one cell. All tuning curves were normalized to have the same maximum and minimum for plotting. Cells were sorted by best ITD for plotting. **B.** The histograms of the best ITDs for speech for all significantly tuned cells in our samples. **C.** The same A1 tuning curves shown in panel A with cells sorted either by BF or decoding performance. Decoding performance was measured as the percentage of single trial responses that were assigned to the correct ITD by a spike rate decoder (the chance level was 1/9).

b. ITD tuning is consistent across different sounds

We next investigated whether ITD tuning was consistent across sounds with different spectrotemporal properties by comparing ITD tuning curves for speech and broadband noise in both IC and A1, as well as for two different speech tokens in A1. Figure 4A shows the raster plots for the responses of an example A1 cell to the different sounds at 5 different ITDs, along with the tuning curves showing the mean spike rate as a function of ITD. The ITD tuning for this example cell was consistent across all three sounds, with the strongest responses evoked by ITDs near 0, corresponding to locations near the midline.

To quantify the similarity of ITD tuning across sounds for each cell, we measured the correlation coefficient between ITD tuning curves. As shown in Figure 4B, ITD tuning curves were highly similar across sounds for nearly all cells in both IC and A1; the median correlation between ITD tuning curves for speech and noise was 0.97 in IC and 0.87 in A1, and the median correlation between ITD tuning curves for two different segments of speech was 0.89 in A1. This suggests that the transformation of the representation of ITDs from IC to A1 is a general phenomenon that will be evident for any complex sound.

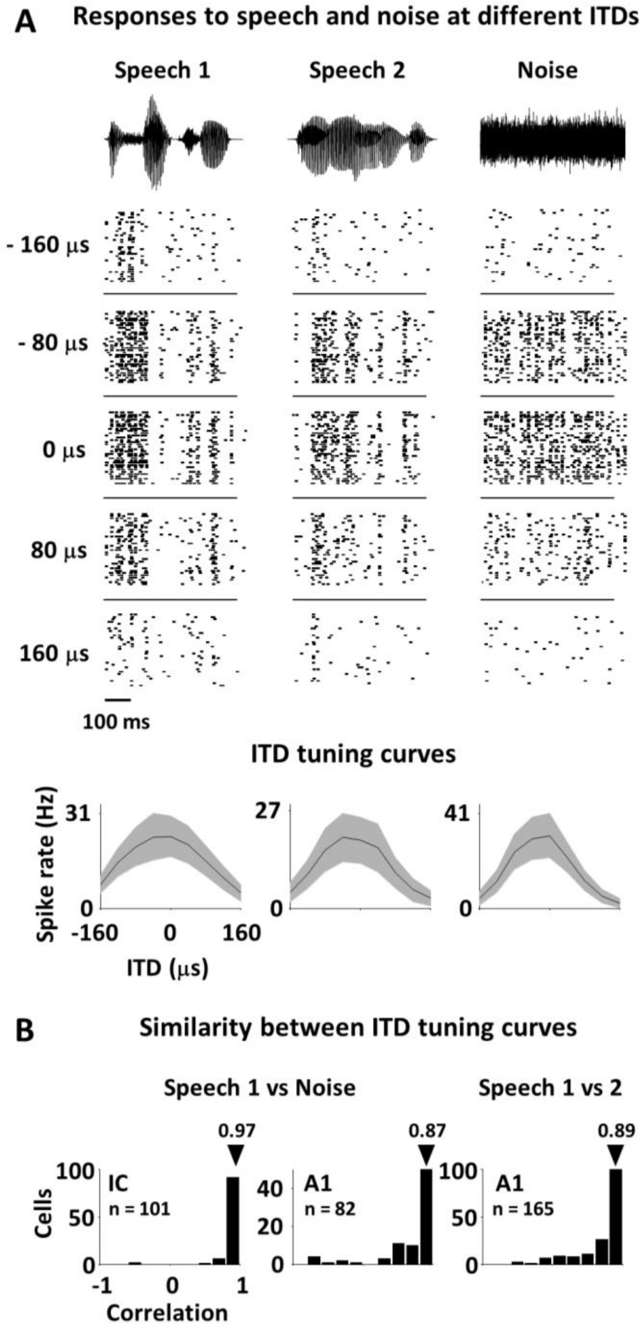

Figure 4: ITD tuning is consistent across different sounds. **A.** The responses of an example cell from A1 with significant ITD tuning for two different segments of speech and broadband noise. Each column shows the raster plots for the responses to one sound at 5 different ITDs spanning the physiological range and the tuning curve showing the mean spike rate as a function of ITD. The black line and gray bands on the tuning curve plots indicate the mean ±1 standard deviation. **B.** The histograms of the correlation coefficients between each cell's ITD tuning curves for speech segment 1 and broadband noise in the IC and A1, and the two speech segments in A1. Only cells with significant ITD tuning for both of the sounds being compared were included. The median value across each sample of cells is noted on each histogram.

c. Spike timing carries relatively little information about ITDs

Studies of IC and A1 responses have shown that for the coding of spectral notches and interaural level differences (ILDs) in high frequency sounds, spike timing contains substantial information beyond that in spike rate alone (Chase and Young 2006; Furukawa and Middlebrooks 2002), but the role of spike timing in coding ITDs in low-frequency sounds is not yet clear. To assess the role of spike timing in coding ITDs, we compared the performance of a decoder based on spike rate alone with that of a decoder that used a distance metric to consider the full spike train at the optimal timescale for each cell (J D Victor and Purpura 1996).

Figure 5A shows the raster plots for the responses of an example A1 cell to speech at 3 different ITDs. For this cell, the timing of some spiking events varied with ITD (see arrows), and considering spike timing in addition to spike rate resulted in a 50% improvement in decoding performance (21% correct for the timing decoder and 14% correct for the rate decoder, for 9 possible ITDs). This cell was, however, not typical of either IC or A1; as shown in Figure 5B, the improvement in decoding in both IC and A1 that resulted from considering spike timing in addition to spike rate was relatively small for both speech and noise. In IC, the improvement in the performance of the timing decoder over the rate decoder was significant for 55% of cells for speech and 60% of cells for noise, but the median improvement for those cells with significant improvement was only 11% for speech and 10% for noise. The improvement in A1 was higher than in IC for speech (50% of cells significant, median improvement 18%), and similar for noise (57% significant, median improvement 12%). These results suggest that spike timing is unlikely to play a major role in the coding of ITDs in either IC or A1.

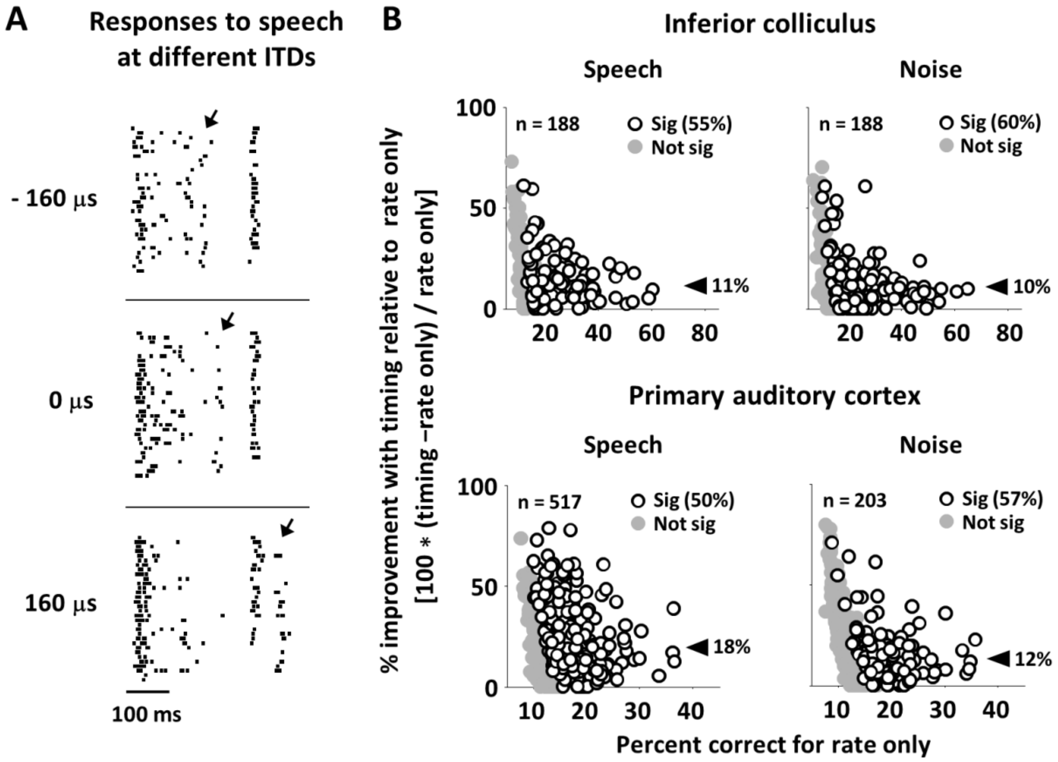

Figure 5: Spike timing carries relatively little information about ITDs. **A.** The responses of an example cell from A1 with significant information about ITD in spike timing. The raster plots show the responses to speech at three different ITDs. Spiking events that are unique to a particular ITD are marked with arrows. **B.** Scatter plots showing the percent improvement in decoding ITD that resulted from considering spike timing in addition to spike rate versus the percent correct for spike rate alone for responses to speech and noise in IC and A1. The values for cells for which the improvement was statistically significant are shown as black circles. The percent of cells for which the improvement was significant and the median improvement across those cells is noted on each histogram.

d. ITD tuning in A1 is qualitatively similar under different anesthesias

All of the responses described above were recorded under a mix of fentanyl, medetomidine, and midazolam (FMM). As the responses of neurons in gerbil A1 are known to vary with brain state (Ter-Mikaelian, Sanes, and Semple 2007b), we also made recordings under a mix of ketamine and xylazine (KX) to determine whether our observations of ITD tuning in A1 were dependent on our choice of anesthesia. In general, ITD tuning in A1 was much weaker under KX: only 39 of 289 cells (13%) had significant ITD tuning for speech (compared to 46% under FMM), and only 37 of 289 cells (12%) had significant ITD tuning for broadband noise (compared to 65% under FMM). However, as shown in Figure 6, the qualitative nature of ITD tuning under KX was similar to that under FMM: for those cells with significant ITD tuning for speech, best ITDs were distributed across the physiological range with no bias toward ITDs corresponding to locations in the contralateral hemifield (Figures 6A and B), and spike timing carried relatively little information about ITD (Figure 6C). All further analyses of ITD tuning in A1 described below were performed only on responses recorded under FMM.

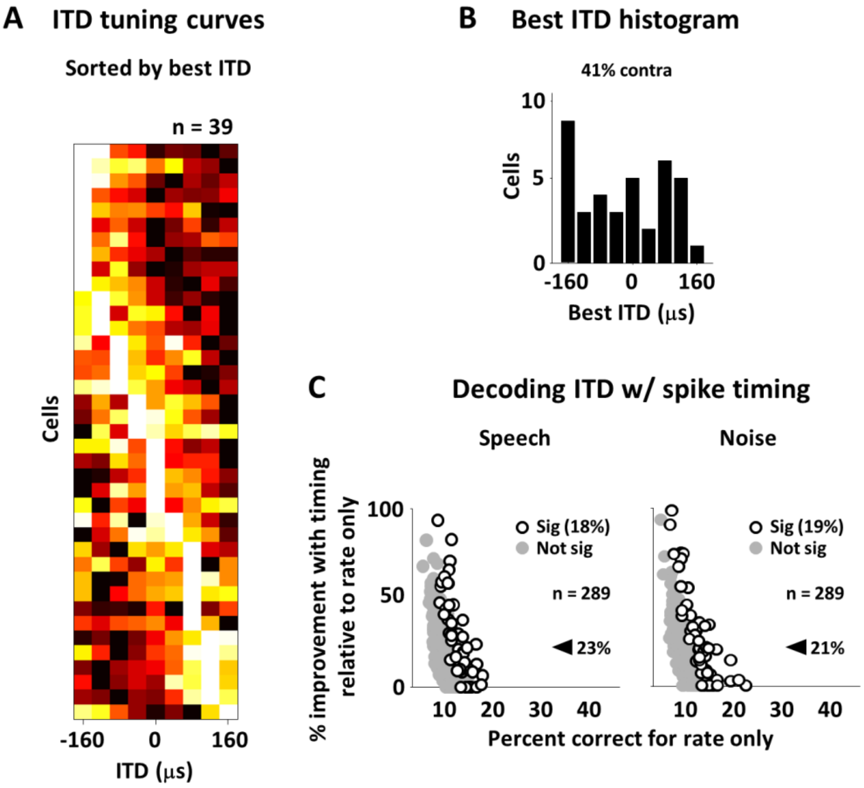

Figure 6: ITD tuning in A1 is qualitatively similar under different anesthesias. **A.** The ITD tuning curves for speech for all significantly tuned cells in our sample of A1 cells recorded under ketamine and xylazine, plotted as in figure 3A. **B.** The histogram of the best ITDs for speech for all significantly tuned cells in our sample of A1 cells recorded under ketamine and xylazine. **C.** Scatter plots showing the percent improvement in decoding ITD that resulted from considering spike timing in addition to spike rate versus the percent correct for spike rate alone for responses to speech and noise in A1 recorded under ketamine and xylazine, plotted as in figure 5B

e. ITD tuning in left and right A1 are similar

All of the responses described above were recorded from the left A1. To verify that ITD tuning in A1 was similar in both brain hemispheres, we made additional recordings from the right A1. ITD tuning in the right A1 was somewhat weaker than that in the left A1: 32 of 100 cells (32%) had significant ITD tuning for speech (compared to 46% in the left A1), and 46 of 100 cells (46%) had significant ITD tuning for broadband noise (compared to 65% in the left A1). As in the left A1, the best ITDs for cells in the right A1 were distributed across the physiological range with only a weak bias toward ITDs corresponding to locations in the contralateral hemifield (Figures 7A and B), and spike timing carried relatively little information about ITD (Figure 7C).

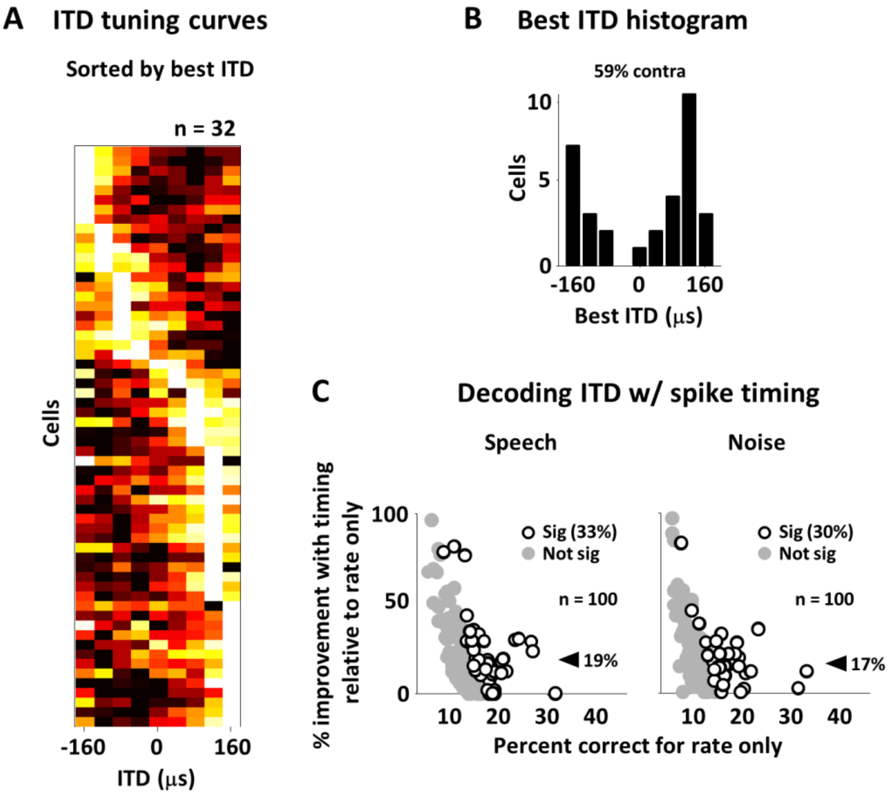

Figure 7: ITD tuning in left and right A1 are similar. **A.** The ITD tuning curves for speech for all significantly tuned cells in our sample of cells recorded in the right A1, plotted as in figure 3A**.** **B.** The histogram of the best ITDs for speech for all significantly tuned cells in our sample of cells recorded in the right A1. **C.** Scatter plots showing the percent improvement in decoding ITD that resulted from considering spike timing in addition to spike rate versus the percent correct for spike rate alone for responses to speech and noise in the right A1, plotted as in figure 5B.

f. Two-channel decoding of population responses in A1 results in a loss of information

The difference in the distributions of best ITDs in IC and A1 suggest a fundamental difference in the coding of ITDs at the population level. We considered two different population codes for ITD: a 'two-channel' code (Day and Delgutte 2013b; D McAlpine, Jiang, and Palmer 2001; Lüling et al. 2011) that considers only the total spike rate in each brain hemisphere (also known as a summed code (Nicholas A Lesica, Lingner, and Grothe 2010) or hemispheric code (Goodman, Benichoux, and Brette 2013)), and a 'labeled-line' code that considers the tuning of individual cells (also known as a distributed code (Nicholas A Lesica, Lingner, and Grothe 2010) or pattern code

We have shown previously that because the ITD tuning curves of most cells in gerbil IC are similar, a two-channel decoder performs almost as well as a labeled-line decoder at inferring the ITD of the sound that evoked a particular single trial population response (Nicholas A Lesica, Lingner, and Grothe 2010). However, for a population with more heterogeneous tuning curves, considering only the total spike rate in each hemisphere can impair decoding performance (Day and Delgutte 2013b; Goodman, Benichoux, and Brette 2013). We compared the performance of labeled-line and two-channel decoders on IC and A1 responses to speech and noise for populations of increasing size. Rather than constrain decoding to a particular computation (e.g. the difference in total spike rate between the two hemispheres), we used a maximum likelihood approach to infer which of 9 possible ITDs evoked each single trial population response based on the joint distributions of spike rates in each hemisphere (two-channel) or in individual cells (labeled line) (L. M. Miller and Recanzone 2009; Day and Delgutte 2013b). Because the noise correlations between pairs of simultaneously recorded cells in both IC and A1 were extremely weak (IC: 0.006±0.018, n = 9112; A1: 0.004±0.012, n = 3606), we assumed that the spike rates of individual cells were conditionally independent (Garcia-Lazaro, Belliveau, and Lesica 2013b).

As shown in Figures 8A and B, the labeled-line decoder performed well on both IC and A1 responses to speech and noise, with no loss of information between the two areas (IC: median performance of 94% correct for speech for the largest populations, 93% for noise; A1: 93% for speech, 91% for noise). In contrast, the performance of the two-channel decoder on A1 responses was much worse than its performance on IC responses (IC: 74% for speech, 73% for noise; A1: 49% for speech, 54% for noise). Thus, for a labeled line code based on the spike rates of individual cells, the information about ITDs that is present in IC is preserved in A1, but for a two-channel population code based on the total spike rate in each hemisphere, there is a substantial loss of information about ITDs between IC and A1.

It should be noted that although the performance of the two-channel decoder on A1 responses was relatively poor, it was still well above chance; although the ITD tuning curve peaks for A1 cells are evenly distributed throughout the physiological range, there is still a significant monotonic modulation of the total population spike rate with ITD, though this modulation is much weaker than that in the IC (see Figure 9A). It should also be noted that the labeled-line decoder significantly outperformed the two-channel decoder not only on A1 responses, but also on IC responses. This result is consistent with recent studies suggesting that the heterogeneity of tuning curves in IC can carry significant information about ITD (Day and Delgutte 2013b; Goodman, Benichoux, and Brette 2013).

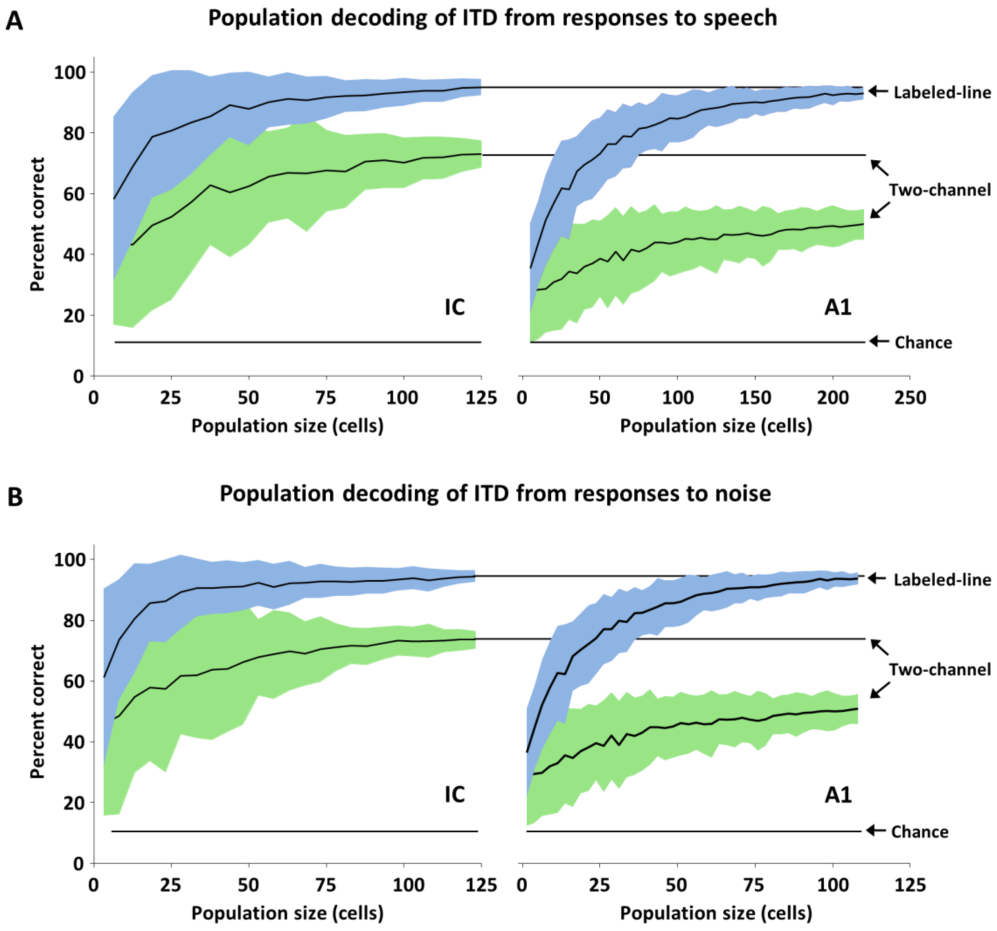

Figure 8: Two-channel and labeled-line decoding of ITD from population responses. **A.** The performance of labeled-line and two-channel decoders on IC and A1 responses to speech for populations of increasing size. Performance was measured as the percentage of single trial responses that were assigned to the correct ITD by the decoder. The chance level (1/9) is indicated. The black line and colored bands indicate the mean ±2 standard deviations of the performance for 100 different random subpopulations of each size drawn from the full sample of cells. **B.** The performance of labeled-line and two-channel decoders on IC and A1 responses to noise, plotted as in A.

g. Both two-channel and labeled-line decoding of population responses are sufficient to explain behavior

Our decoding results demonstrate that a labeled-line code carries substantially more information about ITDs than a two-channel code in gerbil A1. However, there is no guarantee that the code that is most informative about a particular sound feature is the one that underlies its perception, especially in the cortex where, presumably, the same neural circuitry is used to analyze many different features (Brette 2010). One approach to rule in or rule out different candidate codes for a particular feature is to determine whether they are sufficient to account for behavioral performance (Jacobs et al. 2009).

There have only been a few behavioral studies of sound localization in gerbils (R S Heffner and Heffner 1988; Maier and Klump 2006; Maier et al. 2008; Nicholas A Lesica, Lingner, and Grothe 2010; Lingner, Wiegrebe, and Grothe 2012b; Carney et al. 2011). Figure 9B shows the accuracy with which gerbils lateralized low-frequency noise bursts as a function of the difference in ITD approximated from the angle of separation between two speakers across the midline (median performance for 7 gerbils from Lesica et al. (2010)). We used the labeled-line and two-channel decoders described above to simulate the same behavioral task and infer which of two possible ITDs centered on the midline evoked each single trial population response to noise. Surprisingly, we found that both labeled-line and two-channel decoding were sufficient to reproduce this performance, even for A1 responses (see Figure 9B).

Though ability of gerbils to use ITDs to localize sounds has only been tested for pairs of sounds centered on the midline, it is known that behavioral acuity tends to decrease for sounds centered on more lateral locations in many mammals, for example, humans (Mossop and Culling 1998) and rabbits (Ebert et al. 2008). We examined the ability of the labeled-line and two-channel decoders to infer which of two possible ITDs evoked each single trial population response as a function of the ITD on which the sounds were centered (different Δ ITD were used for IC and A1 so that the performance of the labeled-line decoder for pairs of sounds centered on the midline was approximately 90% correct for both brain areas). As shown in Figure 9C, while decoder performance on IC responses decreased for more lateral sounds as expected, decoder performance on A1 responses was relatively consistent across the physiological range.

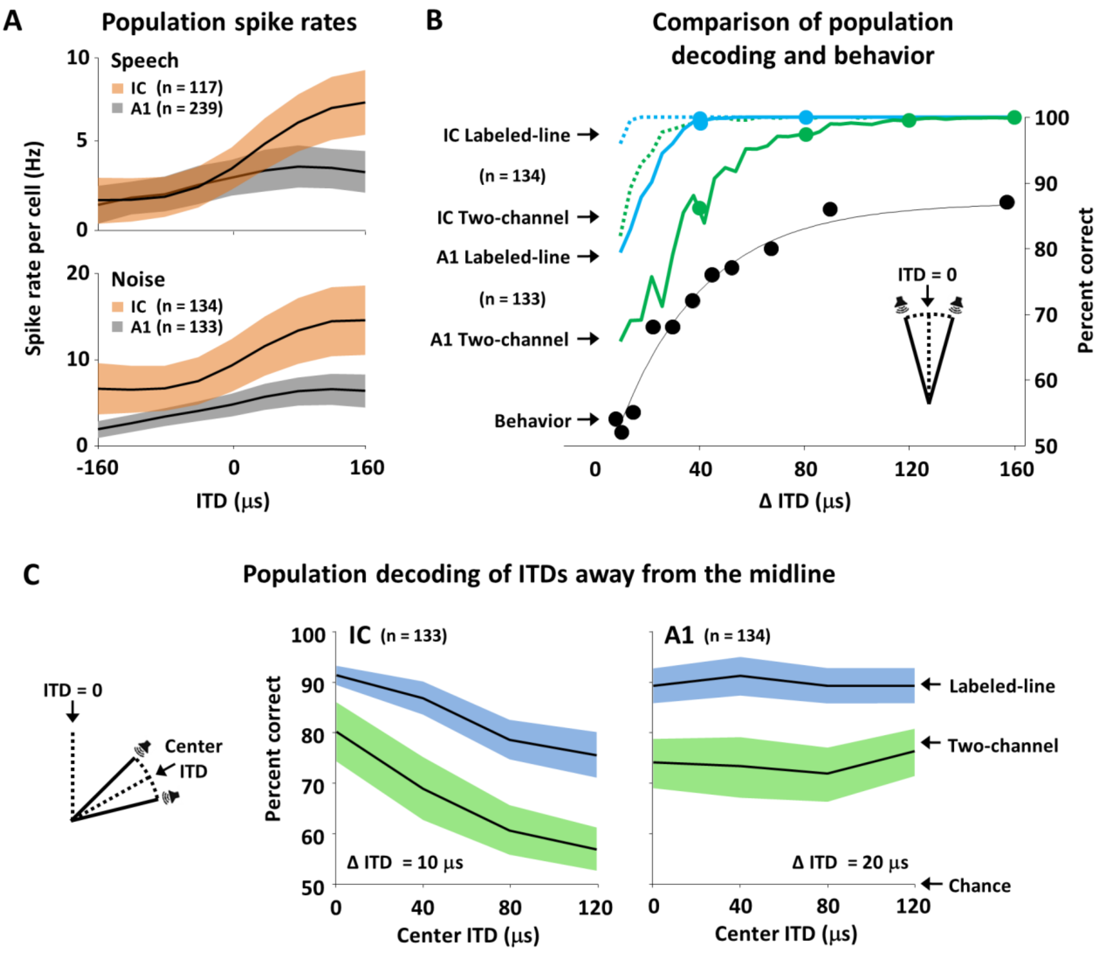

Figure 9: Comparing decoding of population responses and behavior**. A.** The tuning curve showing the total mean spike rate of our populations of cells in IC and A1 in response to speech and noise as a function of ITD. The black lines and colored bands on the tuning curve plots indicate the mean ±1 standard deviation. **B.** A comparison of the performance of labeled-line and two-channel decoders to gerbil behavior. The black dots show the actual performance of gerbils in lateralizing low-frequency noise bursts as a function of the difference in ITD approximated from the angle of separation between two speakers centered on the midline (median performance for 7 gerbils from Lesica et al. (2010)). The colored lines show the performance of labeled-line and two-channel decoders on IC and A1 responses to pairs of noise bursts centered on the midline. The dots on the colored lines correspond to ΔITDs that were actually tested experimentally, the remainder of the values were obtained after interpolating ITD tuning curves as described in the Methods. Performance for ΔITD = X μs was assessed by decoding responses to noise with ITD = ± (X/2) μs. **C.** The colored lines show the performance of labeled-line and two-channel decoders on IC and A1 responses to pairs of noise bursts centered on different ITDs. Performance for center ITD = X μs was assessed by decoding responses to noise with ITD = X ± 10 μs for the IC and ITD = X ± 20 μs for A1. The black line and colored bands indicate the mean ±2 standard deviations of the performance for 100 different bootstrap samples of cells from the full populations.

5. Discussion

We have shown that the neural representation of ITDs in gerbils is transformed from IC to A1. In the IC, we found that most cells responded maximally to ITDs corresponding to the contralateral edge of the physiological range, consistent with previous studies of ITD processing in different sub-cortical stages in gerbils (Pecka et al. 2008; Siveke et al. 2006; Nicholas A Lesica, Lingner, and Grothe 2010; Spitzer and Semple 1995). In contrast, the preferred ITDs of A1 cells were distributed evenly throughout the physiological range, with an equal number of cells preferring ITDs corresponding to ipsilateral and contralateral locations. This transformation in the distribution of preferred ITDs resulted in a loss of information between IC and A1 when using a two-channel decoder that considered only the total spike rate in each brain hemisphere, but not when using a labeled-line decoder that considered the tuning of individual cells. However, despite this loss of information, the two-channel decoder was still sufficient to reproduce gerbil behavioral performance.

Our analysis has revealed several aspects of the neural representation of ITDs in A1 that appear inconsistent with existing behavioral data. First, decoding of ITD from A1 activity results in performance that is far better than that observed behaviorally. Behavioral performance may be expected to be worse than decoder performance for activity from the early stages of peripheral processing, but differences of this degree in cortex are more surprising. Second, both left and right A1 appear to have a complete representation of azimuthal space (i.e. best ITDs span the full physiological range). Thus, it is unclear why a lesion to either the left or right A1 would cause a deficit in the localization of only contralateral sounds, as is the case in several mammals (Malhotra, Hall, and Lomber 2004a; W M Jenkins and Masterton 1982; W M Jenkins and Merzenich 1984; Kavanagh and Kelly 1987). Third, while decoder performance on IC responses decreased for sounds with ITDs corresponding to more lateral locations, consistent with behavioral observations in several mammals (Mossop and Culling 1998; Ebert et al. 2008), decoder performance on A1 responses was similar for ITDs corresponding to medial and lateral locations.

One possible explanation for the apparent mismatch between the neural representation of ITDs in A1 and the existing behavioral data in gerbils is that A1 is not actually required for or involved in the localization of single sound sources in a quiet background. While A1 seems to play a role in localization in most mammals that have been tested (A. J. King and Middlebrooks 2011), it does not appear necessary for sound localization in rats (J B Kelly and Kavanagh 1986). It may also be that if localization were tested under more difficult (e.g. reverberant) conditions, a better match between A1 activity and behavioral performance would become apparent. Another possible explanation for the disconnect between the neural representation of ITDs in A1 and behavior arises when one considers that the role of cortex is presumably to combine information about different acoustic features for the analysis of complex auditory scenes. While a given population of sub-cortical cells can be specialized for the processing of a particular feature, cortical populations may need to process multiple stimulus features simultaneously. Thus, the representation of ITDs in A1 (and/or the manner in which information from A1 is decoded in higher cortical areas, which may differ from the decoders we tested) may not be specialized for sound localization per se, but rather for the general processing of complex scenes, allowing, for example, sound sources from different locations to be processed by different subpopulations of cells (John C Middlebrooks and Bremen 2013) and facilitating the allocation of attentional resources to enhance or suppress activity related to a given source (Lee and Middlebrooks 2011b).

a. How does the transformation of ITD tuning between IC and A1 in gerbils compare with that in other species?

The change in the distribution of ITD tuning curve peaks from strongly contralaterally biased in IC to unbiased in A1 makes gerbils unique among species for which ITD tuning in midbrain and cortex has been systematically studied (Vonderschen and Wagner 2014). There is a clear transformation of the representation of ITDs in the auditory pathway of barn owls, but in the opposite direction, with narrowly-tuned midbrain inputs converging to form broadly-tuned channels in the forebrain (Vonderschen and Wagner 2009, 2012). The differences in ITD tuning between IC and A1 in other mammals are not as clear as those in gerbils, and results differ across species. Studies in cats have reported a relatively strong contralateral bias in both IC and A1 (Reale and Brugge 1990; Yin and Chan 1990a), while in rabbits, best ITDs span the physiological range with a relatively weak contralateral bias throughout the entire auditory pathway (Fitzpatrick, Kuwada, and Batra 2000), though tuning curves get sharper in more central stations (Fitzpatrick et al. 1997). Our results are perhaps most similar to those from chinchillas and primates; in chinchillas, there appears to be a strong contralateral bias in the IC (Bremen and Joris 2013), but only a relatively weak contralateral bias in A1 (Benson and Teas 1976). While there have been no systematic studies of ITD tuning in the primate midbrain, studies of spatial sensitivity in the IC suggest a strong contralateral bias (J M Groh et al. 2001; Jennifer M Groh, Kelly, and Underhill 2003; Zwiers, Versnel, and Van Opstal 2004), while ITD tuning in A1 exhibits a relatively weak contralateral bias (Scott, Malone, and Semple 2009).

Our results may also have implications for the study of ITD processing in humans. EEG and MEG studies in humans investigated the representation of ITDs based on measurements of the change in overall cortical activity in response to a change in ITD (Salminen et al. 2010; Magezi and Krumbholz 2010). In a labeled-line representation, a change in ITD in either direction should cause an increase in overall activity (as the sensory drive is directed toward an unadapted neuronal subpopulation), while in a two-channel representation, overall activity within a given hemisphere should increase with change in ITD in one direction, and decrease with a change in ITD in the other direction. Both studies found that the sign of the change in overall activity depended on the direction of the change in ITD, and, thus, argued for a two-channel representation. However, our A1 data demonstrate that a relatively coarse two-channel representation can coexist with a much more sensitive labeled-line representation and suggest that more detailed studies may be required to determine the true nature of the cortical representation of ITDs in humans.

b. What neural mechanisms underlie the transformation between IC and A1?

The neural circuitry that facilitates the transformation of the neural representation of ITDs between IC and A1 is not yet clear. It is possible to transform a distribution of best ITDs with a strong contralateral bias into an unbiased one either through the addition of inputs with opposing preferences or through the subtraction of inputs with similar preferences but different tuning curves (Jennifer M Groh, Kelly, and Underhill 2003). In principle, either of these possibilities could be implemented between IC and A1, even within a single brain hemisphere, using either the small subpopulation of cells in each IC with best ITDs corresponding to locations in the ipsilateral hemifield, or the heterogeneity of tuning curves in the majority of cells with best ITDs corresponding to locations in the contralateral hemifield.

If the transformation between IC and A1 does involve integration across the two brain hemispheres, it is likely through callosal connections (Budinger, Heil, and Scheich 2000), as the projections from the IC to the auditory thalamus and from the thalamus to

A1 are predominantly ipsilateral (Jeffery A Winer and Schreiner 2005b; Andrew J King and Schnupp 2011). The possibility that callosal connections play a role in shaping ITD tuning in A1 could explain why a unilateral cortical lesion results in a behavioral deficit for only contralateral locations (Malhotra, Hall, and Lomber 2004a; W M Jenkins and Masterton 1982; W M Jenkins and Merzenich 1984); without callosal inputs, the residual sensitivity in the remaining A1 may be for ITDs corresponding to ipsilateral locations only. Efforts to identify the neural circuitry that underlie the transformation from IC to A1 should begin by determining the highest stage at which the distribution of best ITDs still has a strong contralateral bias. While we cannot be certain of the layer in which our A1 recordings were made, it was most likely layer V; it is possible that the distribution of best ITDs in layer IV still has a strong contralateral bias (indeed, recent studies have suggested that the responses of cells in layer IV of A1 are simply amplified versions of their thalamic inputs (Li et al. 2013)), and that the transformation takes place between layer IV and layer II/III, or between layer II/III and layer V.

c. How does the ITD tuning in gerbil IC observed in this study compare with that observed previously?

The results of our population decoding analysis of IC responses differ somewhat from those of a similar analysis that we performed in a previous study (Nicholas A Lesica, Lingner, and Grothe 2010). In the previous study, the performance of the labeled-line and two-channel decoders was nearly identical, whereas in the current study, the performance of the labeled-line decoder was substantially better than that of the two-channel decoder. The difference in the performance of the two-channel decoder in the two studies is consistent with the differences in the distributions of best ITDs in the two populations of cells that were studied. In the previous study, we found that nearly all IC cells had best ITDs corresponding to the contralateral edge of the physiological range and, because of this homogeneity, very little information was lost when ignoring the tuning of individual cells and decoding only the total activity in the population. In the current study, however, 17% of the cells in our IC sample had best ITDs corresponding to locations in the ipsilateral hemifield, and this heterogeneity affected the performance of the two-channel decoder.

We believe that the difference in the distributions of best ITDs in the two studies is due to a difference in the fraction of the IC that was sampled during our recordings. In the previous study, we used a bundle of concentrically arranged electrodes that spanned a relatively small area and targeted the recordings to the rostromedial quadrant of the IC where the dominant input is provided by the MSO (Cant and Benson 2006), while in the current study, we used a much larger electrode array and sampled a larger fraction of the IC. Thus, the results of the present study are likely a more accurate reflection of the processing of ITDs in the IC as a whole, and are consistent with other recent studies suggesting that differences in the ITD tuning of IC cells carry significant information (Goodman, Benichoux, and Brette 2013; Day and Delgutte 2013b).

6. References

Agapiou, John P., and David McAlpine. 2008. “Low-Frequency Envelope Sensitivity Produces Asymmetric Binaural Tuning Curves.” *Journal of Neurophysiology* 100 (4): 2381–96. doi:10.1152/jn.90393.2008.

Aitkin, L. M., W. R. Webster, J. L. Veale, and D. C. Crosby. 1975a. “Inferior Colliculus. I. Comparison of Response Properties of Neurons in Central, Pericentral, and External Nuclei of Adult Cat.” *Journal of Neurophysiology* 38 (5): 1196–1207.

Aitkin, L M, W. R. Webster, J. L. Veale, and D. C. Crosby. 1975b. “Inferior Colliculus. I. Comparison of Response Properties of Neurons in Central, Pericentral, and External Nuclei of Adult Cat.” *Journal of Neurophysiology* 38 (5): 1196–1207.

Akeroyd, Michael A. 2004. “The across Frequency Independence of Equalization of Interaural Time Delay in the Equalization-Cancellation Model of Binaural Unmasking.” *The Journal of the Acoustical Society of America* 116 (2): 1135. doi:10.1121/1.1768959.

———. 2014. “An Overview of the Major Phenomena of the Localization of Sound Sources by Normal-Hearing, Hearing-Impaired, and Aided Listeners.” *Trends in Hearing* 18 (October): 2331216514560442. doi:10.1177/2331216514560442.

Akeroyd, Michael A., Stuart Gatehouse, and Julia Blaschke. 2007. “The Detection of Differences in the Cues to Distance by Elderly Hearing-Impaired Listeners.” *The Journal of the Acoustical Society of America* 121 (2): 1077–89.

Algazi, V. Ralph, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang. 2002. “Approximating the Head-Related Transfer Function Using Simple Geometric Models of the Head and Torso.” *The Journal of the Acoustical Society of America* 112 (5): 2053–64. doi:10.1121/1.1508780.

Arbogast, Tanya L., Christine R. Mason, and Gerald Kidd. 2002. “The Effect of Spatial Separation on Informational and Energetic Masking of Speech.” *The Journal of the Acoustical Society of America* 112 (5): 2086. doi:10.1121/1.1510141.

Baer, Thomas, and Brian C. J. Moore. 1994. “Effects of Spectral Smearing on the Intelligibility of Sentences in the Presence of Interfering Speech.” *The Journal of the Acoustical Society of America* 95 (4): 2277–80. doi:10.1121/1.408640.

Batra, R., S. Kuwada, and T. R. Stanford. 1989. “Temporal Coding of Envelopes and Their Interaural Delays in the Inferior Colliculus of the Unanesthetized Rabbit.” *Journal of Neurophysiology* 61 (2): 257–68.

Beaton, Randal, and Josef M. Miller. 1975. “Single Cell Activity in the Auditory Cortex of the Unanesthetized, Behaving Monkey: Correlation with Stimulus Controlled Behavior.” *Brain Research* 100 (3): 543–62. doi:10.1016/0006-8993(75)90157-2.

Belliveau, Lucile AC, Dmitry R. Lyamzin, and Nicholas A. Lesica. 2014. “The Neural Representation of Interaural Time Differences in Gerbils Is Transformed from Midbrain to Cortex.” *The Journal of Neuroscience* 34 (50): 16796–16808.

Benson, D. A., and R. D. Hienz. 1978. “Single-Unit Activity in the Auditory Cortex of Monkeys Selectively Attending Left vs. Right Ear Stimuli.” *Brain Research* 159 (2): 307–20. doi:10.1016/0006-8993(78)90537-1.

Benson, D. A., R. D. Hienz, and M. H. Goldstein Jr. 1981. “Single-Unit Activity in the Auditory Cortex of Monkeys Actively Localizing Sound Sources: Spatial Tuning and Behavioral Dependency.” *Brain Research* 219 (2): 249–67. doi:10.1016/0006-8993(81)90290-0.

Benson, D. A., and D. C. Teas. 1976. “Single Unit Study of Binaural Interaction in the Auditory Cortex of the Chinchilla.” *Brain Res.* 103: 313–38.

Bentler, Ruth, Yu-Hsiang Wu, Jerrica Kettel, and Richard Hurtig. 2008. “Digital Noise Reduction: Outcomes from Laboratory and Field Studies.” *International Journal of Audiology* 47 (8): 447–60. doi:10.1080/14992020802033091.

Bernstein, Leslie R., and Constantine Trahiotis. 1985. “Lateralization of Low-frequency, Complex Waveforms: The Use of Envelope-based Temporal Disparities.” *The Journal of the Acoustical Society of America* 77 (5): 1868–80. doi:10.1121/1.391938.

Bertoli, Sibylle, Katharina Staehelin, Elisabeth Zemp, Christian Schindler, Daniel Bodmer, and Rudolf Probst. 2009. “Survey on Hearing Aid Use and Satisfaction in Switzerland and Their Determinants.” *International Journal of Audiology* 48 (4): 183–95. doi:10.1080/14992020802572627.

Best, V., E. Ozmeral, F. J. Gallun, K. Sen, and B. G. Shinn-Cunningham. 2005. “Spatial Unmasking of Birdsong in Human Listeners: Energetic and Informational Factors.” *The Journal of the Acoustical Society of America* 118: 3766.

Best, Virginia, Simon Carlile, Norbert Kopčo, and André van Schaik. 2011. “Localization in Speech Mixtures by Listeners with Hearing Loss.” *The Journal of the Acoustical Society of America* 129 (5): EL210–EL215.

Beutelmann, Rainer, and Thomas Brand. 2006. “Prediction of Speech Intelligibility in Spatial Noise and Reverberation for Normal-Hearing and Hearing-Impaired Listeners.” *The Journal of the Acoustical Society of America* 120 (1): 331–42. doi:10.1121/1.2202888.

Beutelmann, Rainer, Thomas Brand, and Birger Kollmeier. 2009. “Prediction of Binaural Speech Intelligibility with Frequency-Dependent Interaural Phase Differences.” *The Journal of the Acoustical Society of America* 126 (3): 1359. doi:10.1121/1.3177266.

Bizley, Jennifer K., Fernando R. Nodal, Carl H. Parsons, and Andrew J. King. 2007. “Role of Auditory Cortex in Sound Localization in the Midsagittal Plane.” *Journal of Neurophysiology* 98 (3): 1763–74. doi:10.1152/jn.00444.2007.

Bizley, Jennifer K., Kerry M. M. Walker, Andrew J. King, and Jan W. H. Schnupp. 2013. “Spectral Timbre Perception in Ferrets: Discrimination of Artificial Vowels under Different Listening Conditions.” *The Journal of the Acoustical Society of America* 133 (1): 365–76. doi:10.1121/1.4768798.

Blauert, Jens. 1997. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press. http://books.google.co.uk/books?hl=en&lr=&id=wBiEKPhw7r0C&oi=fnd&pg=PA1&dq=3+!+.+Spatial+Hearing:+The+Psychophysics+of+Human+Sound+Localization+~+M&ots=4Iu44V7V6c&sig=BXGq9zC45Is_E0rYkGZa4HWuGpI.

Boymans, Monique, S. Theo Goverts, Sophia E. Kramer, Joost M. Festen, and Wouter A. Dreschler. 2008. “A Prospective Multi-Centre Study of the Benefits of Bilateral Hearing Aids.” *Ear and Hearing* 29 (6): 930–941.

Brand, Antje, Oliver Behrend, Torsten Marquardt, David McAlpine, and Benedikt Grothe. 2002. “Precise Inhibition Is Essential for Microsecond Interaural Time Difference Coding.” *Nature* 417 (6888): 543–47. doi:10.1038/417543a.

Bremen, Peter, and Philip X Joris. 2013. “Axonal Recordings from Medial Superior Olive Neurons Obtained from the Lateral Lemniscus of the Chinchilla (Chinchilla Laniger).” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (44): 17506–18. doi:10.1523/JNEUROSCI.1518-13.2013.

Brette, Romain. 2010. “On the Interpretation of Sensitivity Analyses of Neural Responses.” *The Journal of the Acoustical Society of America* 128 (5).

Brimijoin, W. Owen, William M. Whitmer, David McShefferty, and Michael A. Akeroyd. 2014. “The Effect of Hearing Aid Microphone Mode on Performance in an Auditory Orienting Task.” *Ear and Hearing* 35 (5): e204-212. doi:10.1097/AUD.0000000000000053.

Broadbent, D. E., and Peter Ladefoged. 1957. “On the Fusion of Sounds Reaching Different Sense Organs.” *The Journal of the Acoustical Society of America* 29 (6): 708–10. doi:10.1121/1.1909019.

Bronkhorst, Adelbert W. 2000. “The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions.” *Acta Acustica United with Acustica* 86 (1): 117–28.

Bronkhorst, AW. 2000. “The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions.” *Acta Acustica United with Acustica*.

Bronkhorst, and R. Plomp. 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): 1508–16. doi:10.1121/1.395906.

Brons, I., R. Houben, and W. A. Dreschler. 2014. “Effects of Noise Reduction on Speech Intelligibility, Perceived Listening Effort, and Personal Preference in Hearing-Impaired Listeners.” *Trends in Hearing* 18 (0). doi:10.1177/2331216514553924.

Brown, Andrew D., Francisco A. Rodriguez, Cory D. F. Portnuff, Matthew J. Goupell, and Daniel J. Tollin. 2016. “Time-Varying Distortions of Binaural Information by Bilateral Hearing Aids: Effects of Nonlinear Frequency Compression.” *Trends in Hearing* 20 (October). doi:10.1177/2331216516668303.

Brugge, J. F., N. A. Dubrovsky, L M Aitkin, and D. J. Anderson. 1969. “Sensitivity of Single Neurons in Auditory Cortex of Cat to Binaural Tonal Stimulation ; Effects of Varying Interaural Time and Intensity.” *Journal of Neurophysiology* 32: 1005–24.

Brugge, J. F., and M Merzenich. 1973. “Responses of Neurons in Auditory Cortex of the Macaque Monkey to Monaural and Binaural Stimulation.” *Journal of Neurophysiology* 36 (6): 1138–58.

Brugge, John F., and Michael M. Merzenich. 1973. “Response of Neurons in Auditory Cortex of the Macaque Monkey to Monaural and Binaural Stimulation.” *Journal of Neurophysiology*. http://psycnet.apa.org/psycinfo/1974-26872-001.

Brungart, Douglas S., and Nandini Iyer. 2012. “Better-Ear Glimpsing Efficiency with Symmetrically-Placed Interfering Talkers.” *The Journal of the Acoustical Society of America* 132 (4): 2545–56. doi:10.1121/1.4747005.

Budinger, E, P Heil, and H Scheich. 2000. “Functional Organization of Auditory Cortex in the Mongolian Gerbil (Meriones Unguiculatus). IV. Connections with Anatomically

Characterized Subcortical Structures.” *The European Journal of Neuroscience* 12 (7): 2452–74.

Byrne, Denis, and William Noble. 1998. “Optimizing Sound Localization with Hearing Aids.” *Trends in Amplification* 3 (2): 51. doi:10.1177/108471389800300202.

Caird, D. M., A. R. Palmer, and A. Rees. 1991. “Binaural Masking Level Difference Effects in Single Units of the Guinea Pig Inferior Colliculus.” *Hearing Research* 57 (1): 91–106. doi:10.1016/0378-5955(91)90078-N.

Cant, Nell B, and Christina G Benson. 2006. “Organization of the Inferior Colliculus of the Gerbil (Meriones Unguiculatus): Differences in Distribution of Projections from the Cochlear Nuclei and the Superior Olivary Complex.” *The Journal of Comparative Neurology* 495 (5): 511–28. doi:10.1002/cne.20888.

Carney, Laurel H, Srijata Sarkar, Kristina S Abrams, and Fabio Idrobo. 2011. “Sound-Localization Ability of the Mongolian Gerbil (Meriones Unguiculatus) in a Task with a Simplified Response Map.” *Hearing Research* 275 (1–2). Elsevier B.V.: 89–95. doi:10.1016/j.heares.2010.12.006.

Caspary, Donald M., Lynne Ling, Jeremy G. Turner, and Larry F. Hughes. 2008. “Inhibitory Neurotransmission, Plasticity and Aging in the Mammalian Central Auditory System.” *Journal of Experimental Biology* 211 (11): 1781–91. doi:10.1242/jeb.013581.

Chandrasekaran, Bharath, Erika Skoe, and Nina Kraus. 2013. “An Integrative Model of Subcortical Auditory Plasticity.” *Brain Topography* 27 (4): 539–52. doi:10.1007/s10548-013-0323-9.

Chase, Steven M, and Eric D Young. 2006. “Spike-Timing Codes Enhance the Representation of Multiple Simultaneous Sound-Localization Cues in the Inferior Colliculus.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 26 (15): 3889–98. doi:10.1523/JNEUROSCI.4986-05.2006.

Cherry, EC. 1953. “Some Experiments on the Recognition of Speech, with One and with Two Ears.” *The Journal of the Acoustical Society of America* 5 (25): 975–79.

Choi, Joon Hwan, Hae Kyung Jung, and Taejeong Kim. 2006a. “A New Action Potential Detector Using the MTEO and Its Effects on Spike Sorting Systems at Low Signal-to-Noise Ratios.” *Biomedical Engineering, IEEE Transactions on* 53 (4): 738–746.

———. 2006b. “A New Action Potential Detector Using the MTEO and Its Effects on Spike Sorting Systems at Low Signal-to-Noise Ratios.” *IEEE Transactions on Bio-Medical Engineering* 53 (4): 738–46. doi:10.1109/TBME.2006.870239.

Coffey, Charles S., Charles S. Ebert Jr., Allen F. Marshall, John D. Skaggs, Stephanie E. Falk, William D. Crocker, James M. Pearson, and Douglas C. Fitzpatrick. 2006. “Detection of Interaural Correlation by Neurons in the Superior Olivary Complex, Inferior Colliculus and Auditory Cortex of the Unanesthetized Rabbit.” *Hearing Research* 221 (1–2): 1–16. doi:10.1016/j.heares.2006.06.005.

Costalupes, J. A., E. D. Young, and D. J. Gibson. 1984. “Effects of Continuous Noise Backgrounds on Rate Response of Auditory Nerve Fibers in Cat.” *Journal of Neurophysiology* 51 (6): 1326–44.

Cranford, J. L. 1975. “Auditory Masking-Level Differences in the Cat.” *Journal of Comparative and Physiological Psychology* 89 (3): 219–23. doi:10.1037/h0076824.

Culling, J. F., and Q. Summerfield. 1995. “Perceptual Separation of Concurrent Speech Sounds: Absence of across-Frequency Grouping by Common Interaural Delay.” *The Journal of the Acoustical Society of America* 98: 785.

Culling, John F., H. Steven Colburn, and Matthew Spurchise. 2001. “Interaural Correlation Sensitivity.” *The Journal of the Acoustical Society of America* 110 (2): 1020–29. doi:10.1121/1.1383296.

Culling, John F., and C. J. Darwin. 1993. “Perceptual Separation of Simultaneous Vowels: Within and Across-formant Grouping by F0.” *The Journal of the Acoustical Society of America* 93 (6): 3454–67. doi:10.1121/1.405675.

Culling, John F., Kathryn I. Hodder, and Chaz Yee Toh. 2003. “Effects of Reverberation on Perceptual Segregation of Competing Voices.” *The Journal of the Acoustical Society of America* 114 (5): 2871–76. doi:10.1121/1.1616922.

Dannenbring, Gary L., and Albert S. Bregman. 1978. “Streaming vs. Fusion of Sinusoidal Components of Complex Tones.” *Perception & Psychophysics* 24 (4): 369–376.

Darwin, C J. 2008. “Listening to Speech in the Presence of Other Sounds.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1493): 1011–21. doi:10.1098/rstb.2007.2156.

Darwin, C. J., and R. W. Hukin. 1997. “Perceptual Segregation of a Harmonic from a Vowel by Interaural Time Difference and Frequency Proximity.” *The Journal of the Acoustical Society of America* 102: 2316.

———. 1998. “Perceptual Segregation of a Harmonic from a Vowel by Interaural Time Difference in Conjunction with Mistuning and Onset Asynchrony.” *The Journal of the Acoustical Society of America* 103: 1080.

———. 1999. “Auditory Objects of Attention: The Role of Interaural Time Differences.” *Journal of Experimental Psychology: Human Perception and Performance* 25 (3): 617.

Darwin, Christopher J., Douglas S. Brungart, and Brian D. Simpson. 2003. “Effects of Fundamental Frequency and Vocal-Tract Length Changes on Attention to One of Two Simultaneous Talkers.” *The Journal of the Acoustical Society of America* 114 (5): 2913–22. doi:10.1121/1.1616924.

Darwin, C.J. 2008. “Listening to Speech in the Presence of Other Sounds.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1493): 1011–21. doi:10.1098/rstb.2007.2156.

Dawes, Piers, Kevin J. Munro, Sridhar Kalluri, and Brent Edwards. 2013. “Unilateral and Bilateral Hearing Aids, Spatial Release from Masking and Auditory Acclimatization.” *The Journal of the Acoustical Society of America* 134 (1): 596–606.

Day, Mitchell L., and Bertrand Delgutte. 2013a. “Decoding Sound Source Location and Separation Using Neural Population Activity Patterns.” *The Journal of Neuroscience* 33 (40): 15837–47. doi:10.1523/JNEUROSCI.2034-13.2013.

Day, Mitchell L, and Bertrand Delgutte. 2013b. “Decoding Sound Source Location and Separation Using Neural Population Activity Patterns.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (40): 15837–47. doi:10.1523/JNEUROSCI.2034-13.2013.

Dillon, Harvey, Gitte Keidser, Anna O’brien, and Heidi Silberstein. 2003. “Sound Quality Comparisons of Advanced Hearing Aids.” *The Hearing Journal* 56 (4): 30–32.

Dirks, Donald D., and Richard H. Wilson. 1969. “The Effect of Spatially Separated Sound Sources on Speech Intelligibility.” *Journal of Speech and Hearing Research* 12 (1): 5–38.

Drennan, W. R., S. Gatehouse, and C. Lever. 2003. “Perceptual Segregation of Competing Speech Sounds: The Role of Spatial Location.” *The Journal of the Acoustical Society of America* 114: 2178.

Drennan, Ward R., Stuart Gatehouse, Patrick Howell, Dianne Van Tasell, and Steven Lund. 2005. “Localization and Speech-Identification Ability of Hearing-Impaired Listeners Using Phase-Preserving Amplification.” *Ear and Hearing* 26 (5): 461–72.

Dubno, Judy R., Donald D. Dirks, and Donald E. Morgan. 1984. “Effects of Age and Mild Hearing Loss on Speech Recognition in Noise.” *The Journal of the Acoustical Society of America* 76 (1): 87–96. doi:10.1121/1.391011.

Durlach, N. I. 1963. “Equalization and Cancellation Theory of Binaural Masking-Level Differences.” *The Journal of the Acoustical Society of America* 35 (8): 1206–18. doi:10.1121/1.1918675.

Durlach, N. I., K. J. Gabriel, H. S. Colburn, and C. Trahiotis. 1986. “Interaural Correlation Discrimination: II. Relation to Binaural Unmasking.” *The Journal of the Acoustical Society of America* 79 (5): 1548–57. doi:10.1121/1.393681.

Durlach, N. I., C. L. Thompson, and H. S. Colburn. 1981. “Binaural Interaction in Impaired Listeners: A Review of Past Research.” *Audiology* 20 (3): 181–211. doi:10.3109/00206098109072694.

Durlach, Nathaniel I. 1972. “Binaural Signal Detection-Equalization and Cancellation Theory.” http://ntrs.nasa.gov/search.jsp?R=19730045482.

Ebert, Charles S, Deidra A Blanks, Mihir R Patel, Charles S Coffey, Allen F Marshall, and Douglas C Fitzpatrick. 2008. “Behavioral Sensitivity to Interaural Time Differences in the Rabbit.” *Hearing Research* 235 (1–2): 134–42. doi:10.1016/j.heares.2007.11.003.

Edmonds, Barrie A., and John F. Culling. 2005a. “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.

———. 2005b. “The Spatial Unmasking of Speech: Evidence for within-Channel Processing of Interaural Time Delay.” *The Journal of the Acoustical Society of America* 117 (5): 3069. doi:10.1121/1.1880752.

Ehret, Günter, and Michael M. Merzenich. 1988. “Complex Sound Analysis (Frequency Resolution, Filtering and Spectral Integration) by Single Units of the Inferior Colliculus of the Cat.” *Brain Research Reviews* 13 (2): 139–63. doi:10.1016/0165-0173(88)90018-5.

Engineer, Crystal T., Claudia A. Perez, YeTing H. Chen, Ryan S. Carraway, Amanda C. Reed, Jai A. Shetake, Vikram Jakkamsetti, Kevin Q. Chang, and Michael P. Kilgard. 2008. “Cortical Activity Patterns Predict Speech Discrimination Ability.” *Nature Neuroscience* 11 (5): 603–8. doi:10.1038/nn.2109.

Fitzpatrick, D C, R Batra, T R Stanford, and S Kuwada. 1997. “A Neuronal Population Code for Sound Localization.” *Nature* 388 (6645): 871–74. doi:10.1038/42246.

Fitzpatrick, D C, S Kuwada, and R Batra. 2000. “Neural Sensitivity to Interaural Time Differences: Beyond the Jeffress Model.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 20 (4): 1605–15.

Fritz, Jonathan B., Mounya Elhilali, Stephen V. David, and Shihab A. Shamma. 2007. “Auditory Attention—focusing the Searchlight on Sound.” *Current Opinion in Neurobiology* 17 (4): 437–455.

Fritz, Jonathan B., Mounya Elhilali, and Shihab A. Shamma. 2007. “Adaptive Changes in Cortical Receptive Fields Induced by Attention to Complex Sounds.” *Journal of Neurophysiology* 98 (4): 2337–2346.

Furman, Adam C., Sharon G. Kujawa, and M. Charles Liberman. 2013. “Noise-Induced Cochlear Neuropathy Is Selective for Fibers with Low Spontaneous Rates.” *Journal of Neurophysiology* 110 (3): 577–86. doi:10.1152/jn.00164.2013.

Furukawa, Shigeto, and John C Middlebrooks. 2002. “Cortical Representation of Auditory Space: Information-Bearing Features of Spike Patterns.” *Journal of Neurophysiology* 87 (4): 1749–62. doi:10.1152/jn.00491.2001.

Garcia-Lazaro, Jose A., Lucile A. C. Belliveau, and Nicholas A. Lesica. 2013a. “Independent Population Coding of Speech with Sub-Millisecond Precision.” *The Journal of Neuroscience* 33 (49): 19362–72. doi:10.1523/JNEUROSCI.3711-13.2013.

Garcia-Lazaro, Jose A, Lucile A C Belliveau, and Nicholas A Lesica. 2013b. “Independent Population Coding of Speech with Sub-Millisecond Precision.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (49): 19362–72. doi:10.1523/JNEUROSCI.3711-13.2013.

Glasberg, Brian R, and Brian C.J Moore. 1990. “Derivation of Auditory Filter Shapes from Notched-Noise Data.” *Hearing Research* 47 (1–2): 103–38. doi:10.1016/0378-5955(90)90170-T.

Goldberg, David H, Jonathan D Victor, Esther P Gardner, and Daniel Gardner. 2009. “Spike Train Analysis Toolkit: Enabling Wider Application of Information-Theoretic Techniques to Neurophysiology.” *Neuroinformatics* 7 (3): 165–78. doi:10.1007/s12021-009-9049-y.

Goldberg, J. M., and P. B. Brown. 1969. “Response of Binaural Neurons of Dog Superior Olivary Complex to Dichotic Tonal Stimuli: Some Physiological Mechanisms of Sound Localization.” *Journal of Neurophysiology* 32 (4): 613–36.

Goodman, Dan FM, Victor Benichoux, and Romain Brette. 2013. “Decoding Neural Responses to Temporal Cues for Sound Localization.” *eLife* 2 (January): e01312. doi:10.7554/eLife.01312.

Griffin, Sarah J., Leslie R. Bernstein, Neil J. Ingham, and David McAlpine. 2005. “Neural Sensitivity to Interaural Envelope Delays in the Inferior Colliculus of the Guinea Pig.” *Journal of Neurophysiology* 93 (6): 3463–78. doi:10.1152/jn.00794.2004.

Groh, J M, A S Trause, A M Underhill, K R Clark, and S Inati. 2001. “Eye Position Influences Auditory Responses in Primate Inferior Colliculus.” *Neuron* 29 (2): 509–18.

Groh, Jennifer M, Kristin A Kelly, and Abigail M Underhill. 2003. “A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus.” *Journal of Cognitive Neuroscience* 15 (8): 1217–31. doi:10.1162/089892903322598166.

Grothe, Benedikt, Michael Pecka, and D McAlpine. 2010. “Mechanisms of Sound Localization in Mammals.” *Physiological Reviews*, 983–1012. doi:10.1152/physrev.00026.2009.

Hall, Joseph W., Richard S. Tyler, and Mariano A. Fernandes. 1983. “Monaural and Binaural Auditory Frequency Resolution Measured Using Bandlimited Noise and Notched-noise Masking.” *The Journal of the Acoustical Society of America* 73 (3): 894–98. doi:10.1121/1.389013.

Happel, Max F K, Marcus Jeschke, and Frank W Ohl. 2010. “Spectral Integration in Primary Auditory Cortex Attributable to Temporally Precise Convergence of Thalamocortical and Intracortical Input.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 30 (33): 11114–27. doi:10.1523/JNEUROSCI.0689-10.2010.

Harkins, Judith, and Paula Tucker. 2007. “An Internet Survey of Individuals With Hearing Loss Regarding Assistive Listening Devices.” *Trends in Amplification* 11 (2): 91–100. doi:10.1177/1084713807301322.

Harris, David M., Robert V. Shannon, Russell Snyder, and Edward Carney. 1997. “Multi-Unit Mapping of Acoustic Stimuli in Gerbil Inferior Colliculus.” *Hearing Research* 108 (1–2): 145–56. doi:10.1016/S0378-5955(97)00047-6.

Hartley, David, Elena Rochtchina, Philip Newall, Maryanne Golding, and Paul Mitchell. 2010. “Use of Hearing Aids and Assistive Listening Devices in an Older Australian Population.” *Journal of the American Academy of Audiology* 21 (10): 642–53. doi:10.3766/jaaa.21.10.4.

Häusler, R., S. Colburn, and E. Marr. 1983. “Sound Localization in Subjects with Impaired Hearing.” *Acta Oto-Laryngologica* 96 (sup400): 1–62. doi:10.3109/00016488309105590.

Hawkins, David B., and William S. Yacullo. 1984. “Signal-to-Noise Ratio Advantage of Binaural Hearing Aids and Directional Microphones under Different Levels of Reverberation.” *Journal of Speech and Hearing Disorders* 49 (3): 278–86. doi:10.1044/jshd.4903.278.

Hazan, Lynn, Michaël Zugaro, and György Buzsáki. 2006a. “Klusters, NeuroScope, NDManager: A Free Software Suite for Neurophysiological Data Processing and Visualization.” *Journal of Neuroscience Methods* 155 (2): 207–216.

———. 2006b. “Klusters, NeuroScope, NDManager: A Free Software Suite for Neurophysiological Data Processing and Visualization.” *Journal of Neuroscience Methods* 155 (2): 207–16. doi:10.1016/j.jneumeth.2006.01.017.

Hebrank, Jack, and D. Wright. 1974. “Spectral Cues Used in the Localization of Sound Sources on the Median Plane.” *The Journal of the Acoustical Society of America* 56 (6): 1829–1834.

Heffner, Henry E., Rickye S. Heffner, Christopher Contos, and Tonie Ott. 1994. “Audiogram of the Hooded Norway Rat.” *Hearing Research* 73 (2): 244–247.

Heffner, R S, and H E Heffner. 1988. “Sound Localization and Use of Binaural Cues by the Gerbil (Meriones Unguiculatus).” *Behavioral Neuroscience* 102 (3): 422–28.

Heffner, Rickye S., and Henry E. Heffner. 1985. “Hearing Range of the Domestic Cat.” *Hearing Research* 19 (1): 85–88. doi:10.1016/0378-5955(85)90100-5.

Heijden, Marcel van der, and Constantine Trahiotis. 1998. “Binaural Detection as a Function of Interaural Correlation and Bandwidth of Masking Noise: Implications for Estimates of Spectral Resolution.” *The Journal of the Acoustical Society of America* 103 (3): 1609–14. doi:10.1121/1.421295.

———. 1999. “Masking with Interaurally Delayed Stimuli: The Use of ‘internal’ Delays in Binaural Detection.” *The Journal of the Acoustical Society of America* 105 (1): 388–99. doi:10.1121/1.424628.

Hienz, Robert D., Cynthia M. Aleszczyk, and Bradford J. May. 1996. “Vowel Discrimination in Cats: Acquisition, Effects of Stimulus Level, and Performance in Noise.” *The Journal of the Acoustical Society of America* 99 (6): 3656–68. doi:10.1121/1.414980.

Hienz, Robert D, Phaedra Stiles, and Bradford J May. 1998. “Effects of Bilateral Olivocochlear Lesions on Vowel Formant Discrimination in Cats.” *Hearing Research* 116 (1–2): 10–20. doi:10.1016/S0378-5955(97)00197-4.

Hienz, RobertD., Cynthia M. Aleszczyk, and Bradford J. May. 1996. “Vowel Discrimination in Cats: Thresholds for the Detection of Second Formant Changes in the Vowel /ɛ/.” *The Journal of the Acoustical Society of America* 100 (2): 1052–58. doi:10.1121/1.416291.

Hirsh. 1948a. “The Influence of Interaural Phase on Interaural Summation and Inhibition.” *The Journal of the Acoustical Society of America*, a, 20 (4): 536–44. doi:10.1121/1.1906407.

———. 1948b. “Binaural Summation and Interaural Inhibition as a Function of the Level of Masking Noise.” *The American Journal of Psychology*, b, 61 (2): 205–13. doi:10.2307/1416966.

Hocherman, S., D. A. Benson, M. H. Goldstein Jr., H. E. Heffner, and R. D. Hienz. 1976. “Evoked Unit Activity in Auditory Cortex of Monkeys Performing a Selective Attention Task.” *Brain Research* 117 (1): 51–68. doi:10.1016/0006-8993(76)90555-2.

Hoesel, Richard J. M. van, and Richard S. Tyler. 2003. “Speech Perception, Localization, and Lateralization with Bilateral Cochlear Implants.” *The Journal of the Acoustical Society of America* 113 (3): 1617–30. doi:10.1121/1.1539520.

Holube, Inga, Martin Kinkel, and Birger Kollmeier. 1998. “Binaural and Monaural Auditory Filter Bandwidths and Time Constants in Probe Tone Detection Experiments.” *The Journal of the Acoustical Society of America* 104 (4): 2412–25. doi:10.1121/1.423773.

Hubel, D. H., C. O. Henson, A. Rupert, and R. Galambos. 1959. “Attention Units in the Auditory Cortex.” *Science (New York, N.Y.)* 129 (3358): 1279–80.

Hukin, R. W., and C. J. Darwin. 1995. “Effects of Contralateral Presentation and of Interaural Time Differences in Segregating a Harmonic from a Vowel.” *The Journal of the Acoustical Society of America* 98 (3): 1380–87. doi:10.1121/1.414348.

Jacobs, Adam L, Gene Fridman, Robert M Douglas, Nazia M Alam, Peter E Latham, Glen T Prusky, and Sheila Nirenberg. 2009. “Ruling out and Ruling in Neural Codes.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (14): 5936–41. doi:10.1073/pnas.0900573106.

Jeffress, Lloyd A. 1948. “A Place Theory of Sound Localization.” *Journal of Comparative and Physiological Psychology* 41 (1): 35.

Jenkins, W M, and R B Masterton. 1982. “Sound Localization: Effects of Unilateral Lesions in Central Auditory System.” *Journal of Neurophysiology* 47 (6): 987–1016.

Jenkins, W M, and M M Merzenich. 1984. “Role of Cat Primary Auditory Cortex for Sound-Localization Behavior.” *Journal of Neurophysiology* 52 (5): 819–47.

Jenkins, William M., and R. Bruce. 1982. “Sound Localization: Effects of Unilateral Lesions in Central Auditory System.” *Journal of Neurophysiology* 47 (6): 987–1016.

Jiang, Dan, David McAlpine, and Alan R. Palmer. 1997. “Responses of Neurons in the Inferior Colliculus to Binaural Masking Level Difference Stimuli Measured by Rate-versus-Level Functions.” *Journal of Neurophysiology* 77 (6): 3085–3106.

Joris, P. X., and T. C. Yin. 1995. “Envelope Coding in the Lateral Superior Olive. I. Sensitivity to Interaural Time Differences.” *Journal of Neurophysiology* 73 (3): 1043–62.

Joris, Philip X. 2003. “Interaural Time Sensitivity Dominated by Cochlea-Induced Envelope Patterns.” *The Journal of Neuroscience* 23 (15): 6345–50.

Joris, Philip X., Marcel van der Heijden, Dries H. Louage, Bram Van de Sande, and Cindy Van Kerckhoven. 2005. “Dependence of Binaural and Cochlear ‘best Delays’ on Characteristic Frequency.” In *Auditory Signal Processing*, 477–483. Springer. http://link.springer.com/content/pdf/10.1007/0-387-27045-0_58.pdf.

Joris, Philip, and Tom C.T. Yin. 2007. “A Matter of Time: Internal Delays in Binaural Processing.” *Trends in Neurosciences* 30 (2): 70–78. doi:10.1016/j.tins.2006.12.004.

Kavanagh, GL, and JB Kelly. 1987. “Contribution of Auditory Cortex to Sound Localization by the Ferret (Mustela Putorius).” *J Neurophysiol* 57 (6).

Keidser, Gitte, Anna O’Brien, Jens-Uwe Hain, Margot McLelland, and Ingrid Yeend. 2009. “The Effect of Frequency-Dependent Microphone Directionality on Horizontal

Localization Performance in Hearing-Aid Users." *International Journal of Audiology* 48 (11): 789–803. doi:10.3109/14992020903036357.

Keidser, Gitte, Kristin Rohrseitz, Harvey Dillon, Volkmar Hamacher, Lyndal Carter, Uwe Rass, and Elizabeth Convery. 2006. "The Effect of Multi-Channel Wide Dynamic Range Compression, Noise Reduction, and the Directional Microphone on Horizontal Localization Performance in Hearing Aid Wearers." *International Journal of Audiology* 45 (10): 563–79. doi:10.1080/14992020600920804.

Kelly, J B, and G L Kavanagh. 1986. "Effects of Auditory Cortical Lesions on Pure-Tone Sound Localization by the Albino Rat." *Behavioral Neuroscience* 100 (4): 569–75.

Kelly, Jack B., Gerard L. Kavanagh, and James C. H. Dalton. 1986. "Hearing in the Ferret (Mustela Putorius): Thresholds for Pure Tone Detection." *Hearing Research* 24 (3): 269–75. doi:10.1016/0378-5955(86)90025-0.

Kidd, G., C. Mason, V. Richards, F. Gallun, and N. Durlach. 2007. "Informational Masking." *Auditory Perception of Sound Sources*, 143–189.

Kidd, Gerald, Christine R. Mason, Andrew Brughera, and William M. Hartmann. 2005. "The Role of Reverberation in Release from Masking Due to Spatial Separation of Sources for Speech Identification." *Acta Acustica United with Acustica* 91 (3): 526–36.

Kidd, Gerald, Christine R. Mason, Tanya L. Rohtla, and Phalguni S. Deliwala. 1998. "Release from Masking due to Spatial Separation of Sources in the Identification of Nonspeech Auditory Patterns." *The Journal of the Acoustical Society of America* 104 (1): 422–31. doi:10.1121/1.423246.

King, A. J., and J. C. Middlebrooks. 2011. "Cortical Representation of Auditory Space." In *The Auditory Cortex*, edited by J. A. Winer and C. E. Schreiner, 329–40. New York, New York, USA: Springer.

King, Andrew J, and Jan W H Schnupp. 2011. *The Auditory Cortex*. Edited by Jeffery A Winer and Christoph E Schreiner. *Current Biology : CB*. Vol. 17. New York: Springer. doi:10.1016/j.cub.2007.01.046.

King, Andrew J., Jan WH Schnupp, and Timothy P. Doubell. 2001. "The Shape of Ears to Come: Dynamic Coding of Auditory Space." *Trends in Cognitive Sciences* 5 (6): 261–270.

Kittel, Malte, Eva Wagner, and Georg M Klump. 2002. "An Estimate of the Auditory-Filter Bandwidth in the Mongolian Gerbil." *Hearing Research* 164 (1–2): 69–76. doi:10.1016/S0378-5955(01)00411-7.

Klasen, Thomas J., Tim Van den Bogaert, Marc Moonen, and Jan Wouters. 2007. "Binaural Noise Reduction Algorithms for Hearing Aids That Preserve Interaural Time Delay Cues." *IEEE Transactions on Signal Processing* 55 (4): 1579–1585.

Klatt, Dennis H. 1980. "Software for a Cascade/Parallel Formant Synthesizer." *The Journal of the Acoustical Society of America* 67 (3): 971–95. doi:10.1121/1.383940.

Klumpp, R. G., and H. R. Eady. 1956. "Some Measurements of Interaural Time Difference Thresholds." *The Journal of the Acoustical Society of America* 28 (5): 859–860.

Kuwada, SHIGEYUKI, TERRENCE R. Stanford, and RANJAN Batra. 1987. "Interaural Phase-Sensitive Units in the Inferior Colliculus of the Unanesthetized Rabbit: Effects of Changing Frequency." *Journal of Neurophysiology* 57 (5): 1338–1360.

Langford, Ted L., and Lloyd A. Jeffress. 1964. "Effect of Noise Crosscorrelation on Binaural Signal Detection." *The Journal of the Acoustical Society of America* 36 (8): 1455–58. doi:10.1121/1.1919224.

Larson, V. D., D. W. Williams, W. G. Henderson, L. E. Luethke, L. B. Beck, D. Noffsinger, R. H. Wilson, et al. 2000. "Efficacy of 3 Commonly Used Hearing Aid Circuits: A

Crossover Trial. NIDCD/VA Hearing Aid Clinical Trial Group." *JAMA* 284 (14): 1806–13.

Lee, Chen-Chung, and John C Middlebrooks. 2011a. "Auditory Cortex Spatial Sensitivity Sharpens during Task Performance." *Nature Neuroscience* 14 (1): 108–14. doi:10.1038/nn.2713.

———. 2011b. "Auditory Cortex Spatial Sensitivity Sharpens during Task Performance." *Nature Neuroscience* 14 (1). Nature Publishing Group: 108–14. doi:10.1038/nn.2713.

Lee, Chen-Chung, and John C. Middlebrooks. 2013. "Specialization for Sound Localization in Fields A1, DZ, and PAF of Cat Auditory Cortex." *Journal of the Association for Research in Otolaryngology* 14 (1): 61–82. doi:10.1007/s10162-012-0357-9.

Lesica, N. A., A. Lingner, and B. Grothe. 2010. "Population Coding of Interaural Time Differences in Gerbils and Barn Owls." *Journal of Neuroscience* 30 (35): 11696–702. doi:10.1523/JNEUROSCI.0846-10.2010.

Lesica, Nicholas A, Andrea Lingner, and Benedikt Grothe. 2010. "Population Coding of Interaural Time Differences in Gerbils and Barn Owls." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 30 (35): 11696–702. doi:10.1523/JNEUROSCI.0846-10.2010.

Levitt, H. 1971. "Transformed Up-Down Methods in Psychoacoustics." *The Journal of the Acoustical Society of America* 49 (2B): 467–77. doi:10.1121/1.1912375.

Levitt, H., and L. R. Rabiner. 1967. "Binaural Release From Masking for Speech and Gain in Intelligibility." *The Journal of the Acoustical Society of America*, a, 42 (3): 601–8. doi:10.1121/1.1910629.

Li, Ling-yun, Ya-tang Li, Mu Zhou, Huizhong W Tao, and Li I Zhang. 2013. "Intracortical Multiplication of Thalamocortical Signals in Mouse Auditory Cortex." *Nature Neuroscience* 16 (9): 1179–81. doi:10.1038/nn.3493.

Licklider, J. C. R. 1948. "The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise." *The Journal of the Acoustical Society of America* 20 (2): 150–59. doi:10.1121/1.1906358.

Lin, Frank R., Kristine Yaffe, Jin Xia, Qian-Li Xue, Tamara B. Harris, Elizabeth Purchase-Helzner, Suzanne Satterfield, Hilsa N. Ayonayon, Luigi Ferrucci, and Eleanor M. Simonsick. 2013. "Hearing Loss and Cognitive Decline Among Older Adults." *JAMA Internal Medicine* 173 (4). doi:10.1001/jamainternmed.2013.1868.

Lingner, Andrea, Lutz Wiegrebe, and Benedikt Grothe. 2012a. "Sound Localization in Noise by Gerbils and Humans." *JARO - Journal of the Association for Research in Otolaryngology* 13 (2): 237–48. doi:10.1007/s10162-011-0301-4.

———. 2012b. "Sound Localization in Noise by Gerbils and Humans." *Journal of the Association for Research in Otolaryngology : JARO* 13 (2): 237–48. doi:10.1007/s10162-011-0301-4.

Litovsky, Ruth Y., H. Steven Colburn, William A. Yost, and Sandra J. Guzman. 1999. "The Precedence Effect." *The Journal of the Acoustical Society of America* 106 (4): 1633–54. doi:10.1121/1.427914.

Litovsky, Ruth Y., Gary L. Jones, Smita Agrawal, and Richard van Hoesel. 2010. "Effect of Age at Onset of Deafness on Binaural Sensitivity in Electric Hearing in Humans." *The Journal of the Acoustical Society of America* 127 (1): 400–414. doi:10.1121/1.3257546.

Loizou, Philipos C., Yi Hu, Ruth Litovsky, Gongqiang Yu, Robert Peters, Jennifer Lake, and Peter Roland. 2009. "Speech Recognition by Bilateral Cochlear Implant Users in a

Cocktail-Party Setting.” *The Journal of the Acoustical Society of America* 125 (1): 372–83. doi:10.1121/1.3036175.

Lorenzi, C., S. Gatehouse, and C. Lever. 1999b. “Sound Localization in Noise in Hearing-Impaired Listeners.” *The Journal of the Acoustical Society of America* 105 (6): 3454–3463.

Lorenzi, Christian, Stuart Gatehouse, and Catherine Lever. 1999a. “Sound Localization in Noise in Normal-Hearing Listeners.” *The Journal of the Acoustical Society of America* 105 (3): 1810–1820.

Lüling, Hannes, Ida Siveke, Benedikt Grothe, and Christian Leibold. 2011. “Frequency-Invariant Representation of Interaural Time Differences in Mammals.” *PLoS Computational Biology* 7 (3): e1002013. doi:10.1371/journal.pcbi.1002013.

Maddox, Ross K, Cyrus P Billimoria, Ben P Perrone, Barbara G Shinn-Cunningham, and Kamal Sen. 2012. “Competing Sound Sources Reveal Spatial Effects in Cortical Processing.” *PLoS Biology* 10 (5): e1001319. doi:10.1371/journal.pbio.1001319.

Magezi, David A, and Katrin Krumbholz. 2010. “Evidence for Opponent-Channel Coding of Interaural Time Differences in Human Auditory Cortex.” *Journal of Neurophysiology* 104 (4): 1997–2007. doi:10.1152/jn.00424.2009.

Maier, Julia K, Teresa Kindermann, Benedikt Grothe, and Georg M Klump. 2008. “Effects of Omni-Directional Noise-Exposure during Hearing Onset and Age on Auditory Spatial Resolution in the Mongolian Gerbil (Meriones Unguiculatus) -- a Behavioral Approach.” *Brain Research* 1220 (July): 47–57. doi:10.1016/j.brainres.2008.01.083.

Maier, Julia K., and Georg M. Klump. 2006. “Resolution in Azimuth Sound Localization in the Mongolian Gerbil (Meriones Unguiculatus).” *The Journal of the Acoustical Society of America* 119 (2): 1029. doi:10.1121/1.2159429.

Maki, K., and S. Furukawa. 2005. “The Responses of Neurons in the Gerbil Inferior Colliculus to Virtual Acoustic Space Stimuli.” *Acoustical Science and Technology* 26 (1): 82–84.

Maki, Katuhiro, and Shigeto Furukawa. 2005. “Acoustical Cues for Sound Localization by the Mongolian Gerbil, Meriones Unguiculatus.” *The Journal of the Acoustical Society of America* 118 (2): 872. doi:10.1121/1.1944647.

Malhotra, Shveta, Amee J Hall, and Stephen G Lomber. 2004a. “Cortical Control of Sound Localization in the Cat: Unilateral Cooling Deactivation of 19 Cerebral Areas.” *Journal of Neurophysiology* 92 (3): 1625–43. doi:10.1152/jn.01205.2003.

Malhotra, Shveta, Amee J. Hall, and Stephen G. Lomber. 2004b. “Cortical Control of Sound Localization in the Cat: Unilateral Cooling Deactivation of 19 Cerebral Areas.” *Journal of Neurophysiology* 92 (3): 1625–43. doi:10.1152/jn.01205.2003.

Malone, Brian J., Brian H. Scott, and Malcolm N. Semple. 2002. “Context-Dependent Adaptive Coding of Interaural Phase Disparity in the Auditory Cortex of Awake Macaques.” *Journal of Neuroscience* 22 (11): 4625–38.

Marquardt, Torsten, and David McAlpine. 2009. “Masking with Interaurally ‘double-Delayed’ Stimuli: The Range of Internal Delays in the Human Brain.” *The Journal of the Acoustical Society of America* 126 (6): EL177-EL182. doi:10.1121/1.3253689.

May, Bradford J., Glenn S. Le Prell, and Murray B. Sachs. 1998. “Vowel Representations in the Ventral Cochlear Nucleus of the Cat: Effects of Level, Background Noise, and Behavioral State.” *Journal of Neurophysiology* 79 (4): 1755–67.

McAlpine, D, D Jiang, and A R Palmer. 2001. “A Neural Code for Low-Frequency Sound Localization in Mammals.” *Nature Neuroscience* 4 (4): 396–401. doi:10.1038/86049.

McAlpine, David, Dan Jiang, and Alan R. Palmer. 1996. “Binaural Masking Level Differences in the Inferior Colliculus of the Guinea Pig.” *The Journal of the Acoustical Society of America*, b, 100 (1): 490–503. doi:10.1121/1.415862.

———. 2001. “A Neural Code for Low-Frequency Sound Localization in Mammals.” *Nature Neuroscience* 4 (4): 396–401. doi:10.1038/86049.

McCormack, Abby, and Heather Fortnum. 2013. “Why Do People Fitted with Hearing Aids Not Wear Them?” *International Journal of Audiology* 52 (5): 360–68. doi:10.3109/14992027.2013.769066.

McFadden, Dennis, and Edward G. Pasanen. 1976. “Lateralization at High Frequencies Based on Interaural Time Differences.” *The Journal of the Acoustical Society of America* 59 (3): 634–39. doi:10.1121/1.380913.

Metzger, Ryan R., Nathaniel T. Greene, Kristin Kelly Porter, and Jennifer M. Groh. 2006. “Effects of Reward and Behavioral Context on Neural Activity in the Primate Inferior Colliculus.” *The Journal of Neuroscience* 26 (28): 7468–76. doi:10.1523/JNEUROSCI.5401-05.2006.

Mick, Paul, Ichiro Kawachi, and Frank R. Lin. 2014. “The Association between Hearing Loss and Social Isolation in Older Adults.” *Otolaryngology -- Head and Neck Surgery* 150 (3): 378–84. doi:10.1177/0194599813518021.

Mickey, Brian J., and John C. Middlebrooks. 2003. “Representation of Auditory Space by Cortical Neurons in Awake Cats.” *The Journal of Neuroscience* 23 (25): 8649–63.

Middlebrooks, John C., and Peter Bremen. 2013. “Spatial Stream Segregation by Auditory Cortical Neurons.” *The Journal of Neuroscience* 33 (27): 10986–1. doi:10.1523/JNEUROSCI.1065-13.2013.

Middlebrooks, John C, and Peter Bremen. 2013. “Spatial Stream Segregation by Auditory Cortical Neurons.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (27): 10986–1. doi:10.1523/JNEUROSCI.1065-13.2013.

Middlebrooks, JOHN C., and JOHN D. Pettigrew. 1981. “Functional Classes of Neurons in Primary Auditory Cortex of the Cat Distinguished by Sensitivity to Sound Location.” *J Neurosci* 1 (1): 107220.

Miller, J. M., D. Sutton, B. Pfingst, A. Ryan, R. Beaton, and G. Gourevitch. 1972. “Single Cell Activity in the Auditory Cortex of Rhesus Monkeys: Behavioral Dependency.” *Science* 177 (4047): 449–51. doi:10.1126/science.177.4047.449.

Miller, Lee M, and Gregg H Recanzone. 2009. “Populations of Auditory Cortical Neurons Can Accurately Encode Acoustic Space across Stimulus Intensity.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (14): 5931–35. doi:10.1073/pnas.0901023106.

Moore, Brian CJ. 1996. “Perceptual Consequences of Cochlear Hearing Loss and Their Implications for the Design of Hearing Aids.” *Ear and Hearing* 17 (2): 133–161.

Mossop, J E, and J F Culling. 1998. “Lateralization of Large Interaural Delays.” *The Journal of the Acoustical Society of America* 104 (3 Pt 1): 1574–79.

Mulrow, Cynthia D. 1990. “Quality-of-Life Changes and Hearing Impairment: A Randomized Trial.” *Annals of Internal Medicine* 113 (3): 188. doi:10.7326/0003-4819-113-3-188.

Neher, Tobias, Søren Laugesen, Niels Søgaard Jensen, and Louise Kragelund. 2011. “Can Basic Auditory and Cognitive Measures Predict Hearing-Impaired Listeners’ Localization and Spatial Speech Recognition Abilities? A.” *The Journal of the Acoustical Society of America* 130 (3): 1542–1558.

Noble, William, Denis Byrne, and Bernadette Lepage. 1994. “Effects on Sound Localization of Configuration and Type of Hearing Impairment.” *The Journal of the Acoustical Society of America* 95 (2): 992–1005.

Nodal, F. R., V. M. Bajo, C. H. Parsons, J. W. Schnupp, and A. J. King. 2008. “Sound Localization Behavior in Ferrets: Comparison of Acoustic Orientation and Approach-to-Target Responses.” *Neuroscience*, From Cochlea to Cortex: Recent Advances in Auditory Neuroscience, 154 (1): 397–408. doi:10.1016/j.neuroscience.2007.12.022.

Oliver, Douglas L. 2000. “Ascending Efferent Projections of the Superior Olivary Complex.” *Microscopy Research and Technique* 51 (4): 355–363. doi:10.1002/1097-0029(20001115)51:4<355::AID-JEMT5>3.0.CO;2-J.

Oliver, Douglas L., Gretchen E. Beckius, and Amiram Shneiderman. 1995. “Axonal Projections from the Lateral and Medial Superior Olive to the Inferior Colliculus of the Cat: A Study Using Electron Microscopic Autoradiography.” *The Journal of Comparative Neurology* 360 (1): 17–32. doi:10.1002/cne.903600103.

Orton, Llwyd David, and Adrian Rees. 2014. “Intercollicular Commissural Connections Refine the Representation of Sound Frequency and Level in the Auditory Midbrain.” *eLife* 3 (November): e03764. doi:10.7554/eLife.03764.

Palmer, Alan R., Dan Jiang, and David McAlpine. 1999. “Desynchronizing Responses to Correlated Noise: A Mechanism for Binaural Masking Level Differences at the Inferior Colliculus.” *Journal of Neurophysiology* 81 (2): 722–734.

———. 2000. “Neural Responses in the Inferior Colliculus to Binaural Masking Level Differences Created by Inverting the Noise in One Ear.” *Journal of Neurophysiology* 84 (2): 844–52.

Parsons, Carl H., Richard G. Lanyon, Jan W. H. Schnupp, and Andrew J. King. 1999. “Effects of Altering Spectral Cues in Infancy on Horizontal and Vertical Sound Localization by Adult Ferrets.” *Journal of Neurophysiology* 82 (5): 2294–2309.

Patterson, Roy D. 1976. “Auditory Filter Shapes Derived with Noise Stimuli.” *The Journal of the Acoustical Society of America* 59 (3): 640–54. doi:10.1121/1.380914.

Pecka, Michael, Antje Brand, Oliver Behrend, and Benedikt Grothe. 2008. “Interaural Time Difference Processing in the Mammalian Medial Superior Olive: The Role of Glycinergic Inhibition.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (27): 6914–25. doi:10.1523/JNEUROSCI.1660-08.2008.

Perez, Claudia A., Crystal T. Engineer, Vikram Jakkamsetti, Ryan S. Carraway, Matthew S. Perry, and Michael P. Kilgard. 2012. “Different Timescales for the Neural Coding of Consonant and Vowel Sounds.” *Cerebral Cortex*, March. doi:10.1093/cercor/bhs045.

———. 2013. “Different Timescales for the Neural Coding of Consonant and Vowel Sounds.” *Cerebral Cortex* 23 (3): 670–83. doi:10.1093/cercor/bhs045.

Peterson, Gordon E., and Harold L. Barney. 1952. “Control Methods Used in a Study of the Vowels.” *The Journal of the Acoustical Society of America* 24 (2): 175–84. doi:10.1121/1.1906875.

Pfingst, B. E., and T. A. O’Connor. 1981. “Characteristics of Neurons in Auditory Cortex of Monkeys Performing a Simple Auditory Task.” *Journal of Neurophysiology* 45 (1): 16–34.

Pfingst, B. E., T. A. O’Connor, and J. M. Miller. 1977. “Response Plasticity of Neurons in Auditory Cortex of the Rhesus Monkey.” *Experimental Brain Research* 29 (3–4): 393–404. doi:10.1007/BF00236178.

Picou, Erin M., Elizabeth Aspell, and Todd A. Ricketts. 2014. “Potential Benefits and Limitations of Three Types of Directional Processing in Hearing Aids.” *Ear and Hearing* 35 (3): 339–52. doi:10.1097/AUD.0000000000000004.

Plomp, R., and H. J. M. Steeneken. 1969. “Effect of Phase on the Timbre of Complex Tones.” *The Journal of the Acoustical Society of America* 46 (2B): 409–21. doi:10.1121/1.1911705.

Pollack, Irwin, and W. J. Trittipoe. 1959a. “Binaural Listening and Interaural Noise Cross Correlation.” *The Journal of the Acoustical Society of America*, a, 31 (9): 1250–52. doi:10.1121/1.1907852.

Pollack, Irwin, and William Trittipoe. 1959b. “Interaural Noise Correlations: Examination of Variables.” *The Journal of the Acoustical Society of America*, b, 31 (12): 1616–18. doi:10.1121/1.1907669.

Quaranta, Nicola, Francesco Coppola, Mara Casulli, Maria Rosaria Barulli, Francesco Panza, Rosanna Tortelli, Vincenzo Solfrizzi, Carlo Sabbà, and Giancarlo Logroscino. 2015. “Epidemiology of Age Related Hearing Loss: A Review.” *Hearing, Balance and Communication* 13 (2): 77–81. doi:10.3109/21695717.2014.994869.

Rakerd, Brad, and W. M. Hartmann. 1986. “Localization of Sound in Rooms, III: Onset and Duration Effects.” *The Journal of the Acoustical Society of America* 80 (6): 1695–1706. doi:10.1121/1.394282.

Reale, RA, and JF Brugge. 1990. “Auditory Cortical Neurons Are Sensitive to Static and Continuously Changing Interaural Phase Cues.” *J Neurophysiol* 103: 313–38.

Rébillat, Marc, Victor Benichoux, Makoto Otani, Renaud Keriven, and Romain Brette. 2014. “Estimation of the Low-Frequency Components of the Head-Related Transfer Functions of Animals from Photographs.” *The Journal of the Acoustical Society of America* 135 (5).

Recanzone, Gregg H., Darren C. Guard, and Mimi L. Phan. 2000. “Frequency and Intensity Response Properties of Single Neurons in the Auditory Cortex of the Behaving Macaque Monkey.” *Journal of Neurophysiology* 83 (4): 2315–31.

Rose, J. E., N. B. Gross, C. D. Geisler, and J. E. Hind. 1966. “Some Neural Mechanisms in the Inferior Colliculus of the Cat Which May Be Relevant to Localization of a Sound Source.” *Journal of Neurophysiology* 29 (2): 288–314.

Ryan, A., and J. Miller. 1978. “Single Unit Responses in the Inferior Colliculus of the Awake and Performing Rhesus Monkey.” *Experimental Brain Research* 32 (3): 389–407.

Ryan, Allen. 1976. “Hearing Sensitivity of the Mongolian Gerbil, Merionesunguiculatis.” *The Journal of the Acoustical Society of America* 59 (5): 1222–26. doi:10.1121/1.380961.

Ryan, ALLEN F., JOSEF M. Miller, BRYAN E. Pfingst, and G. K. Martin. 1984. “Effects of Reaction Time Performance on Single-Unit Activity in the Central Auditory Pathway of the Rhesus Macaque.” *J Neurosci* 4: 298–308.

Sachs, Murray B., Ian C. Bruce, Roger L. Miller, and Eric D. Young. 2002. “Biological Basis of Hearing-Aid Design.” *Annals of Biomedical Engineering* 30 (2): 157–68. doi:10.1114/1.1458592.

Salminen, Nelli H, Hannu Tiitinen, Santeri Yrttiaho, and Patrick J C May. 2010. “The Neural Code for Interaural Time Difference in Human Auditory Cortex.” *The Journal of the Acoustical Society of America* 127 (2): EL60-5. doi:10.1121/1.3290744.

Schebesch, Gabriele, Andrea Lingner, Uwe Firzlaff, Lutz Wiegrebe, and Benedikt Grothe. 2010. “Perception and Neural Representation of Size-Variant Human Vowels in the Mongolian Gerbil (Meriones Unguiculatus).” *Hearing Research* 261 (1–2): 1–8. doi:10.1016/j.heares.2009.12.016.

Schilling, John R., Roger L. Miller, Murray B. Sachs, and Eric D. Young. 1998. "Frequency-Shaped Amplification Changes the Neural Representation of Speech with Noise-Induced Hearing Loss." *Hearing Research* 117 (1–2): 57–70. doi:10.1016/S0378-5955(98)00003-3.

Schmitzer-Torbert, N., J. Jackson, D. Henze, K. Harris, and A. D. Redish. 2005a. "Quantitative Measures of Cluster Quality for Use in Extracellular Recordings." *Neuroscience* 131 (1): 1–11.

Schmitzer-Torbert, N, J Jackson, D Henze, K Harris, and A D Redish. 2005b. "Quantitative Measures of Cluster Quality for Use in Extracellular Recordings." *Neuroscience* 131 (1): 1–11. doi:10.1016/j.neuroscience.2004.09.066.

Scott, Brian H., Brian J. Malone, and Malcolm N. Semple. 2007. "Effect of Behavioral Context on Representation of a Spatial Cue in Core Auditory Cortex of Awake Macaques." *Journal of Neuroscience* 27 (24): 6489–99. doi:10.1523/JNEUROSCI.0016-07.2007.

Scott, Brian H, Brian J Malone, and Malcolm N Semple. 2009. "Representation of Dynamic Interaural Phase Difference in Auditory Cortex of Awake Rhesus Macaques." *Journal of Neurophysiology* 101 (4): 1781–99. doi:10.1152/jn.00678.2007.

Shackleton, Trevor M., Robert H. Arnott, and Alan R. Palmer. 2005. "Sensitivity to Interaural Correlation of Single Neurons in the Inferior Colliculusof Guinea Pigs." *Journal of the Association for Research in Otolaryngology* 6 (3): 244–59. doi:10.1007/s10162-005-0005-8.

Shackleton, Trevor M., Ray Meddis, and Michael J. Hewitt. 1992. "Across Frequency Integration in a Model of Lateralization." *The Journal of the Acoustical Society of America* 91 (4): 2276–79. doi:10.1121/1.403663.

Shamma, Shihab A., Naiming Shen, and Preetham Gopalaswamy. 1989. "Stereausis: Binaural Processing without Neural Delays." *The Journal of the Acoustical Society of America* 86 (3): 989–1006. doi:10.1121/1.398734.

Shera, Christopher A., John J. Guinan, and Andrew J. Oxenham. 2002. "Revised Estimates of Human Cochlear Tuning from Otoacoustic and Behavioral Measurements." *Proceedings of the National Academy of Sciences* 99 (5): 3318–23. doi:10.1073/pnas.032675099.

Shore, S. E., C. J. Sumner, S. C. Bledsoe, and J. Lu. 2003. "Effects of Contralateral Sound Stimulation on Unit Activity of Ventral Cochlear Nucleus Neurons." *Experimental Brain Research* 153 (4): 427–435.

Simpson, Andrea. 2009. "Frequency-Lowering Devices for Managing High-Frequency Hearing Loss: A Review." *Trends in Amplification* 13 (2): 87–106. doi:10.1177/1084713809336421.

Sinnott, Joan M., Stacy L. Street, Kelly W. Mosteller, and Trina L. Williamson. 1997. "Behavioral Measures of Vowel Sensitivity in Mongolian Gerbils (Meriones Unguiculatus) : Effects of Age and Genetic Origin." *Hearing Research* 112 (1–2): 235–46. doi:10.1016/S0378-5955(97)00125-1.

Siveke, Ida, Christian Leibold, and Benedikt Grothe. 2007. "Spectral Composition of Concurrent Noise Affects Neuronal Sensitivity to Interaural Time Differences of Tones in the Dorsal Nucleus of the Lateral Lemniscus." *Journal of Neurophysiology* 98 (5): 2705–15. doi:10.1152/jn.00275.2007.

Siveke, Ida, Michael Pecka, Armin H Seidl, Sylvie Baudoux, and Benedikt Grothe. 2006. "Binaural Response Properties of Low-Frequency Neurons in the Gerbil Dorsal Nucleus of the Lateral Lemniscus." *Journal of Neurophysiology* 96 (3): 1425–40. doi:10.1152/jn.00713.2005.

Sivian, L. J., and S. D. White. 1933. “ON MINIMUM AUDIBLE SOUND FIELDS.” *The Journal of the Acoustical Society of America* 4 (4): 288–321. doi:10.1121/1.1915608.

Soede, Wim, Frans A. Bilsen, and Augustinus J. Berkhout. 1993. “Assessment of a Directional Microphone Array for Hearing-impaired Listeners.” *The Journal of the Acoustical Society of America* 94 (2): 799–808. doi:10.1121/1.408181.

Sondhi, Man Mohan, and Newman Guttman. 1966. “Width of the Spectrum Effective in the Binaural Release of Masking.” *The Journal of the Acoustical Society of America* 40 (3): 600–606. doi:10.1121/1.1910124.

Spitzer, M W, and M N Semple. 1995. “Neurons Sensitive to Interaural Phase Disparity in Gerbil Superior Olive: Diverse Monaural and Temporal Response Properties.” *Journal of Neurophysiology* 73 (4): 1668–90.

Stecker, G. Christopher, Ian A. Harrington, and John C. Middlebrooks. 2005. “Location Coding by Opponent Neural Populations in the Auditory Cortex.” *PLOS Biology* 3 (3): e78. doi:10.1371/journal.pbio.0030078.

Stern, Richard M., Andrew S. Zeiberg, and Constantine Trahiotis. 1988. “Lateralization of Complex Binaural Stimuli: A Weighted-image Model.” *The Journal of the Acoustical Society of America* 84 (1): 156–65. doi:10.1121/1.396982.

Syka, Josef, Jiří Popelář, Eugen Kvašňák, and Jaromír Astl. 2000. “Response Properties of Neurons in the Central Nucleus and External and Dorsal Cortices of the Inferior Colliculus in Guinea Pig.” *Experimental Brain Research* 133 (2): 254–66. doi:10.1007/s002210000426.

Tanaka, H, K Hachisuka, and H Ogata. 1999. “Sound Lateralisation in Patients with Left or Right Cerebral Hemispheric Lesions: Relation with Unilateral Visuospatial Neglect.” *Journal of Neurology, Neurosurgery, and Psychiatry* 67 (4): 481–86.

Ter-Mikaelian, Maria, Dan H. Sanes, and Malcolm N. Semple. 2007a. “Transformation of Temporal Properties between Auditory Midbrain and Cortex in the Awake Mongolian Gerbil.” *The Journal of Neuroscience* 27 (23): 6091–6102. doi:10.1523/JNEUROSCI.4848-06.2007.

Ter-Mikaelian, Maria, Dan H Sanes, and Malcolm N Semple. 2007b. “Transformation of Temporal Properties between Auditory Midbrain and Cortex in the Awake Mongolian Gerbil.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 27 (23): 6091–6102. doi:10.1523/JNEUROSCI.4848-06.2007.

Thomas, Hardy, Jochen Tillein, Peter Heil, and Henning Scheich. 1993. “Functional Organization of Auditory Cortex in the Mongolian Gerbil (Meriones Unguiculatus). I. Electrophysiological Mapping of Frequency Representation and Distinction of Fields.” *The European Journal of Neuroscience* 5 (7): 882–97.

Tollin, Daniel J., and Tom C. T. Yin. 2005. “Interaural Phase and Level Difference Sensitivity in Low-Frequency Neurons in the Lateral Superior Olive.” *The Journal of Neuroscience* 25 (46): 10648–57. doi:10.1523/JNEUROSCI.1609-05.2005.

Van den Bogaert, Tim, Simon Doclo, Jan Wouters, and Marc Moonen. 2008. “The Effect of Multimicrophone Noise Reduction Systems on Sound Source Localization by Users of Binaural Hearing Aids.” *The Journal of the Acoustical Society of America* 124 (1): 484–497.

Van den Bogaert, Tim, Thomas J. Klasen, Marc Moonen, Lieselot Van Deun, and Jan Wouters. 2006. “Horizontal Localization with Bilateral Hearing Aids: Without Is Better than with.” *The Journal of the Acoustical Society of America* 119 (1): 515–526.

Vanderweele, Dennis A., and Robert M. Abelson. 1973. “Selected Schedules of Reinforcement in the Mongolian Gerbil.” *Psychological Reports* 33 (1): 99–104. doi:10.2466/pr0.1973.33.1.99.

Vickers, Deborah A., Brian CJ Moore, and Thomas Baer. 2001. “Effects of Low-Pass Filtering on the Intelligibility of Speech in Quiet for People with and without Dead Regions at High Frequencies.” *The Journal of the Acoustical Society of America* 110 (2): 1164–1175.

Victor, J D, and K P Purpura. 1996. “Nature and Precision of Temporal Coding in Visual Cortex: A Metric-Space Analysis.” *Journal of Neurophysiology* 76 (2): 1310–26.

Victor, Jonathan D., and Keith P. Purpura. 1997. “Metric-Space Analysis of Spike Trains: Theory, Algorithms and Application.” *Network: Computation in Neural Systems* 8 (2): 127–164.

Vonderschen, Katrin, and Hermann Wagner. 2009. “Tuning to Interaural Time Difference and Frequency Differs between the Auditory Arcopallium and the External Nucleus of the Inferior Colliculus.” *Journal of Neurophysiology* 101 (5): 2348–61. doi:10.1152/jn.91196.2008.

———. 2012. “Transformation from a Pure Time Delay to a Mixed Time and Phase Delay Representation in the Auditory Forebrain Pathway.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 32 (17): 5911–23. doi:10.1523/JNEUROSCI.5429-11.2012.

———. 2014. “Detecting Interaural Time Differences and Remodeling Their Representation.” *Trends in Neurosciences* 37 (5). Elsevier Ltd: 289–300. doi:10.1016/j.tins.2014.03.002.

Vuorialho, Arja, Petri Karinen, and Martti Sorri. 2006. “Counselling of Hearing Aid Users Is Highly Cost-Effective.” *European Archives of Oto-Rhino-Laryngology and Head & Neck* 263 (11): 988–95. doi:10.1007/s00405-006-0104-0.

Wakeford, Onslow S., and Donald E. Robinson. 1974. “Detection of Binaurally Masked Tones by the Cat.” *The Journal of the Acoustical Society of America* 56 (3): 952–56. doi:10.1121/1.1903354.

Walton, Joseph P. 2010. “Timing Is Everything: Temporal Processing Deficits in the Aged Auditory Brainstem.” *Hearing Research*, Special Issue: Aging and Hearing, 264 (1–2): 63–69. doi:10.1016/j.heares.2010.03.002.

Warren, Edus H, and M. Charles Liberman. 1989. “Effects of Contralateral Sound on Auditory-Nerve Responses. I. Contributions of Cochlear Efferents.” *Hearing Research* 37 (2): 89–104. doi:10.1016/0378-5955(89)90032-4.

Whitton, Jonathon P., Kenneth E. Hancock, and Daniel B. Polley. 2014. “Immersive Audiomotor Game Play Enhances Neural and Perceptual Salience of Weak Signals in Noise.” *Proceedings of the National Academy of Sciences* 111 (25): E2606–E2615.

Wightman, F L, and D J Kistler. 1992. “The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization.” *The Journal of the Acoustical Society of America* 91 (3): 1648–61.

Wilbanks, W. A., and John K. Whitmore. 1968. “Detection of Monaural Signals as a Function of Interaural Noise Correlation and Signal Frequency.” *The Journal of the Acoustical Society of America* 43 (4): 785–97. doi:10.1121/1.1910897.

Winer, Jeffery A., and Christoph E. Schreiner. 2005a. *The Inferior Colliculus*. Springer. http://books.google.com/books?hl=en&lr=&id=6Ce3PMzcRqsC&oi=fnd&pg=PR7&dq=winer+schreiner+Inferior+colliculus+book&ots=OXBOZlwRrE&sig=F5NfdGumWM4h0b1XQhVsQRzShQM.

Winer, Jeffery A, and Christoph E Schreiner, eds. 2005b. *The Inferior Colliculus*. New York: Springer.

Yamada, K, K Kaga, A Uno, and M Shindo. 1996. “Sound Lateralization in Patients with Lesions Including the Auditory Cortex: Comparison of Interaural Time Difference (ITD) Discrimination and Interaural Intensity Difference (IID) Discrimination.” *Hearing Research* 101 (1–2): 173–80.

Yin, T C, and J C Chan. 1990a. “Interaural Time Sensitivity in Medial Superior Olive of Cat.” *Journal of Neurophysiology* 64 (2): 465–88.

Yin, T. C., and J. C. Chan. 1990b. “Interaural Time Sensitivity in Medial Superior Olive of Cat.” *Journal of Neurophysiology* 64 (2): 465–88.

Yin, T. C., J. C. Chan, and L. H. Carney. 1987. “Effects of Interaural Time Delays of Noise Stimuli on Low-Frequency Cells in the Cat’s Inferior Colliculus. III. Evidence for Cross-Correlation.” *Journal of Neurophysiology* 58 (3): 562–83.

Yin, T. C., J. C. Chan, and D. R. Irvine. 1986. “Effects of Interaural Time Delays of Noise Stimuli on Low-Frequency Cells in the Cat’s Inferior Colliculus. I. Responses to Wideband Noise.” *Journal of Neurophysiology* 55 (2): 280–300.

Yin, T. C., and S. Kuwada. 1983. “Binaural Interaction in Low-Frequency Neurons in Inferior Colliculus of the Cat. III. Effects of Changing Frequency.” *Journal of Neurophysiology* 50 (4): 1020–42.

Zurek, Patrick M., Richard L. Freyman, and Uma Balakrishnan. 2004. “Auditory Target Detection in Reverberation.” *The Journal of the Acoustical Society of America* 115 (4): 1609–20. doi:10.1121/1.1650333.

Zurek, Paul M. 1993. “Binaural Advantages and Directional Effects in Speech Intelligibility.” *Acoustical Factors Affecting Hearing Aid Performance* 2: 255–275.

Zwiers, Marcel P, Huib Versnel, and A John Van Opstal. 2004. “Involvement of Monkey Inferior Colliculus in Spatial Hearing.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 24 (17): 4145–56. doi:10.1523/JNEUROSCI.0199-04.2004.

Zwislocki, J., and R. S. Feldman. 1956. “Just Noticeable Differences in Dichotic Phase.” *The Journal of the Acoustical Society of America* 28: 860.

V. Awake electrophysiological and behavioral recordings

1. Introduction

Sensory information processing was and is still often studied through neural recordings in anesthetized animals. While this allows the collection of neuronal responses in great variety of detail and yield (single cell patch clamp to multi-electrode recording or wide field imaging), we don't know how similar the brain activity is when the animal is anesthetized or actively processing stimuli.

There have been few studies of neural activity in the inferior colliculus of awake animals. In the gerbil, Ter-Mikaelian, Sanes, and Semple (2007) found that neural responses were less variable in the inferior colliculus than in the auditory cortex when switching from anesthetized to awake animals, but they did not study any binaural properties. Other studies have measured ITD and interaural coherence sensitivity in the awake passive rabbit (Coffey et al. 2006; Day and Delgutte 2013) and the basic properties of neurons seem to be similar to those measured in anesthetized recordings. We recorded neuronal activity in the IC in awake passive gerbils and could show a proof of concept that electrophysiological recordings are possible.

In these studies, the animals were not involved in a task so the effects of attention could not be measured. We trained gerbils to perform an auditory task in a free field arena, which could later be used for a wide variety of auditory paradigms and coupled with electrophysiological recordings. We showed that gerbils could indeed perform in a very simple frequency discrimination task and laid the bases for gerbil training.

2. Methods

a. Surgery for awake electrophysiological recordings

Adult male gerbils (60–80 g, P60–P120) were habituated to handling and laboratory environment for one week before the surgery. On the surgery day, they were anesthetized with an initial injection of a mix of fentanyl, medeteomidine and midazolam, and the same solution was injected every 1.5h during the surgery. Metacam was injected prior to the surgery and once a day for 5 days afterwards. The animals were secured on a stereotactic frame with hollow ear bars and a bite bar, and 5 screws were mounted on the skull to secure the implant. A craniotomy was performed over the inferior colliculus, an incision was made in the dura mater, and a tetrode bundle was lowered into the brain. Auditory stimulus was played while the tetrodes were lowered so that they could be implanted in the region where auditory driven responses started to appear. The craniotomy was protected with sterile jelly, the implant was sealed with dental cement and the skin was sutured around the implant. Anesthesia was reversed with a mix of Atipamezol, Anexate and Naloxone, and the animals were monitored during the recovery period.

Chronic implants were made of a bundle of 4 tetrodes, each of which was made of 4 wires of insulated tungsten wire twisted together. The electrodes were protected by a cannula and were connected to a custom made printed circuit board and to a lightweight 1-screw microdrive (Axona) allowing the movement of the whole electrode bundle along the depth axis after the implantation. Only the tips of the electrodes were de-insulated by a sharp scissor cut. They were not gold plated. The single electrode impedances varied from 0.8 to 2MOhm.

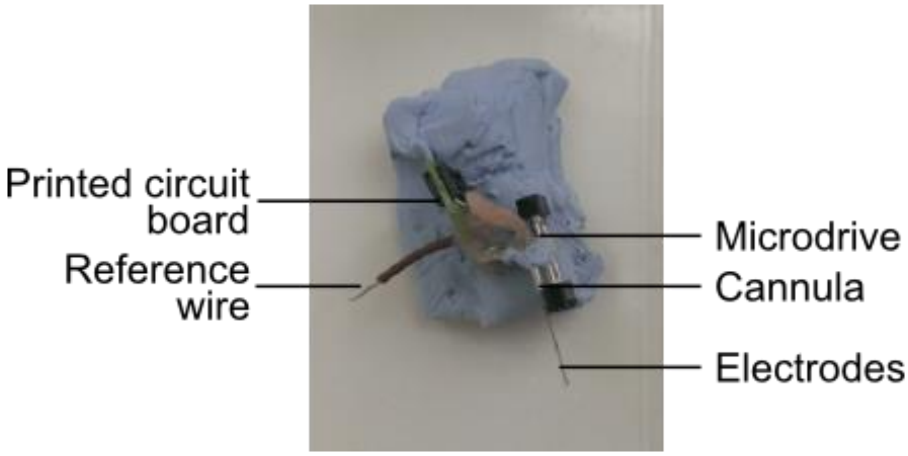

Figure 36: Microdrive before implantation.

b. Awake passive electrophysiological recordings

Gerbils were habituated to stay on an elevated platform of 30cm diameter in a sound proof booth before the surgery. After the chronic implant surgery and 5 days of recovery, they were placed on the platform, the implant was connected to the computer via a Zif-Clip (TDT), a commutator (Dragonfly) and an amplifier (TDT) and electrophysiological responses were measured. The same stimuli as in the anesthetized experiment were played, except that two diametrically opposite free field speakers (TDT) were used for spatial cues instead of ITDs applied through headphones. The position of the head of the animal was tracked by a high speed camera (Epix) through two LEDs present on the Zif-Clip, allowing for post hoc reconstruction of the speaker positions relative to the head of the animal and hence of the spatial cues it experienced during auditory stimulation.

c. Behavioral procedure

Gerbils were put individually in an arena of 90cm diameter delimited by high plastic walls. Their position was tracked in real time with a camera (FireFly) placed over the center of the arena. Pure tones were presented through one free field speaker (Audio3) placed on the outer edge of the arena. The frequency of the pure tone was changed according to the animal's position in the arena, creating an auditory virtual environment that the animals could explore. The pure tones were 50ms long with a 5ms on and off cosine ramp. They were presented as a continuous stream of single tones with pauses of 50ms between each tone. The mean tone intensity was 70dB SPL with a random +/-5dB roving level.

The animals were trained to find a small circular area of 15cm diameter where a 660Hz tone was played that we will call target island. The animals were trained to find this target island within three different environments:

- Island: environment where a 20kHz tone was played everywhere except in the target island (Figure 39A);
- Gradient: environment where a circular gradient of increasing frequency was around the target island, with steps of 7.5cm. The gradient frequencies were 660, 4528, 8396, 12264, 16132 and 20000Hz (Figure 39B);
- Multi-island: environment where 20kHz was played everywhere except in the target island and in 3 additional non-overlapping islands of the same diameter. The frequencies played in the additional islands were chosen without replacement among 2 sets of frequencies defining two environments (Figure 39C):
    - Multi-island gradient: frequencies we used in the Gradient environment: 4528, 8396, 12264 and 16132Hz,
    - Multi-island near target: frequencies close to that presented in the target island: 460, 860 and 1060Hz.

When the animal stayed for three consecutive seconds in the target island, the sound presentation stopped and the animal was rewarded by a sunflower seed dropped randomly in the arena. The next trial was started as soon as the animal had finished eating the reward. If the animal did not resolve the task within 2min the current trial was stopped and a new trial with a different target island location was started immediately.

At the beginning of the training, the diameter of the target island was increased and the required time to stay inside the target island to get the reward was decreased so that the animals could learn the task with easier parameters. The diameter was then progressively decreased and the time increased to reach the final parameters.

d. Behavioral analysis

The position of each gerbil in the arena was recorded every 100ms as the position of its center of mass. To assess whether the animals could resolve the task by chance, we looked at whether the animals would have resolved the task if the target island had been at another random location. For each trial, we chose a random location for the fake target island and looked at whether the animal would have stayed for enough time in this island to complete the task, using the real trajectory that was recorded for this trial. For the real trials, the animals had 2min to complete the task but often completed it in less time. This is a problem for the fake target analysis because we do not have many full 2min trajectories to analyse. For example, if the animal resolved the real task in 1min, we can only evaluate whether it would have resolved the fake task during this 1min but not if it would have resolved it if it had the full 2min. Hence, we chose to perform this analysis using different time bins. For each time bin, we analysed the animal's success in the real and fake task, hence analysing the behavior in both cases for the same amount of time.

We recorded the gerbil's position every 100ms and could hence analyse the direction and speed the animals travelled at for each of these 'steps'. To understand their exploration strategies, we computed the angle and speed biases towards the target island. The angle bias was computed as the proportion of steps the animal took towards the target island. The speed bias was computed as the speed of the animal (i.e. the distance between two consecutive position recordings) when going towards the target island compared to its speed when going in other directions.

3. Results

a. Neuronal population in the awake passive IC

We successfully implanted a bundle of 4 tungsten wire tetrodes in the Inferior Colliculus of 4 adult male gerbils. We were able to perform spike sorting on the electrophysiological data and to record from one to three single units every day. We observed a wide diversity of cells, with slightly different characteristics than what we had observed in our anesthetized recordings.

We could classify our cells in three broad categories based on their frequency receptive field:

- Cells with a sharp frequency receptive field and strong inhibitory side bands (Figure 37 A,B),
- Cells with a wide frequency receptive field (Figure 37C),
- Cells that fired rhythmically at a high rate but were not driven by our auditory stimuli (Figure 37D).

The cells had diverse responses to pure tones with onset, sustained, build-up or mixed post-stimulus time histograms (PSTH). They also responded in different way to our vowels with their firing rate increasing, decreasing or staying the same during the vowel presentation, and with or without an onset response.

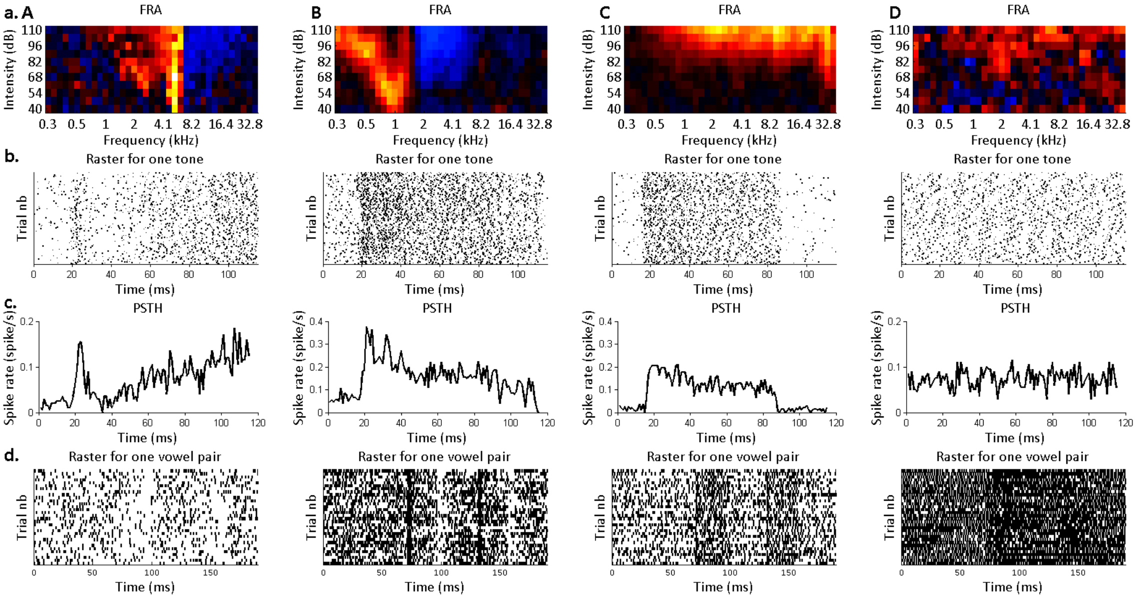

Figure 37: Examples of 4 cells (columns) recorded in awake passive animals. **a.** FRAs measured with 75ms long pure tones with frequencies ranging from 300Hz to 8.2kHz and intensities from 40dB to 110dB. **b.** Raster for one 75ms long pure tone at the cell's best frequency. **c.** PSTHs measured by playing 75ms long pure tones at the cell's best frequency. **d.** Raster for one full trial of our vowel stimulus (masker alone – masker + vowel – masker alone – masker + vowel – masker alone) with the masker played 5dB more intense than the vowel . The masker and vowel were played from diametrically opposite speakers across the platform the animal was on.

We observed that the basic properties of the cells changed with the state of the animal. The population of cells we recorded from awake gerbils had more cells with a best frequency around 5kHz, and had higher firing rates in response to the vowel stimulus (Figure 38). Indeed, the neurons recorded in the awake animals had an average firing rate of 28.5spikes/s (Figure 38Ab) during a full trial (presentation of two vowels with the masker) while the neurons recorded in the anesthetized animals had an average firing rate of 9spikes/s (Figure 38Bb). We observed the same phenomenon when comparing the firing rates to vowels alone in the awake animals (26.2 spikes/s on average, Figure 38Ac), to the firing rates to pure tones of an equivalent intensity in the anesthetized animal (1.0 spikes/s on average, Figure 38Bc).

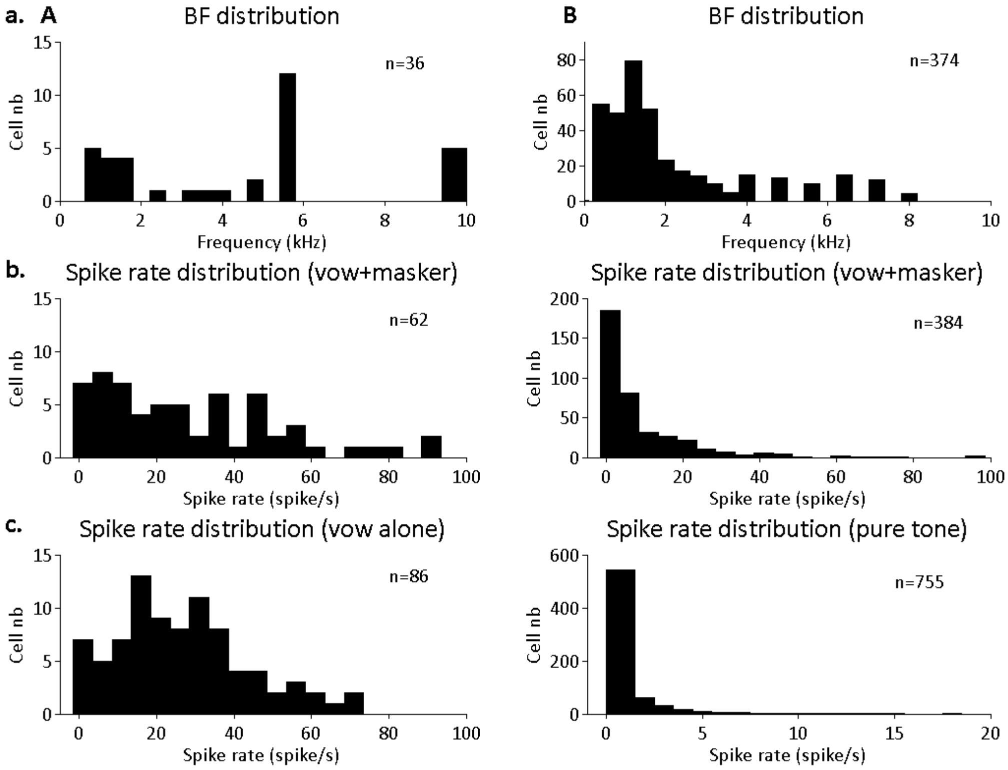

Figure 38: Characteristics of neuronal populations recorded in **A.** awake and **B.** anesthetized gerbils. **a.** Best frequency distributions, defined as the pure tone frequency that elicited the maximum spike rate over all sound intensities. **b.** Spike rate distribution measured during a full trial in the Opposite condition (masker alone – vowel – masker alone – vowel – masker alone). **c.** Spike rate distribution measured over the presentation of one vowel alone from one free field speaker (A) or one pure tone alone of equivalent intensity from two earphones (B).

In conclusion, we observed that cells had a higher firing rate and more diverse responses in the awake animal compared to the anesthetized animal. Even though we could record from only very few cells which makes our results preliminary, this indicates that the processing of sensory information might be different in these two brain states and stresses the importance of studying neural activity in awake animals.

b. Development of a behavioral task

We developed a new behavioral paradigm using a large circular arena that gerbils could freely explore. We created an acoustic virtual environment inside this arena by presenting various sounds from a single speaker placed on the side of the arena. We tracked the animal's position in real time using a camera placed over the arena and modified the sounds that were presented in function of the animal's position. The animal was rewarded if it could find a circular area inside the arena where a specific sound was played, and stay in that area for 3s. The location of this target island was chosen randomly at the beginning of each trial so that the animals had to find it using only auditory cues and could not rely on spatial cues.

We started by training gerbils with a similar task as the one we used for the human psychophysics and physiological experiment. We played a stream of two alternating vowels and the animals had to find the target island where the two vowels were the same. To increase the repetition rate of the stimulus, we used 50ms long vowels with a 50ms pause between them. After 2 months of training, we realized that the gerbils could not learn this task. Moreover, we found that we would not be able to study ITD cues in our arena due to its geometry. We hence decided to focus on creating a simple task that the gerbils could learn and that would allow us to have a proof of concept that our free field arena can be used for behavioral training.

Instead of using vowels which were made of 4 harmonics, we used pure tones of 50ms duration. We presented them as a continuous stream with 50ms pauses between each tone. The tones' frequencies were chosen in function of the position of the animal in the arena following three environments (Figure 39) and the animals were trained to find the target island in which a 660Hz tone was played. Our three environments were:

- Island: a tone of 20kHz is played everywhere in the arena except in the target island (Figure 39A);
- Gradient: tones of increasing frequencies are played as the animal gets closer to the target island, creating a discrete frequency gradient around the island (Figure 39B);
- Multi-island: a tone of 20kHz is played everywhere in the arena except in the target island and in three additional island in which other frequencies are played (Figure 39C).

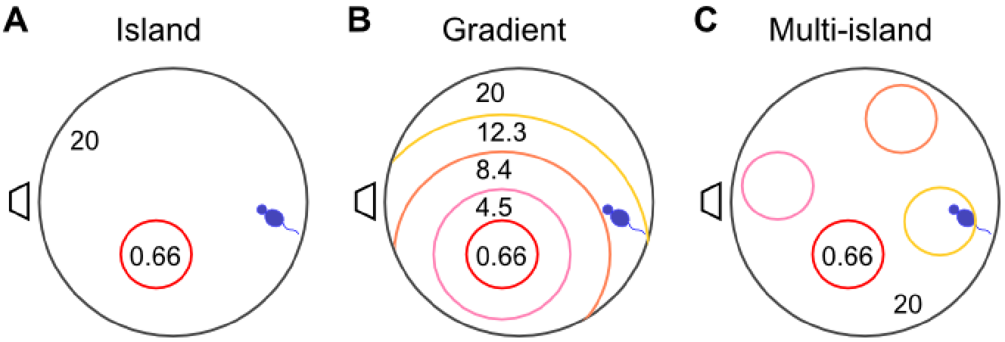

Figure 39: Types of virtual acoustic environments in our free field arena with one speaker. The frequencies of the 50ms long pure tones played in each region are indicated in kHz. **A.** Island environment: 0.66kHz tones were played in the target island and 20kHz tones everywhere else. **B.** Gradient environment: 0.66kHz tones were played in the target island and 4.528, 8.396, 12.264, 16.132Hz and 20kHz tones were played in circular rings of increasing distance to the target island, creating a discrete linear frequency gradient centered on the target island. **C.** Multi-island environment: 0.66kHz tones were played in the target island, the frequencies of the tones played in the 3 non-target islands were chosen without replacement within 4.528, 8.396, 12.264 and 16.132kHz (same frequencies as in the gradient environment) or 0.46, 0.86 and 1.06kHz (frequencies close to 0.66kHz in 200Hz steps) and 20kHz tones were played in the rest of the arena.

With these simpler paradigms, we successfully trained gerbils to find the target island and stay inside it for 3s to obtain a reward. However, we observed that small variations in the training paradigm could influence their behaviour. For example, we first used peanut butter as a reward and fed it to the gerbils through a reward spout located on one side of the arena. The animals then spent a lot of time around the reward spout and caused an important bias in their exploration strategy. We were able to resolve this by dropping sunflower seeds inside the arena from above so that the rewards would arrive at random locations. We also observed that the gerbils were not always engaged in the task during the training session. By monitoring them by eye during the training, we could see that they sometimes resolved the task very quickly and used a very efficient trajectory but sometimes they rested or groomed on the side of the arena for several minutes. We tried to make the gerbils more motivated by starving them to 90% of their initial body weight, but we found that it made them more aggressive and agitated but not more efficient in resolving the task.

In conclusion, we were able to train gerbils to perform a simple auditory discrimination task in a free field arena they could freely explore. We improved the training paradigm through trial and error to find the most efficient way to train gerbils, but we would like to refine it further so that the gerbils are more motivated and have a stable performance during each training session.

c. Insights on explorative strategies of gerbils

We successfully trained 2 gerbils to solve our task in the Gradient environment, and 2 gerbils in the Island environment. We observed that on the first training day all gerbils performed at chance level, while on the last training day they performed significantly better than chance (Figure 40A). This showed that the animals did learn to perform the task and stayed for 3s in the target island to obtain their reward.

We wondered which exploratory strategies the gerbils used to solve the task. One hypothesis was that the gerbils could use the information from the frequency gradient to find the target island. For example, if they heard a succession of tones that decreased in frequency they could know that they were headed towards the target island where the lowest frequency was played. On the contrary, gerbils in the Island environment had no information available about the target island location until they went inside it by chance.

We observed that the gerbils took the same amount of time to resolve the task in both environments, with an average of 20.4+/-13.8s for the Island environment and of 19.9+/-14.0s for the gradient environment (Figure 40B). This indicated that the gerbils might not use the information contained in the gradient as we expected then to resolve the task faster if they used that information. To assess the exploration strategies more precisely, we computed the angle bias towards the target island: we measured the percentage of steps the animal took towards the target or towards other directions relative to the target. A step was defined as the difference in the position of the gerbil we observed between two frames recorded by our camera. We observed that the animals took the same proportion of steps in all directions in the gradient and in the island environment (Figure 40C, left column), which contributes to showing that the gerbils did not use the frequency gradient.

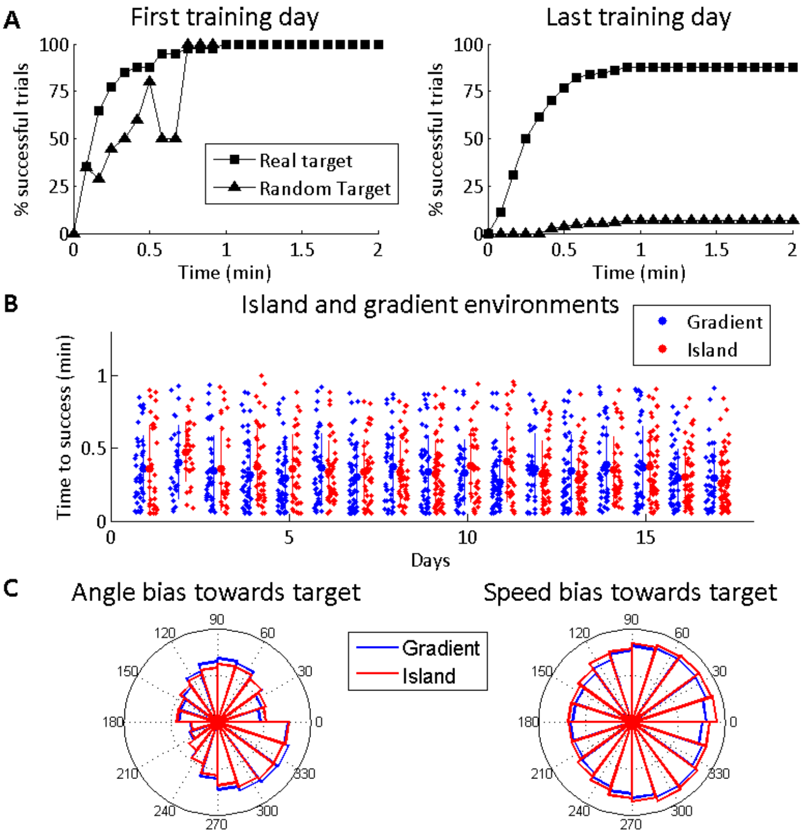

Figure 40: Behavioral results for the Island and Gradient environments. **A.** Percentage of trials successfully completed by all the animals during the first and last day of training for the real task with the real target island and for a fake task with random target islands. We consider the same trajectories for the first seconds (time, x-axis) of each trial to compare the performance in the fake and real tasks over the same amount of time (see methods). **B.** Time taken to resolve the task by animals in the Gradient and in the Island environments over 17 training days. **C.** *(left column)* Proportion of steps taken towards the target island (0 in the polar plot) or in other directions relative to the target island on the last day of training. Angle biases towards the target island: Gradient: r=0.32 max=5.81; Island: r=0.31, max=6.13. *(right column)* Speed of every step taken towards the target island (0 in the polar plot) or in other directions relative to the target island on the last day of training. Speed biases: Gradient: r=0.00 max=5.18; Island: r=0.00, max=0.16.

A recent study (Whitton, Hancock, and Polley 2014) suggested that rodents might not show an angle bias towards the target location but rather a speed bias by running faster when they are oriented towards the target. We hence quantified the speed bias in our experiment by looking at the distance the animals covered when they stepped towards the target island or in other directions relative to the target island. We observed that animals in the Gradient and Island environment showed the same uniform distributions of walking speed relative to the target direction (Figure 40C, right column). Hence, we concluded that the gerbils did not use the frequency gradient information for this task.

The gerbils we used for these analyses were all initially trained with the Gradient environment, and 2 of them were subsequently switched to the Island environment. When we started training the next 4 gerbils, we trained 2 of them directly using the Island environment and it seemed that they learned the task faster than the previous gerbils. This practical observation indicated that the greater variety of tone frequencies presented in the Gradient environment might have confused the gerbils and prevented learning rather than helped them to find the target island. Overall, it seemed that using a very simple auditory stimulus is essential when training gerbils in this free field arena.

d. Electrophysiological recordings stability and yield

We also observed differences in neural activity between awake and anesthetized gerbils on the single neuron scale. In the awake animals, we observed that single neuron activity sometimes changed drastically over short periods of time. For example, some cells had a higher firing rate during 2min and then went back to their initial activity levels (Figure 41).

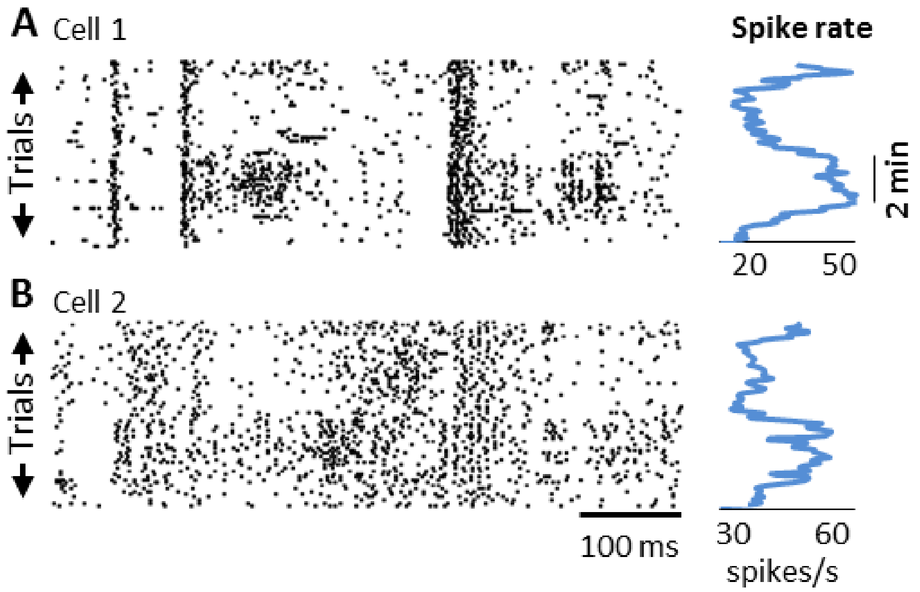

Figure 41: Example of recordings from two single cells (A and B) recorded simultaneously from the Inferior Colliculus of an awake gerbil during the presentation of two vowels in silence. Rasters (left column) and average spike rate over time (right column).

We could not observe any correlations between these activity changes and the animals' behavior as recorded by a camera placed over the platform. It is possible that we could not detect more subtle behavioral states of the animal with these simple observations. It is also possible that the activity changes were due to noise or instability in the electrophysiological recordings.

We observed a large difference in the yield of single units recorded per tetrode in the awake and anesthetized recordings. We recorded a mean of 1.3 cells per tetrode in the awake experiments (Figure 42A) and of 1.9 cells in the anesthetized experiments (Figure 42B). More strikingly, we never recorded more than 4 cells per tetrode in the awake experiments while we recorded as many as 12 cells per tetrode in the anesthetized ones. We also had days where we could not isolate any single units during the awake recordings that do not appear on this plot.

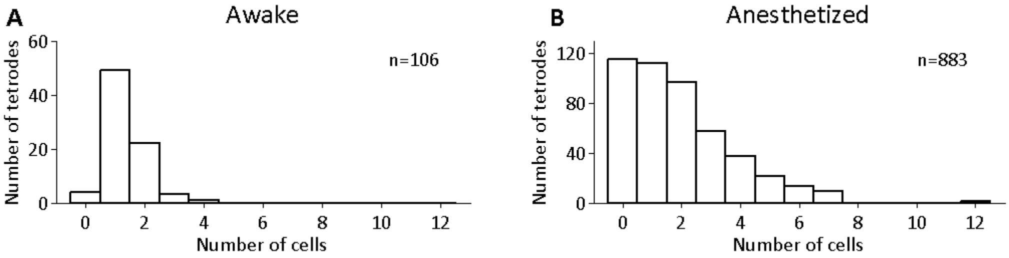

Figure 42: Number of cells recorded for each implanted tetrode recorded in the **A.** awake passive gerbil with tungsten tetrodes and **B.** anesthetized gerbil with silicon probes.

These differences were probably due in part to the different electrodes we used. We used Neuronexus silicon probes for the anesthetized experiments and self-made tungsten wire tetrodes for the awake recordings. The Neuronexus probes were more rigid and the geometry of the electrodes was optimized for single unit isolation whereas the tungsten wire bundles were sometimes slightly bent during the implantation and their geometry was not exactly reproducible from one surgery to another. The impedance of the Neuronexus electrodes was also better controlled and more equal across sites whereas we chose not to gold plate our tungsten electrodes which sometimes led to big differences in impedances. The fact that the awake recordings were made on an awake animal also probably made the recordings less stable over time, and we might have recorded from some neurons for a too short period of time to be able to isolate it as a single unit.

We could show that it was possible to record from isolated single units in the awake gerbil. We observed differences in the recordings that seemed to be in part due to a different brain state in awake animals and in part to the greater technical difficulty of these experiments.

e. Gerbils can recognize absolute pure tone frequencies

We observed that gerbils could learn to find a target island where a 660Hz tone was played and stay within that island for 3s to obtain a reward. In all the paradigms we used previously, we used 660Hz as the target frequency and that was also the lowest frequency that was presented in the whole environment. Hence, the gerbils could learn to find the area where the lowest frequency was presented or learn to recognize an absolute frequency of 660Hz. In the Island environment, they could also simply learn to recognize a change in frequency from a high frequency tone (20kHz) to a low frequency tone (660Hz) without learning to recognize the absolute frequencies.

To distinguish between these possibilities, we trained the 4 gerbils we used previously in a Multi-island environment where a 20KHz tone was played everywhere except in the target island and in 3 additional islands of same diameter where tones different frequencies were played (Figure 39C). First, we trained the gerbils in a Multi-island environment where the frequencies played in the additional islands were drawn randomly without replacement from the frequencies used for the gradient in the Gradient environment. We observed that all the animals could do the task immediately in the new Multi-island environment (Figure 43A). This showed that the learning was transferred between the two environments, and that the gerbils might be able to recognize the absolute frequency of the 660Hz tone. Indeed, stepping inside any of the islands would produce a drop of frequency from 20kHz to a lower frequency, yet the gerbils were able to find the correct target island.

To test whether the gerbils confused the other islands for the target island, we computed the number of times they stayed for 3s or more in islands of different frequencies. This corresponds to the number of times the animal would have successfully resolved the task if the island it stayed in was the target island. Gerbils trained in the Gradient environment seemed to stay in the 8.4kHz island more often in failed trials (i.e. trials where the animal did not stay 3s in the target island in the imparted 2min, Figure 43B, left column), suggesting they might confuse it for the target island. Gerbils trained in the Island environment stayed more often in the 8.4kHz and 12.3kHz island (Figure 43C, left column). This is surprising because we would have expected the gerbils to stay more often in the 4.5kHz island which is the closest frequency to the target island. This could be an artefact due to the small number of trials, to the geometry of the arena that might reflect different frequencies in different ways or to other factors we did not anticipate.

To test whether the animals had learned to find the island where the lowest frequency was played or learned to recognize the absolute frequency of the target island, we tested the same animals on a Multi-Island paradigm where the frequencies of the additional islands where closer and sometimes lower to that of the target island. We chose to use 460Hz, 860Hz and 1060Hz tones in the additional islands so that the frequencies would still be discriminable by the gerbils and the lowest frequency would still be well within their hearing range.

Again, we observed that all gerbils could successfully resolve the task as soon as we changed the environment (Figure 43A), showing that the learning was transferred. This showed that the gerbils did learn to recognize the absolute frequency played in the target island and not only to find the island with the lowest frequency. To see whether the gerbils showed a preference for the lower frequency islands, we computed the number of times they stayed 3s in the non-target islands. We observed that they had a preference for the 860Hz island (Figure 43B and C, right columns), suggesting they might confuse this island with the target island. However, they did not display a strong preference for the 460Hz island, confirming that they learned to recognize the 660Hz target tone and not the absolute lowest frequency.

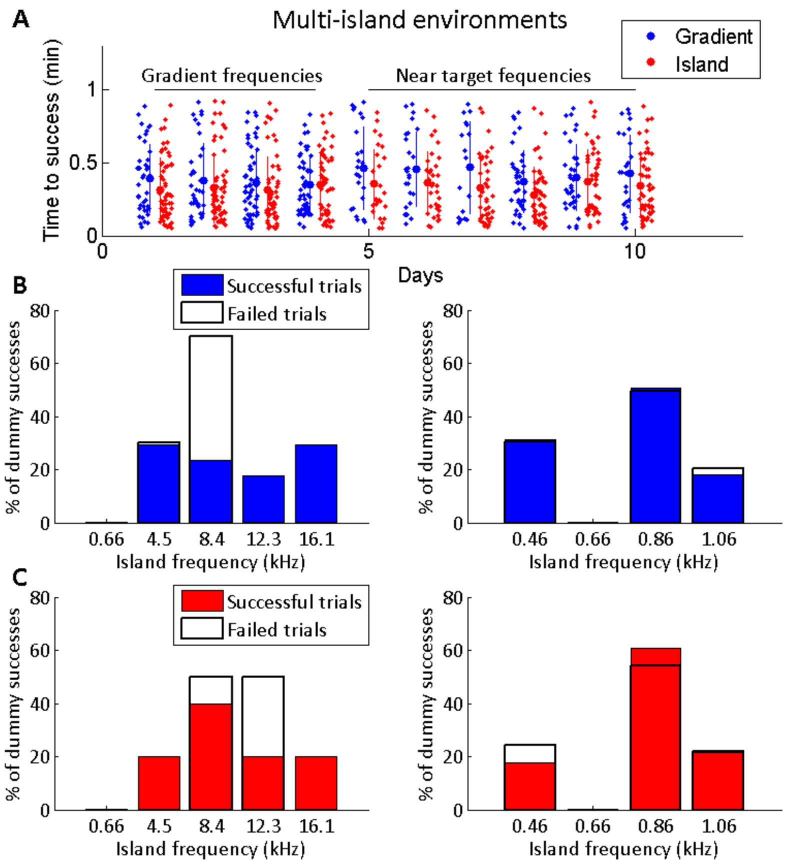

Figure 43: Behavioral results from the Multi-island environments, for gerbils previously trained in the Gradient and Island environments. **A.** Time taken to complete the task in the Multi-island environments (first 4 days with the frequencies from the Gradient environment, last 6 days with the frequencies near the target frequency). **B.** Percentage of times the gerbils spent 3s or more in islands of each frequency in successful trials where they spent 3s in the target island and were rewarded (full bars) and in failed trials where the trial was stopped after 2min (empty bars), for gerbils that had been trained in the Gradient environment. **C.** Same as B for gerbils that had been trained in the Island environment.

To conclude, we observed that the gerbils could transfer their knowledge of the task to a new multi-island environment. Even the gerbils trained in the Island environment that had only been exposed to 2 pure tone frequencies could resolve the task in a more complex environment. We observed that the gerbils did not stay for 3s in the islands where the pure tone was closest or lower in frequency than the target tone, which showed that they learned to discriminate the absolute target tone frequency. These results were encouraging because they showed that after the initial training period where the animals learned to perform an auditory discrimination task, they could quickly be trained to similar task thus allowing us to gain insight more quickly into their exploration strategies.

4. Discussion

a. Electrophysiology

We successfully recorded from single units in the awake gerbil IC, which is an important proof of concept that this chronic tetrode implantation technique can be implemented in gerbils, and that spike sorting is possible in the awake IC. We observed that the cells we recorded in awake passive animals had different properties than the ones we recorded in anesthetized animals. One of the most striking differences was that cells we recorded in awake animals had a much higher firing rate in all conditions. Even though it is hard to compare the sound intensity levels between the speakers we used in the awake experiments and the headphones we used in the anesthetized ones, it is probably safe to assume that the cells would spike more at an equivalent intensity level given the magnitude of the changes we observed. This indicates that the neural activity and maybe the global state of the IC is different in awake animals, and validates the need for such studies of more natural neuronal activity.

We observed more inhibitory effects of sounds in the awake passive animals. Indeed, we observed inhibitory side bands in some FRAs and the vowel presentations seemed to lower firing rates of some cells. These effects could be due to a different state of the brain or simply be revealed by the higher global firing rate. If a cell was firing more spikes in responses to certain sounds or had a higher spontaneous activity, we would be able to detect a drop in firing rate due to inhibition rather than observing a constant null firing rate.

We also observed that cells we recorded in the awake animals had higher best frequencies. This might have been due to a different location of the recording sites, as the implantation procedure was less precise for the chronic recordings. The sound stimulation was also less precise as it was done in free field instead of through headphones placed directly at the beginning of the ear canal.

We noticed differences in single neuron activity over time. While the activity of each cell seemed very stable over one hour or more in the anesthetized experiments, we observed changes in the course of several minutes in the awake animals. These changes in single neuron activity were not obviously correlated to the animals' behavioral states. To analyse this in detail we would need more advanced techniques to record the animals' behavior and head position. Ideally, we would like to track the body of the animals so that we can recognize different behaviours such as quiet rest, exploration or grooming. We would also like to be able to track the head precisely enough to infer the exact ITD and ILD cues the animal experienced during auditory stimulation.

These results are to some extent contradictory with those obtained by Ter-Mikaelian, Sanes, and Semple (2007a). They compared the activity in the IC and in A1 in anesthetized and awake gerbils and concluded that the activity in the IC was similar in both states. However, they measured different variables (the temporal precision of neuronal responses to SAM tones) and used a different anesthetic (Ketamine-Pentobarbital) which might explain the different results. By observing their data, we noticed that the neurons in the awake IC seemed to have more sustained responses to SAM tones and to fire more spikes at all phases of the envelope of the SAM tones. Although the authors did not use this angle in their analysis, it seems that they might also have observed an increase in the overall firing rate in the awake IC.

The single unit yield was lower in our awake recordings, and we think it could be improved by fine tuning the methods for constructing the electrodes and for implanting them. We could try gold plating the electrodes which could improve signal detection by making the geometry of the site of electrical contact more regular. Spike sorting could then become easier because the impedances would be more uniform across electrodes. But it could also become harder because the impedances would be lowered by the gold plating and each electrode might detect too much signal from the neuronal population and we would not be able to isolate single units. We could also try to using commercial chronic silicon probes which should have a similar yield to the ones we used in the anesthetized experiments. The problem is that the surgery procedure to implant them is much longer due to having to partially assemble the microdrive during the surgery and might be detrimental to the health of the animal. They are also very expensive, making the cost of each experiment significantly higher. Overall, technical improvements are possible to have a better yield and we would also probably obtain better results by repeating the same procedure more times and becoming more efficient and experienced.

To conclude, we established an important proof of concept that single unit recordings are possible in the awake gerbil IC. We still need to refine the techniques for recording the animals' behavior, the exact sound location cues they experience in our free field arena and for recording single unit activity but we believe this can be achieved.

b. Behavior

We trained gerbils using a novel technique that allowed animals to freely explore a virtual auditory environment rather than constraining them to a few alternative forced choices. This might correspond to a more ecologic behavior where animals can explore their environment and find which conditions bring them rewards. We observed that training animals with the simplest stimuli was the most efficient, as the animals were trained faster and more reliably using only two pure tones in the environment rather than more pure tones or harmonic complexes. Interestingly, we found that the learning could then be readily transferred to more complex environments where pure tones of different frequencies were presented.

We presented the animals with the task of finding a circular target area where a 660Hz pure tone was played. They performed significantly better than chance after a month of training and we tried to understand which exploration strategies they used to perform the task. We concluded that they learned to recognize the absolute target tone frequency and that they did not use the information contained in a frequency gradient towards the target area when it was available.

We found that the behavioral data remained noisy even after 2 months of training. Informal observation of the animals during the experiment showed that they were sometimes engaged in the task, but that they also often explored the environment or rested independently of the auditory stimulation. We could try tracking the behavior of the animals more precisely to be able to separate between rest and exploratory behavior, but it would still be difficult to exclude part of the data. We could also try to make the animals more motivated by finding a more attractive reward or using an aversive stimulus as punishment for failed trials. We could also have animals initiate trials themselves by nose poking in a ring in the middle of the arena. The trials would then be initiated only when they are ready to perform rather than automatically and they might perform the task right away. This might be a good solution but would increase the training time significantly as the animals would have to learn to nose poke before they start training on the auditory task.

Our initial aim was to conduct electrophysiological recordings during behavior to study the neuronal activity in the IC while animals were engaged in or learning an auditory task. We proved that this is possible as we established a technique for single unit recordings in awake freely moving animals and found an auditory task that gerbils can learn. Unfortunately, we did not have enough time to combine the two techniques, as both still needed further refinements to yield enough reliable data. One major problem is that it is difficult to know when the animals are really attending to the sounds, so analysing the physiological data in regards to attention might be difficult. Nevertheless, our study is a good proof of concept that interesting behavioral and electrophysiological data can be collected in the awake gerbil.

VI. Discussion

1. Comparison of our work with BMLD

a. Psychophysics and BMLD

In keeping with previous research on speech intelligibility in noise (Edmonds and Culling 2005b; Beutelmann, Brand, and Kollmeier 2009), we found that simple vowels composed of 4 harmonics forming two formants were discriminated better when a babble noise masker was presented with an opposite ITD from that of the formants. Our experimental paradigm is similar to the one used to study binaural masking level differences (BMLD). In BMLD paradigms, the intelligibility of speech or the detectability of pure tones is measured in function of the phase relationship between the target sound and a noise masker. The phase shift between target and masker is usually 0 or π (Hirsh 1948a; Licklider 1948).

For our paradigm, we can calculate the phase difference between each of our formants and the noise at that formant's frequency. As we saw in Chapter II (II.1.d), this phase difference can be expressed as:

$$\Delta \mathrm{P} = 4\pi \mathrm{F} * ITD_M$$

In the Opposite and Split conditions where F is the center frequency of the formant and ITD_M is the ITD of the formant. And

$\Delta \mathrm{P} = 0$ in the Same condition.

We can hence compute the phase difference between the noise and each formant for the mean value of the formant frequency for each vowel:

- Reference vowel:
    - Formant 1: $F1_R = 630\mathrm{Hz}$, $\Delta \mathrm{P} = 4.75 rad$, and the closest rational fraction of pi is $\Delta \mathrm{P} \approx 3\pi/2$
    - Formant 2: $F2_R = 1230\mathrm{Hz}$, $\Delta \mathrm{P} = 3.00 rad \approx \pi$

  Formant 2 is hence close to the optimal phase shift of π, and Formant 1 is less phase shifted but still has a notable phase difference with the noise, so one might expect a noticeable BMLD.

- Different 1:
    - $\mathrm{F1} = 510\mathrm{Hz}$, $\Delta \mathrm{P} = 3.85 rad \approx 5\pi/4$

- $\mathrm{F2} = \mathrm{F2r},\ \Delta\mathrm{P} \approx \pi$

  Both formants have an optimal phase shift close to π.

- Different 2:
  - $\mathrm{F1} = F1_R,\ \Delta\mathrm{P} \approx 3\pi/2$
  - $\mathrm{F2} = 1110\mathrm{Hz},\ \Delta\mathrm{P} = 2.09 rad\ \approx 2\pi/3$

  Both formants have a phase shift close to π/2, so there might be less BMLD effect.

- Different 3:
  - $\mathrm{F1} = 510\mathrm{Hz},\ \Delta\mathrm{P} = 3.85 rad\ \approx 5\pi/4$
  - $\mathrm{F2} = 1050\mathrm{Hz},\ \Delta\mathrm{P} = 1.63 rad\ \approx \pi/2$

Formant 1 is close to the optimal phase shift of π, and Formant 2 is less phase shifted so one might expect a noticeable BMLD.

We chose to use ITDs in our experiment to be able to interpret the data with the angle of sound source location. We hence have phase shifts between the formants and the masker that vary continuously between 0 and π. For a target sound source in front of the head, we know that varying the location of the noise source between a 0° and a 60° angle (hence making the ITD of the noise vary in a frequency-dependent way) makes the speech reception performance increase. Moreover, making the ITD of a fixed ITD noise vary between values that correspond to a 0° to a 90° source location angle makes the subject's speech reception performance vary between its minimum and maximum (Bronkhorst and Plomp 1988). We hypothesized that in a similar way, intermediate phase shifts would allow an intermediate BMLD and an intermediate discriminability level of our vowels.

It is interesting to notice that the phase difference between each formant and the noise is between π/2 and π for the Reference, Different 1 and Different 3 vowels, but of π/2 or smaller for Different 2. This might explain why this vowel was more problematic for subjects even in the Opposite condition. For 5 out of 6 subjects that could not discriminate Different 2 from the Reference vowel, the frequency of $\mathrm{F2}_{\mathrm{D2}}$ was made as distinct as possible from $\mathrm{F2}_{\mathrm{R}}$ during the adaptive procedure. For these subjects:

$F2_{D2} = F2_R - 300\mathrm{Hz} = 930\mathrm{Hz}$

so $\Delta\mathrm{P} = 0.73 rad\ \approx \pi/4$

Hence, while the frequency of the second formant was further away from $F2_R$, the phase difference between the second formant and the masker was smaller. This might explain why these subjects could not discriminate Different 2 in the Opposite and Split conditions.

Our experiment can hence be interpreted in terms of BMLD, and studying the phase differences between the vowels and the masker at corresponding frequencies brought additional insight in the data. The phase shifts we computed for our experimental paradigm were valid only when comparing one formant frequency to the masker at the same frequency, but would be different if we considered the masker at other frequencies. It would be interesting to test an alternative paradigm where the whole vowels have a single phase shift to the masker and see whether that changes the discriminability performance. This modification will lead to a loss of the perception of localized sounds, but might increase discriminability.

It would also be interesting to do another experiment where the phase differences between the vowels and the masker are of π for the Opposite condition. We could compare the discriminability of the vowels when their phase is maximally different from that of the masker to their discriminability when their location (or ITD) is maximally distinct from that of the masker. It is probable that the maximal phase shift will lead to a better performance as it was shown that a sentence presented with a phase shift of π from a masker was more intelligible than presented with a single ITD (Levitt and Rabiner, 1967a).

b. Physiology and BMLD

In our physiological experiment, we used the same frequencies for the formant of the vowels and the same masking sounds. However, we used a different ITD value so that the sounds would be localized at 90° from the gerbil head's midline. Using the same method, we can define the phase shifts that were present in our physiological experiments:

$$\Delta \mathrm{P} = 4\pi \mathrm{F} * ITD_M$$

With ITD_M = 160µs to account for the smaller gerbil head size. We hence have:

- Reference vowel:
    - Formant 1: $F1_R = 630\mathrm{Hz}$, $\Delta \mathrm{P} = 1.37 rad$, and the closest rational fraction of pi is $\Delta \mathrm{P} \approx \pi/2$
    - Formant 2: $F2_R = 1230\mathrm{Hz},\ \Delta \mathrm{P} = 2.47 rad\ \approx 4/5\pi$

Formant 1 is phase shifted by less than π/2 and Formant 2 is very close to being in phase with the noise, so there might not be an appreciable BMLD.

- Different 1:
  - $\mathrm{F1} = 510\mathrm{Hz},\ \Delta\mathrm{P} = 1.03 rad\ \approx \pi/3$
  - $\mathrm{F2} = F2_R,\ \Delta\mathrm{P} \approx 4\pi/5$
- Different 2:
  - $\mathrm{F1} = F1_R,\ \Delta\mathrm{P} \approx \pi/2$
  - $\mathrm{F2} = 1110\mathrm{Hz},\ \Delta\mathrm{P} = 2.23 rad\ \approx 2\pi/3$
- Different 3:
  - $\mathrm{F1} = 510\mathrm{Hz}, \Delta\mathrm{P} = 1.03 rad\ \approx \pi/3$
  - $\mathrm{F2} = 1050\mathrm{Hz}, \Delta\mathrm{P} = 2.11 rad\ \approx 2\pi/3$

We notice that the phase shift between the formants and the noise at the same frequency is smaller than in the psychophysical experiment. If phase differences are a more important factor than sound localization for the discriminability of the vowels, that might cause an artificially low performance of our neuronal populations.

Several physiological studies have studied single neuron responses to tones of different phase shifts and ITDs in noise. For example, David McAlpine, Jiang, and Palmer (1996) recorded the activity of single neurons in the IC of anesthetized guinea pigs. They devised an experiment where a noise masker was always presented diotically (N_0) and a pure tone was added either diotically with the same phase (S_0), reversed at both ears (S_π) or at the neuron's best delay (S_{BD}). They observed that the firing rate of each neuron in presence of the tone was more different to their firing rate for noise alone in the N_0S_{BD} condition than in the N_0S_π condition. Hence, there was more unmasking when the tone was played at the neuron's best delay and not when it was maximally phase shifted from the masker. This shows that the relationship between the phase shift of a tone of a specific frequency and each neuron's ITD tuning curve at that frequency might be more important than the absolute phase shift value. In our experiment, a majority of neurons had a best ITD around +/-160µs so applying an ITD of 160µs to the vowels might produce a significant release from masking even though the phase difference between the noise and the vowels is close to 0.

In the majority of physiological BMLD studies, the noise is presented diotically in the N_0 condition and the phase of the target sound is varied. In our stimulus, both the masker and the vowels have an ITD and no sound is diotic. It was shown that neurons in the IC show a release from masking when comparing N_0S_0 and $N_\pi S_0$, even though it is smaller than the one observed between N_0S_0 and N_0S_π (Palmer, Jiang, and McAlpine 2000). It would be interesting to redo our experiment with a diotic masker rather than a masker at -160µs ITD. This would be closer to the classical BMLD paradigms but would mean that the masker comes from a sound source localized in front of the head, which might make it more difficult to separate from the vowels. It seems intuitive that for a single neuron, the condition that will lead to the best performance is $N_{WD}S_{BD}$ where the noise is presented at 'worse delay' (WD) which is an ITD where the neuron fires the least. However at a population level where the neurons have very different ITD tuning curves and frequency tuning, it becomes difficult to predict the neuronal responses and performance levels, especially when using a 4 harmonic stimulus rather than one pure tone.

Caird, Palmer, and Rees (1991) studied BMLDs using a more complex target sound in the guinea pig IC. They first used single pure tones optimized to single neurons' best frequencies and delays and observed a good correspondence between physiological and psychophysical BMLDs. They then used fragments of synthetic vowels instead of pure tones, and could not observe BMLDs in a significant part of their neuronal population. They argued that for a tone of any delay and frequency there will be one cell in the IC with matching properties (that wasn't necessarily present in the population of neuron they recorded) that will show a large unmasking. Such optimized neurons would hence be used for detecting or discriminating the tone. However, we notice that their results are similar to ours in that they observed a wide range of neuronal responses even for a quite simple stimulus, which makes it harder to reconcile the physiological data with psychophysical results.

c. Behavior and BMLD

Most behavioral experiments on non-humans subjects are conducted in free field, which makes it difficult to use a pure BMLD paradigm with phase shifts. However, a few studies proved that animals could take advantage of BMLDs. Cranford (1975) used a shock avoidance paradigm in cats where the animals had to detect a 1kHz tone presented only to one ear. He observed that the detectability of the tone was higher when the noise masker was presented diotically than when it was only presented the same ear as the tone. This is similar to Hirsch's results that showed that sentences are more intelligible in the N_0S_0 condition than in the N_mS_m condition.

Wakeford and Robinson (1974) used cats with pinna inserts and a shock avoidance paradigm to measure the detectability of 0.5kHz, 1kHz and 1.5kHz tone bursts in noise using the conditions N_0S_0, N_0S_π and N_0S_m. Consistently to what was observed in humans, they observed a large release for masking in the N_0S_π condition. Hence, it seems that BMLD has a behavioral relevance for cats which is coherent with the readily observable effects such paradigms have at a single neuron level. It would be interesting to measure BMLDs using the exact same paradigms in behavioral and physiological experiments to see whether the results are similar. It would also be interesting to study it at a population level rather than for select neurons. Our study set out with a similar goal of direct comparison of anesthetized and awake neuronal activity and performance, but we focused on the effects of sound localization rather than phase differences. We could envision doing similar experiments using BMLD paradigms. This would make our results more comparable to previous research but less applicable to real-world listening situations.

2. Physiology of the awake brain

a. Properties of the IC in awake animals

Part of the motivation behind this study was to be able to directly compare neuronal activity in anesthetized, awake and behaving animals. Unfortunately, we did not record from behaving animals but we did record preliminary data from awake passive animals. We observed the same global properties in awake and anesthetized IC recordings, with a majority of single units strongly influenced by frequency and ITD. In the awake preparation, we observed higher spike rates, more inhibitory regions in frequency receptive fields and slow variations of the global neuronal activity.

Other studies have measured binaural properties in the IC of awake animals. Coffey et al. (2006) measured ITD tuning and sensitivity to interaural correlation in the IC of un-anesthetized rabbits. They observed that neurons that were tuned to low frequencies and to ITDs were very sensitive to changes in interaural correlations, which was already observed in the anesthetized guinea pigs (Shackleton, Arnott, and Palmer 2005) and cats (Yin, Chan, and Carney 1987). Interestingly, they observed that the ability of the neuronal population to detect changes in interaural correlation matched the human psychophysical acuity only when they used the best neurons and not all the population. Day and Delgutte (2013) also recorded from the IC of un-anesthetized rabbits. Although they did not comment directly on it, it seems that their neuronal population has similar ITD and frequency tuning properties to what has been observed in anesthetized preparations. The un-anesthetized rabbit is arguably not the best model for awake animals as they are not freely moving and might not be in an alert state, but it is a convenient preparation that allows additional insights into brain function.

From our study and previous literature, it hence seems that the basic binaural properties of the IC are conserved between anesthetized and awake animals. We also observed in our second physiological experiment (Belliveau, Lyamzin, and Lesica 2014) that the properties of neurons in the IC of anesthetized gerbils were robust to different types of anesthesia. Indeed, the cells' ITD tuning and ITD population decoding performance was the same under FMM (fentanyl, medetomidine, and midazolam) and ketamine-xylazine anesthesia. This suggests that the IC is robust to changes in recording conditions both when comparing different anesthetized states and anesthetized and awake recordings. On the other hand, our preliminary data shows global changes in the neuronal activity. It would be interesting to record from more single units to understand whether these changes are reproducible and significant and whether they have an impact on the finer binaural tuning of the cells and on their ITD decoding and vowel discrimination performance.

b. Insights from A1 recordings in awake animals

The differences between anesthetized, awake passive and awake behaving neuronal activity in A1 has been much more studied than in the IC. Indeed, Hubel et al. (1959) already noted that some cells in A1 only responded to auditory stimuli when cats were attentive to the sounds. It is interesting to see which changes happen in cortical structures to be able to apply similar concepts to the midbrain and perhaps understand better the information transfer in the auditory pathway.

Mickey and Middlebrooks (2003) observed the same general binaural tuning in the anesthetized and awake cat A1 with most cells tuned to contralateral ITDs, but they observed that this tuning became sharper in awake animals. They could distinguish between two behavioral states: a performing state where cats were actively performing a sound localization task and an idle state where the same sounds were played but the animals were not engaged in the task. They observed that on average the neurons had a greater modulation depth in response to different sound locations and a sharper ITD tuning in the performing state compared to the idle state. However, only few individual neurons showed a significant difference between the two states. They also observed differences in the temporal encoding of the sounds: in the anesthetized recordings the first spikes after the onset of a sound contained the most information, while in the awake recordings the sustained part of the response contained most information. Interestingly, they observed a very diverse neuronal population and comment on the fact that even units recorded simultaneously from a single electrode can show very different responses. In agreement with our data, it seems that recording the responses from as many cells as possible without preselection leads to very diverse and complex data sets.

Lee and Middlebrooks (2013) observed a larger diversity of responses in awake than in anesthetized cats A1. They also observed differences in the responses to sounds over time: in the anesthetized condition neurons showed a 10ms to 30ms onset spike burst at the beginning of a noise burst while in the awake condition neurons showed many different responses (onset-only, complex onset, long latency, offset-only, etc). They compared the ITD tuning in A1 in a sound identity discrimination task (recognize a click train within noise bursts) and a sound location discrimination task (recognize a noise burst coming from a specific location). They observed that the ITD tuning was overall sharper for awake animals and that the sharpening was larger during the task that involved recognizing sound locations. It hence seems that binaural receptive fields are sharpened in A1 for awake animals, and that this sharpening becomes more significant when the animals are actively engaged in a sound localisation task. It would be interesting to see whether this is also the case in the IC. Neuronal activity in the midbrain is generally considered as less attention dependent but there are many feedback loops within the auditory pathway that might lead to top-down modulations of the activity.

The activity of A1 was also extensively studied using anesthetized, passive and behaving monkeys. Basic properties of A1 seem to be conserved when comparing anesthetized and awake passive animals: the range of observed characteristic frequencies was only slightly broader in awake passive animals and the frequency specific anatomical organization was the same (Recanzone, Guard, and Phan 2000). Binaural properties also seem conserved as measures of sensitivity to ILDs and ITDs showed stronger neuronal responses but similar tuning curves (Brugge and Merzenich 1973) and a preferred spatial tuning to contralateral sound sources (Benson, Hienz, and Goldstein 1981). As this basic knowledge about the awake brain was established, an important question was whether paying attention to sounds influences the neuronal activity in A1. When monkeys were engaged in an auditory reaction time task, the same tonotopic organization of A1 was observed but more exceptions to the tonotopy were noticed as well as more non-monotonic rate-level tuning functions (Pfingst and O’Connor 1981). The evoked responses to tones and white noise were stronger (i.e. higher firing rates during sound presentation) during the task but the spontaneous neural activity was the same as during passive wake (Pfingst, O’Connor, and Miller 1977; Ryan et al. 1984). Interestingly, the spontaneous activity seemed to increase when the animal was engaged in the task in lower stations of the auditory pathway (cochlear nucleus, superior olive, lateral lemniscus and IC; Ryan et al. 1984). These results seem to be generalizable to other tasks as stronger evoked responses in A1 were also observed when monkeys performed auditory discrimination tasks (Miller et al. 1972). More recently, experiments were devised to probe the binaural properties of A1 neurons in awake behaving monkeys. When the animals performed an IPD discrimination task, increases in evoked and spontaneous firing rates were observed. The discharge rate at best IPD increased for most neurons, but no shifts in best IPDs or steepness of the tuning function were observed. Surprisingly, a better IPD discrimination was nonetheless observed using a neurometric measure for more than half the recorded population (Scott, Malone, and Semple 2007). For such a task, it seems that the auditory context has more effect on neuronal tuning: a change in the IPD of sounds presented before the discrimination task changed neuronal tuning by as much as 45° (Malone, Scott, and Semple 2002). Indeed, we know that the immediate history of sound stimulation evokes short-term plasticity and dynamic changes in A1 neuronal responses (Andrew J. King, Schnupp, and Doubell 2001; Fritz et al. 2007; Fritz, Elhilali, and Shamma 2007). It would be interesting to study such dynamic changes in the midbrain, and they could be relevant to our experimental paradigm as we are presenting a repetitive stimulus with only two ITDs and a narrow frequency range.

We saw that evoked and sometimes spontaneous firing rates are higher when animals are behaving than when they are passively awake. However, more precise experiments are needed to understand whether this is an effect of general arousal and attention to the auditory modality or of attention to a specific stimulus. Beaton and Miller (1975) devised a pure tone reaction time task where some monkeys were rewarded for detecting pure tones of all frequencies while others were rewarded only for responses to a subset of frequencies. In the latter condition they observed higher evoked activity, altered latencies and altered discharge patterns in response to tones of rewarded frequencies for

25% of the recorded neurons. There was no change in the spontaneous firing rate between the two conditions. Hence, it seems that rewarding a specific stimulus - which might make the animals more attentive to it - modified the responses of only a quarter of the neuronal population. Benson and Hienz (1978) presented pure tones monaurally to the right and left ear in random succession and trained monkeys to report the detection of pure tones presented to one ear only. They observed that 65% of the recorded single units had higher evoked responses when the tone came from the attended ear, but this difference was significant only for 18% of the units. No changes in spontaneous activity or response patterns were observed between tones coming from the attended or non-attended ear. Later, Benson, Hienz, and Goldstein (1981) compared evoked activity in A1 during a sound detection task and a sound localization task and observed that only 8% of the single units showed a different firing rate in response to sounds between the two tasks. Overall, it seems that neuronal activity is dependent on the global attention state of the animal (behaving VS passive) but that only a subset of A1 neurons has different responses in function of the specific task or attended stimulus. However, neuronal activity was mostly assessed in terms of discharge rate in these studies and finer measures of receptive field shapes or temporal properties might be needed to unveil such dependencies. For example, we saw before that the sharpness of ITD tuning curves in the cat A1 was different between a sound localization and a discrimination task (Lee and Middlebrooks 2013).

Interestingly, there seems to be some multi-sensory integration already at play in A1. Hocherman et al. (1976) trained monkeys to associate one sound to pressing a lever on their right and another sound to pressing a lever on their left, and did the same training independently with two visual stimuli. They then presented concurrently the auditory and visual stimuli while the responses to only one sensory modality were rewarded. This meant that the visual and auditory stimuli might be contradictory but that the correct response would depend only on one modality chosen in each training block. They observed that 2/3 of single units in A1 responded more strongly to sounds when both the auditory and the visual cues pointed to the same side. Interestingly, half of these units showed a higher evoked response when the auditory modality was rewarded and the other half when the visual modality was rewarded. It seems that considering other sensory modalities is also important when analysing activity in A1 of awake animals, and that sensory integration might take place before or in the visual cortex.

The current state of research indicates that basic properties of brain regions such as frequency tuning, tonotopy or tuning to binaural cues are preserved between anesthetized and awake animals, with more diverse responses and more exceptions to the rules being observed in awake experiments. It seems that neurons have higher firing rates in awake animals, especially when they are engaged in a behavioral task, which corresponds to our observations of a higher neuronal activity in awake gerbils. When animals are engaged in a task, it seems that the evoked activity of their neurons are higher and that receptive fields shapes and centers can change in function of the specific task at hand, of the short-term stimulus history and even in function of other sensory modalities. These results are encouraging to keep making experiments in anesthetized animals which provide valid basic knowledge about neuronal activity and in awake behaving animals which provide a more accurate measure of the complexity and adaptability of the brain.

3. Behavior

a. Towards improving the reliability of the behavioral data

We trained gerbils to perform a pure tone discrimination task in a free field arena they could freely explore. While we could show that behavioral training was possible using our novel paradigm, we failed to obtain reliable enough behavioral data to be able to proceed towards more complex auditory tasks or to relate the behavioral data to electrophysiological recordings of neuronal activity.

One salient problem was that the animals seemed to be idle for a large percentage of the time they spent in the behavioral arena. Using a real-time monitoring camera, we could observe the gerbils grooming, resting or exploring the arena in an un-related way to the task (scratching the floor, climbing on the walls, etc.). We did not have a rigorous way to define this idle state and hence could not exclude these data from our analysis. Other studies have reported that animals were idle for part of their behavioral experiments. For example, cats trained to perform an auditory task were idle 38% of the time at the beginning of the behavioral sessions and as much as 62% of the time towards the end of the session when they were satiated (Lee and Middlebrooks 2013). However, the structure of the task used in this study allowed the authors to know when the cats were idle: the animals had to press a pedal to initiate a trial, keep pressing the pedal during sound presentation and release it when the target sound was presented. It was hence very clear that when the animals were not pressing the pedal they were not engaged in the task. The authors could then play the same sounds as during the task, record the neuronal activity and analyse it in this well-defined idle state.

For our task, we could have made the gerbils go to a specific point in the arena to initiate a trial. We could also have installed a nose-poke or elevated platform in the arena that the gerbils would have to go to to initiate a trial. No trial would then start without an active action from the animals and this would surely improve their engagement in the task. We chose not to have the animals initiate the trials at the beginning of our study in hopes to reduce training times because informal reports from colleagues indicated that it took two months to train gerbils to use a nose poke, but this might have prevented us from collecting reliable data.

Ideally, we would also like to be able to monitor the engagement of the gerbils during the trial, which is challenging in our behavioral arena. We considered stopping the trials if gerbils spent more than 3s (i.e. the time they had to spend in the target area for the trial to be completed) in the wrong spot. This would stop the task when the gerbils groom, rest, or ‘make a mistake’ by staying in a wrong place for 3 s. However, this rule might be hard to learn for the animals and does not allow us to detect moments where the gerbils are moving but not engaged in the task. The most reliable way to detect idleness would be to change our behavioral task to a task where the animals have to sustain an action to continue the task, but that would force us to abandon our naturalistic paradigm of free exploration. It seems that trying to implement a behavioral task in a more naturalistic environment had a too high impact on the reliability of the behavioral data, at least for our gerbils.

b. Other animal models

We chose to use gerbils in this study because their audiogram includes low frequencies (Ryan 1976, Rickye S. Heffner and Heffner 1985), they can be trained to perform vowel discrimination tasks (Lingner, Wiegrebe, and Grothe 2012), there are already a lot of studies on the binaural properties of their midbrain (Harris et al. 1997; Maki and Furukawa 2005) and we already had well established electrophysiological recording procedures for the IC in our laboratory. However, this might have been a poor choice from a behavioral point of view.

Cats have even better hearing capacities than gerbils at low frequencies (Heffner and Heffner 1985) and have been extensively used for behavioral and physiological studies. They can discriminate between different spatial locations of target sounds (Mickey and Middlebrooks 2003; Lee and Middlebrooks 2013) and between human vowels. Indeed, cats can perceive differences in the formant frequencies of single vowels (Hienz, Aleszczyk, and May 1996a) and can even discriminate between different human vowels in a noisy environment with a comparable performance to human listeners (Hienz, Aleszczyk, and May 1996b). It was shown that this discrimination performance depends on the brainstem, as lesions in the cochlear nucleus impaired discrimination performance (Hienz, Stiles, and May 1998; May, Prell, and Sachs 1998). The responses of the inferior colliculus were also extensively studied in cats, both for low-frequency binaural stimuli (Yin, Chan, and Irvine 1986) and for complex sounds in noise (Ehret and Merzenich 1988).

Although ferrets have a worse hearing sensitivity at low frequencies (Jack B. Kelly, Kavanagh, and Dalton 1986), they are also a widely used animal model for binaural and speech-like sound discrimination and processing research. Parsons et al. (1999) developed a method for an azimuthal plane sound localization task that was extensively used for ferrets (Nodal et al. 2008; Parsons et al. 1999) and even adapted to sound localization in the vertical plane (Bizley et al. 2007). Ferrets were placed in a circular arena with a ring of speakers on the outside of the arena and a water spout inside the arena directly in front of each speaker. They were trained to lick water from a central spout to initiate a trial, which ensured that they were engaged in the task and that their head was positioned in the center of the arena so that they would receive reproducible spatial cues from each speaker. During the task, a sound was played from one of the speakers and they had to go and collect a water reward at the active speaker’s spout. Ferrets were also used successfully for synthetic vowel discrimination tasks where they were able to discriminate between two vowels in a two-alternative forced choice paradigm and to quickly generalize this to other vowels and listening conditions (Bizley et al. 2013).

Rats have worse detection thresholds for low frequency sounds (H. E. Heffner et al. 1994), but are a very common model for behavioral and physiological studies. Rats were trained to perform discrimination tasks between human vowels and human consonants but the stimuli were shifted one octave up to match their hearing range (Engineer et al. 2008; Perez et al. 2013). While this is very valuable to study the mechanisms of speech processing, shifting the sounds in frequency modifies the binaural cues available for their localisation and makes it difficult to measure responses to binaural cues and to compare them to human psychophysical data.

Another practical consideration is how to motivate the animals to perform the behavioral task. Food deprivation is a commonly used option, especially when using cats. We tried to implement this by food depriving our gerbils to 80% of their body weight but this seemed to increase their stress, aggressivity and agitation levels without increasing their motivation to perform the task. Water deprivation is another commonly used method, so that animals are motivated to obtain small water rewards during the behavioral sessions. We were discouraged to use this technique because the gerbils would not be healthy under a water deprivation regimen, but it seems to have been used successfully in the past (Vanderweele and Abelson 1973).

Ultimately, we could and maybe should have considered using another animal model for this project. This would have made the physiological recordings in the IC more challenging as our laboratory was not experienced or equipped for such experiments but might have made the behavioral experiments more successful. Being able to conduct both physiological and behavioral experiments was the goal which would have allowed us to gain more insight into the processing of speech-like sound and binaural cues in the midbrain of anesthetized, awake and behaving animals.

4. Application to hearing loss

a. Hearing loss and real-world listening

Hearing loss is a prevalent condition with 25% of the European population experiencing age-related hearing loss at 70 years old and as much as 50% at 80 years old. Hearing loss can be due to ageing, noise-induced trauma or central auditory processing disorders but the latter has a lower prevalence of about 12% of American residents older than 65 years old (Quaranta et al. 2015). Hearing loss has been linked to social isolation (Mick, Kawachi, and Lin 2014), is correlated with cognitive decline (Lin et al. 2013), and has been shown to be correlated with a lower quality of life (Mulrow 1990), hence severely impacting the lives of concerned people.

The most common treatment for hearing loss is hearing aids, which have been shown to restore some of the patients' hearing capabilities (Mulrow 1990). Despite the measured improvement in some listening situations, several factors cause people not to wear their hearing aids. As reviewed in McCormack and Fortnum (2013), the leading factors are discomfort when wearing the hearing aids and the lack of positive improvement in real-world listening situations. One factor mentioned in several studies that is of particular relevance to us is that hearing aid users are not able to understand speech in a noisy environment (Bertoli et al. 2009; Hartley et al. 2010; Vuorialho, Karinen, and Sorri 2006). Indeed, Larson et al. (2000) showed that speech recognition improvements measured with hearing aids were smaller when the overall sound levels were higher, and so for three commonly used sound processing circuits. They observed a mean improvement of 22% for understanding speech in babble noise when the overall sound level was of 52dB SPL and of only 4% at 74dB SPL. This highlights the need for conducting hearing tests and neuroscientific research in more realistic listening environments including background noise and high sound levels

b. Designing hearing aids in function of the type of deficits

Designing hearing aids is a complex process. Even if the loss of audibility is one of the primary losses due to cochlear damage, it is not sufficient to linearly amplify the sounds that reach the damaged cochlea to restore normal hearing (Moore 1996). The signals reaching the hearing aid can be modified before being sent to the cochlea. For example, it is possible to frequency-shape the signal to improve the contrast between the formants of vowels and other harmonics. However, it was shown that while this restores some of the phase locking of auditory nerve fibers (ANFs) to the formant frequency, it also increases their phase locking to other harmonics in the spectral trough and hence is unlikely to improve speech intelligibility (Schilling et al. 1998). This is particularly relevant in the context of listening to speech in noise as it has been shown that while frequency smearing has little influence on speech intelligibility in quiet, it does significantly reduce speech intelligibility in noise (Baer and Moore 1994).

It is also important to define which ANFs one is targeting with hearing aids. Indeed, ANFs with a low to medium spontaneous firing rate respond only to high intensity sounds and seem to play a major role in responses to high intensity tones in noise (Costalupes, Young, and Gibson 1984). Noise-induced hearing loss might be due to a selective loss of these high threshold ANFs (Furman, Kujawa, and Liberman 2013). This might explain why noise-induced hearing loss is often not detected by standard hearing tests that measure the audibility of single tones in quiet. These low intensity single tones could be perceived through low threshold ANFs that were not be affected by the noise-induced hearing loss. Hence, it seems that more precise studies and models of the responses of healthy and damaged ANFs in response to speech in noise might help to develop better hearing aids

(Sachs et al. 2002). It is likely that different types of ANFs project to different regions of the mid-brain, and it seems necessary to study mid-brain processing in light of these data to understand the repercussions of hearing loss.

It has already been shown that changes in sound processing mechanisms in the brain are linked to hearing loss. For example, temporal processing of speech in the brainstem is less efficient with ageing (Walton 2010). When animals and humans age, there is a shift in the excitation/inhibition balance in the brain and in particular a loss of inhibition in the SOC and IC (Caspary et al. 2008). This might participate in the loss of temporal accuracy in the neural networks and impair the animal’s abilities to localize sounds in their environment as sound localization processing in the brainstem and mid-brain might rely on precisely timed inhibition.

c. Influence of hearing aids on sound localization acuity

Hearing impaired patients often report difficulties in understanding speech in a group setting within a noisy environment, even when using hearing aids they are already accustomed to (Harkins and Tucker 2007). This impairment in understanding might be due to an inability to localize relevant sound sources and to binaural processing deficits. The problem of sound localization in hearing aid users has been discussed for a long time: Durlach, Thompson, and Colburn (1981) already noted the necessity to conduct several types of auditory tests to understand the spatial localization deficits of hearing impaired patients and stressed the importance of understanding whether they came from reduced sensibility to sounds or from sound processing impairments.

Several studies have showed that hearing impaired patients have a lower performance and higher thresholds than normal hearing listeners in experimental tasks that involve binaural cues and spatial localization (see Akeroyd 2014 for a review). For example, minimal audible angles (MAAs) for white-noise stimuli coming from the side of the head’s midline were worse for hearing impaired patients (Häusler, Colburn, and Marr 1983). Interestingly, while all the normal hearing subjects showed MAAs of 12° or less, half the hearing impaired patients showed MAAs superior or equal to 30° but some patients had MAAs of only 7°. They also noted that the MAAs for hearing impaired and normal hearing subjects were comparable when the stimuli came from the front of the head instead of from the sides. This highlights the complexity of the relationship between hearing impairment and sound localization, even for simple paradigms such as MAA measures with white-noise bursts. The fact that the responses of hearing impaired patients varied from normal to severely deficient shows that one has to be careful not to assume the average results for hearing impaired listeners applies to every individual patient.

Sound localization performance in hearing impaired patients was studied more directly by Lorenzi, Gatehouse, and Lever (1999b) and compared to the performance of normal hearing listeners (Lorenzi, Gatehouse, and Lever 1999a). They observed that the horizontal localization of click trains in noise became more difficult for hearing impaired patients at a higher SNR than for normal hearing listeners. Hearing impaired subjects also had a worse and more variable performance when the noise was presented away from the midline rather than in front of the head. Localization performance can also be impaired in terms of distance perception: hearing impaired listeners showed larger just noticeable differences for the distance separating two sounds than normal hearing listeners (Akeroyd, Gatehouse, and Blaschke 2007). Localization impairment can also be manifested in terms of phase shift detection: normal hearing listeners could detect IPDs at higher frequencies than hearing impaired listeners (Neher et al. 2011). In keeping with these results using simple sound stimuli, hearing impaired listeners showed poorer spatial localization of words masked by other words than normal hearing listeners in more naturalistic experimental paradigms (Best et al. 2011).

The type and frequency of hearing loss can be correlated with different sound localization deficits: patients with high-frequency hearing loss seemed to have more difficulties localizing sounds in the vertical plane while patients with low-frequency hearing loss seemed to have more difficulties localizing sounds in the horizontal plane and understanding speech (Noble, Byrne, and Lepage 1994; Häusler, Colburn, and Marr 1983). This corresponds with our understanding of the mechanisms of sound localization: the localization of sounds in the vertical plane using spectral cues relies mostly on frequencies from 4kHz to 16kHz (Hebrank and Wright 1974) and speech intelligibility relies mostly on frequencies below 3kHz (Vickers, Moore, and Baer 2001). Patients with conductive hearing loss had more difficulty localizing sounds than patients with sensory-neural hearing loss, even after normalizing to the same sound reception level impairment (Noble, Byrne, and Lepage 1994; Häusler, Colburn, and Marr 1983). This might be because more auditory information came directly through the bones of the head and skull for patients with conductive hearing loss, thus preventing them from taking advantage of air borne binaural cues.

Overall, hearing impaired patients show deficits in many different aspects of sound localization. Even though the observed impairments can sometimes be correlated to hearing loss in a specific frequency range or to a specific type of hearing impairment, the performances of patients are often highly variable from individual to individual and are hard to predict from other simpler auditory tests, making the diagnostic and optimal treatment recommendation difficult.

d. Bilateral hearing aids

Fitting patients with bilateral hearing aids is a common strategy to try to restore their sound localization performance: they are fitted with one hearing aid on each ear, which might allow them to take advantage of binaural cues. Unfortunately, the positive effects of bilateral fitting for sound localization are still unclear. Byrne and Noble (1998) observed a better localization performance of bursts of pink noise in quiet for patients with severe hearing loss fitted with bilateral hearing aids rather than fitted with monaural hearing aids. However, they did not observe any beneficial effect of bilateral fittings for patients with mild to moderate hearing loss and hypothesized that the benefit observed for patients with severe hearing loss was due to a better sensitivity to sound thanks to having two aided ears rather than to a restoration of binaural cues. Other studies have confirmed that while there are sometimes localization benefits obtained from bilateral hearing aids, they are very variable from patient to patient and cannot be predicted from previous auditory testing using headphones in a laboratory environment (Boymans et al. 2008). The beneficial effects of bilateral hearing aids seems to be even smaller in complex listening situations: the localization performance of hearing impaired patients was sometimes better without than with hearing aids (Van den Bogaert et al. 2006). Yet, bilateral hearing aids seemed to have an overall positive impact on spatial release from masking for patients of a wide range of ages and cognitive abilities (Dawes et al. 2013).

Sound localization in the horizontal plane hence seems to be better in some cases with bilateral hearing aids, depending on the experimental paradigm and on the type of hearing loss, but the beneficial effects of bilateral fittings are still controversial. For localization in the sagittal plane, it seems that using open ear molds in unilateral or bilateral fittings is sufficient to allow the normal patterns of spectral alterations by the pinna and restore vertical sound localization (Byrne and Noble 1998).

A very important issue with bilateral hearing aids is that they in fact do not restore the binaural cues necessary to sound localization. For example, the brain is sensitive to

ITDs as small as 10µs (Klumpp and Eady 1956) yet hearing aids have an internal time delay that varies between 3ms and 10ms in function of the brand and model, and sometimes even vary in function of frequency (Dillon et al. 2003). One can assume that even if the two hearing aids fitted to one patient are exactly the same, this intrinsic delay will not conserve ITDs between the two ears with enough precision. Moreover, many different algorithms are used in hearing aids to process the incoming signals and make speech more intelligible but they tend to distort binaural cues. Frequency-lowering programs are used to reduce the high-frequency content of the input and thus emphasise the low frequencies used for speech reception. This was shown to produce intelligibility increases in some patients (Simpson 2009) but also produces distortions of high-frequency envelope ITDs and reduces interaural coherence (Brown et al. 2016). Digital noise reduction algorithms were shown to improve ease of listening and listening comfort in laboratory conditions but did not improve significantly the listening performance and comfort of patients at home (Bentler et al. 2008). Interestingly, this technique seems to reduce the annoyance due to noise rather than increase speech intelligibility per se (Brons, Houben, and Dreschler 2014).

In general, it was shown that nonlinear wide dynamic range compression algorithms alter the relationship between ITDs and ILDs, sometimes resulting in conflicting localization cues, and thus most likely impairing sound localization (Brown et al. 2016). Phase preserving programs were developed to restore the correct phase relationships between the two ears. Non-linear phase preserving programs did improve the ability of patients to localize and understand single words in noise, but only after a 16-week usage period (Drennan et al. 2005). A noise reduction algorithm that preserves ITD cues when allowing a fraction of the noise signal to be perceived was developed (Klasen et al. 2007) but the authors remarked that while this approach was promising to allow hearing impaired listeners to benefit from binaural cues, there was still a lot of work to be done. Indeed, the algorithm was tested in a quite simple noise situation, in an anechoic environment, with a perfect voice detection algorithm and without taking head movements into account. This might make the benefits of using such algorithm small or non-existent for patients in a real-world listening situation.

e. Directional microphones

A successful strategy to improve speech intelligibility in noise for bilateral hearing aid users is to use directional microphones and beamforming. The directional microphones are directed to the front of the head and hence receive only sounds coming from the front and not the ambient noise coming from other directions. This directly improves the SNR of sounds coming from the front of the head, which are generally the signals of interest. The beamforming algorithms reinforce this directionality by combining the signals from the two ears in such a way that signals coming from the front experience constructive phase interferences and signals coming from other angles experience destructive interferences. Some algorithms do so while trying to preserve natural binaural cues (Van den Bogaert et al. 2008). This approach was successfully tested on hearing impaired patients: using bilateral hearing aids with directional microphones rather than unilateral hearing aids or omnidirectional microphones improved SNR in rooms with low reverberation (Hawkins and Yacullo 1984) and allowed better sentence recognition in noise in reverberant environments (Picou, Aspell, and Ricketts 2014). Multi-array of directional microphones were also built, and they were shown to attenuate reverberation, improve SNR and improve speech reception thresholds in diffuse noise (Soede, Bilsen, and Berkhout 1993).

There is hence strong evidence that using bilateral directional microphones is beneficial for understanding speech in noise but the effects of such technology on sound localization performance is more complex. On the one hand, using directional microphones was shown to reduce front-back localization errors in a noisy environment (Keidser et al. 2006), especially when using frequency-dependent unidirectional microphones (Keidser et al. 2009). Using directional microphones also improved sound localization on the horizontal plane, but only when visual cues were reinforcing the auditory cues (Picou, Aspell, and Ricketts 2014). On the other hand, a lot of localization errors were observed on the horizontal plane (Keidser et al. 2006), especially for sounds presented at an angle superior to 60° away from the midline of the head (Picou, Aspell, and Ricketts 2014). In fact, even the most modern beamforming algorithms still seem to alter binaural cues and produce contradictory ITDs and ILDs that point to different sound source locations (Brown et al. 2016).

For sounds that come from a source on the side of the head, patients might have an orienting behavior towards the sound of interest and hence place it in front of their head where localization, SNR and intelligibility are best. However, this strategy seems to work well only to orient towards sounds that are close to the midline. For off-axis sounds, patients seem to engage in a search behavior that can start with a localization error (starting to orient towards a direction opposite to the sound source) and hence make the time to orient towards the sound source larger when patients use their hearing aids than when they don’t (Brimijoin et al. 2014).

f. Application to hearing aid development

Overall, it seems that bilateral hearing aids can help to restore speech intelligibility in noise, especially when using them in combination with directional microphones and beamforming algorithms. However, this seems to improve sensitivity to speech and SNRs rather than to allow patients to take advantage of binaural cues. Hearing aids are not temporally precise and most algorithms used to improve speech intelligibility distort binaural cues, so even if phase restoring algorithms are being developed there is still a lot of work to do to allow patients fitted with binaural hearing aids to truly take advantage of their two ears.

Based on our research, we can make suggestions on which features to implement with the highest priority in hearing aids. We showed that normal hearing subjects benefitted from spatial release from masking when using very simple stimuli containing only frequency below 1300Hz and applying a constant ITD to all the harmonics. Edmonds and Culling (2005b) also showed that unmasking only the low frequency part of a sentence was sufficient to benefit from spatial release from masking. They observed that applying different ITDs to the sentence and the noise masker only below 750Hz yielded intermediate speech reception thresholds, and doing so only below 1500Hz yielded speech reception thresholds that were almost equal to the fully unmasked condition where the sentence and the masker had different ITDs on the whole frequency range. We would hence suggest focusing on restoring the ITDs of signals below 1.5kHz while making sure that the higher frequency signals don’t give contradictory cues.

In our study, we showed that applying a single ITD to several harmonics was sufficient for subjects to benefit from spatial release from masking. We didn’t need to apply frequency-dependent ITDs to match the ITD pattern from a natural sound source (Algazi et al. 2002). It was indeed already shown that frequency-dependent ITD patterns do not play a significant role for spatial release from masking (Bronkhorst and Plomp 1988). We also showed that the starting phase of the harmonics didn’t influence the performance of our subjects and that only the ongoing phase differences seemed to be relevant.

We computed that applying a single ITD to harmonics of different frequencies created a different phase shift between each harmonic and the noise masker presented at a single ITD. We showed that our subjects benefited from spatial unmasking when presented simultaneously with harmonics that had different phase relationships to the masker, even if an appreciable phase difference between each harmonic and the masker was probably necessary (see VI.1.a on Distractor 2). Most importantly, we showed that spatial unmasking occurred when harmonics falling in the same auditory filter had the same ITDs but also when harmonics falling in different auditory filters had different ITDs. Similarly, it was already shown that normal hearing listeners benefit from spatial release from masking if the IPDs applied to a sentence are coherent within each auditory filter but that the IPDs across auditory filters can be different (Beutelmann, Brand, and Kollmeier 2009). To benefit from spatial release from masking, it is hence crucial to apply coherent ITDs to sounds that fall in the same auditory filter, but it is not necessary to apply coherent ITDs over the whole frequency range or to have a naturalistic frequency-dependent ITD variation. For hearing aid development, we would hence recommend trying to preserve ongoing ITDs for only a few narrow non-overlapping frequency channels. This way, the speech of interest and the ambient noise would have different ITDs on each channel but it would not be necessary for the speech to have the same ITD on all channels. Removing the constraint of coherent ITDs for each sound source over the whole frequency range might make the development of phase preserving algorithms easier and more successful.

Since spatial release from masking seems to be appreciable using only ITDs, we would suggest cancelling all ILDs between the two ears in patients fitted with bilateral hearing aids. Indeed, ILDs and the better ear effect result mostly in better a SNR at one ear. However, directional microphones and beamforming algorithms already optimize the SNR of a frontal signal of interest so the cues provided by ILDs might be less relevant. Equalizing ILDs between the two ears might be an easy algorithm to construct and would prevent contradictions between ILD and ITD cues.

In conclusion, we suggest developing a binaural hearing aid program that preserves ITDs in narrow and sparse frequency channels below 1.5kHz. It would only need to preserve ongoing phase differences without taking into account the starting phases, and would only need to preserve ITDs within each frequency channel without requiring coherent ITDs for each sound source across channels. The program could also cancel ILDs to avoid contradictions between different binaural cues. If this could be achieved, we predict that patients would benefit from spatial release from masking. This would allow for better speech intelligibility in noise but would result in a diffuse perception of sound source location, unfortunately not allowing patients to correctly localize sound sources.

As we can see, understanding which aspects of binaural cues are the most important for spatial release from masking and speech intelligibility in noise is crucial to be able to focus on transmitting the most relevant binaural signals to hearing impaired patients through bilateral hearing aids. The development of hearing aids based on state of the art research will hence hopefully improve the hearing capacities of hearing impaired patients in difficult real-world situations and thus their quality of life.

VII. Bibliography

Agapiou, John P., and David McAlpine. 2008. "Low-Frequency Envelope Sensitivity Produces Asymmetric Binaural Tuning Curves." *Journal of Neurophysiology* 100 (4): 2381–96. doi:10.1152/jn.90393.2008.

Aitkin, L. M., W. R. Webster, J. L. Veale, and D. C. Crosby. 1975a. "Inferior Colliculus. I. Comparison of Response Properties of Neurons in Central, Pericentral, and External Nuclei of Adult Cat." *Journal of Neurophysiology* 38 (5): 1196–1207.

Aitkin, L M, W. R. Webster, J. L. Veale, and D. C. Crosby. 1975b. "Inferior Colliculus. I. Comparison of Response Properties of Neurons in Central, Pericentral, and External Nuclei of Adult Cat." *Journal of Neurophysiology* 38 (5): 1196–1207.

Akeroyd, Michael A. 2004. "The across Frequency Independence of Equalization of Interaural Time Delay in the Equalization-Cancellation Model of Binaural Unmasking." *The Journal of the Acoustical Society of America* 116 (2): 1135. doi:10.1121/1.1768959.

———. 2014. "An Overview of the Major Phenomena of the Localization of Sound Sources by Normal-Hearing, Hearing-Impaired, and Aided Listeners." *Trends in Hearing* 18 (October): 2331216514560442. doi:10.1177/2331216514560442.

Akeroyd, Michael A., Stuart Gatehouse, and Julia Blaschke. 2007. "The Detection of Differences in the Cues to Distance by Elderly Hearing-Impaired Listeners." *The Journal of the Acoustical Society of America* 121 (2): 1077–89.

Algazi, V. Ralph, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang. 2002. "Approximating the Head-Related Transfer Function Using Simple Geometric Models of the Head and Torso." *The Journal of the Acoustical Society of America* 112 (5): 2053–64. doi:10.1121/1.1508780.

Arbogast, Tanya L., Christine R. Mason, and Gerald Kidd. 2002. "The Effect of Spatial Separation on Informational and Energetic Masking of Speech." *The Journal of the Acoustical Society of America* 112 (5): 2086. doi:10.1121/1.1510141.

Baer, Thomas, and Brian C. J. Moore. 1994. "Effects of Spectral Smearing on the Intelligibility of Sentences in the Presence of Interfering Speech." *The Journal of the Acoustical Society of America* 95 (4): 2277–80. doi:10.1121/1.408640.

Batra, R., S. Kuwada, and T. R. Stanford. 1989. "Temporal Coding of Envelopes and Their Interaural Delays in the Inferior Colliculus of the Unanesthetized Rabbit." *Journal of Neurophysiology* 61 (2): 257–68.

Beaton, Randal, and Josef M. Miller. 1975. "Single Cell Activity in the Auditory Cortex of the Unanesthetized, Behaving Monkey: Correlation with Stimulus Controlled Behavior." *Brain Research* 100 (3): 543–62. doi:10.1016/0006-8993(75)90157-2.

Belliveau, Lucile AC, Dmitry R. Lyamzin, and Nicholas A. Lesica. 2014. "The Neural Representation of Interaural Time Differences in Gerbils Is Transformed from Midbrain to Cortex." *The Journal of Neuroscience* 34 (50): 16796–16808.

Benson, D. A., and R. D. Hienz. 1978. "Single-Unit Activity in the Auditory Cortex of Monkeys Selectively Attending Left vs. Right Ear Stimuli." *Brain Research* 159 (2): 307–20. doi:10.1016/0006-8993(78)90537-1.

Benson, D. A., R. D. Hienz, and M. H. Goldstein Jr. 1981. "Single-Unit Activity in the Auditory Cortex of Monkeys Actively Localizing Sound Sources: Spatial Tuning and Behavioral Dependency." *Brain Research* 219 (2): 249–67. doi:10.1016/0006-8993(81)90290-0.

Benson, D. A., and D. C. Teas. 1976. "Single Unit Study of Binaural Interaction in the Auditory Cortex of the Chinchilla." *Brain Res.* 103: 313–38.

Bentler, Ruth, Yu-Hsiang Wu, Jerrica Kettel, and Richard Hurtig. 2008. “Digital Noise Reduction: Outcomes from Laboratory and Field Studies.” *International Journal of Audiology* 47 (8): 447–60. doi:10.1080/14992020802033091.

Bernstein, Leslie R., and Constantine Trahiotis. 1985. “Lateralization of Low-frequency, Complex Waveforms: The Use of Envelope-based Temporal Disparities.” *The Journal of the Acoustical Society of America* 77 (5): 1868–80. doi:10.1121/1.391938.

Bertoli, Sibylle, Katharina Staehelin, Elisabeth Zemp, Christian Schindler, Daniel Bodmer, and Rudolf Probst. 2009. “Survey on Hearing Aid Use and Satisfaction in Switzerland and Their Determinants.” *International Journal of Audiology* 48 (4): 183–95. doi:10.1080/14992020802572627.

Best, V., E. Ozmeral, F. J. Gallun, K. Sen, and B. G. Shinn-Cunningham. 2005. “Spatial Unmasking of Birdsong in Human Listeners: Energetic and Informational Factors.” *The Journal of the Acoustical Society of America* 118: 3766.

Best, Virginia, Simon Carlile, Norbert Kopčo, and André van Schaik. 2011. “Localization in Speech Mixtures by Listeners with Hearing Loss.” *The Journal of the Acoustical Society of America* 129 (5): EL210–EL215.

Beutelmann, Rainer, and Thomas Brand. 2006. “Prediction of Speech Intelligibility in Spatial Noise and Reverberation for Normal-Hearing and Hearing-Impaired Listeners.” *The Journal of the Acoustical Society of America* 120 (1): 331–42. doi:10.1121/1.2202888.

Beutelmann, Rainer, Thomas Brand, and Birger Kollmeier. 2009. “Prediction of Binaural Speech Intelligibility with Frequency-Dependent Interaural Phase Differences.” *The Journal of the Acoustical Society of America* 126 (3): 1359. doi:10.1121/1.3177266.

Bizley, Jennifer K., Fernando R. Nodal, Carl H. Parsons, and Andrew J. King. 2007. “Role of Auditory Cortex in Sound Localization in the Midsagittal Plane.” *Journal of Neurophysiology* 98 (3): 1763–74. doi:10.1152/jn.00444.2007.

Bizley, Jennifer K., Kerry M. M. Walker, Andrew J. King, and Jan W. H. Schnupp. 2013. “Spectral Timbre Perception in Ferrets: Discrimination of Artificial Vowels under Different Listening Conditions.” *The Journal of the Acoustical Society of America* 133 (1): 365–76. doi:10.1121/1.4768798.

Blauert, Jens. 1997. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT press. http://books.google.co.uk/books?hl=en&lr=&id=wBiEKPhw7r0C&oi=fnd&pg=PA1&dq=3+!+.+Spatial+Hearing:+The+Psychophysics+of+Human+Sound+Localization+~+M&ots=4Iu44V7V6c&sig=BXGq9zC45Is_E0rYkGZa4HWuGpI.

Boymans, Monique, S. Theo Goverts, Sophia E. Kramer, Joost M. Festen, and Wouter A. Dreschler. 2008. “A Prospective Multi-Centre Study of the Benefits of Bilateral Hearing Aids.” *Ear and Hearing* 29 (6): 930–941.

Brand, Antje, Oliver Behrend, Torsten Marquardt, David McAlpine, and Benedikt Grothe. 2002. “Precise Inhibition Is Essential for Microsecond Interaural Time Difference Coding.” *Nature* 417 (6888): 543–47. doi:10.1038/417543a.

Bremen, Peter, and Philip X Joris. 2013. “Axonal Recordings from Medial Superior Olive Neurons Obtained from the Lateral Lemniscus of the Chinchilla (Chinchilla Laniger).” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (44): 17506–18. doi:10.1523/JNEUROSCI.1518-13.2013.

Brette, Romain. 2010. “On the Interpretation of Sensitivity Analyses of Neural Responses.” *The Journal of the Acoustical Society of America* 128 (5).

Brimijoin, W. Owen, William M. Whitmer, David McShefferty, and Michael A. Akeroyd. 2014. "The Effect of Hearing Aid Microphone Mode on Performance in an Auditory Orienting Task." *Ear and Hearing* 35 (5): e204-212. doi:10.1097/AUD.0000000000000053.

Broadbent, D. E., and Peter Ladefoged. 1957. "On the Fusion of Sounds Reaching Different Sense Organs." *The Journal of the Acoustical Society of America* 29 (6): 708–10. doi:10.1121/1.1909019.

Bronkhorst, Adelbert W. 2000. "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions." *Acta Acustica United with Acustica* 86 (1): 117–28.

Bronkhorst, AW. 2000. "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions." *Acta Acustica United with Acustica*.

Bronkhorst, and R. Plomp. 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): 1508–16. doi:10.1121/1.395906.

Brons, I., R. Houben, and W. A. Dreschler. 2014. "Effects of Noise Reduction on Speech Intelligibility, Perceived Listening Effort, and Personal Preference in Hearing-Impaired Listeners." *Trends in Hearing* 18 (0). doi:10.1177/2331216514553924.

Brown, Andrew D., Francisco A. Rodriguez, Cory D. F. Portnuff, Matthew J. Goupell, and Daniel J. Tollin. 2016. "Time-Varying Distortions of Binaural Information by Bilateral Hearing Aids: Effects of Nonlinear Frequency Compression." *Trends in Hearing* 20 (October). doi:10.1177/2331216516668303.

Brugge, J. F., N. A. Dubrovsky, L M Aitkin, and D. J. Anderson. 1969. "Sensitivity of Single Neurons in Auditory Cortex of Cat to Binaural Tonal Stimulation ; Effects of Varying Interaural Time and Intensity." *Journal of Neurophysiology* 32: 1005–24.

Brugge, J. F., and M Merzenich. 1973. "Responses of Neurons in Auditory Cortex of the Macaque Monkey to Monaural and Binaural Stimulation." *Journal of Neurophysiology* 36 (6): 1138–58.

Brugge, John F., and Michael M. Merzenich. 1973. "Response of Neurons in Auditory Cortex of the Macaque Monkey to Monaural and Binaural Stimulation." *Journal of Neurophysiology*. http://psycnet.apa.org/psycinfo/1974-26872-001.

Brungart, Douglas S., and Nandini Iyer. 2012. "Better-Ear Glimpsing Efficiency with Symmetrically-Placed Interfering Talkers." *The Journal of the Acoustical Society of America* 132 (4): 2545–56. doi:10.1121/1.4747005.

Budinger, E, P Heil, and H Scheich. 2000. "Functional Organization of Auditory Cortex in the Mongolian Gerbil (Meriones Unguiculatus). IV. Connections with Anatomically Characterized Subcortical Structures." *The European Journal of Neuroscience* 12 (7): 2452–74.

Byrne, Denis, and William Noble. 1998. "Optimizing Sound Localization with Hearing Aids." *Trends in Amplification* 3 (2): 51. doi:10.1177/108471389800300202.

Caird, D. M., A. R. Palmer, and A. Rees. 1991. "Binaural Masking Level Difference Effects in Single Units of the Guinea Pig Inferior Colliculus." *Hearing Research* 57 (1): 91–106. doi:10.1016/0378-5955(91)90078-N.

Cant, Nell B, and Christina G Benson. 2006. "Organization of the Inferior Colliculus of the Gerbil (Meriones Unguiculatus): Differences in Distribution of Projections from the Cochlear Nuclei and the Superior Olivary Complex." *The Journal of Comparative Neurology* 495 (5): 511–28. doi:10.1002/cne.20888.

Carney, Laurel H, Srijata Sarkar, Kristina S Abrams, and Fabio Idrobo. 2011. "Sound-Localization Ability of the Mongolian Gerbil (Meriones Unguiculatus) in a Task with a Simplified Response Map.” *Hearing Research* 275 (1–2). Elsevier B.V.: 89–95. doi:10.1016/j.heares.2010.12.006.

Caspary, Donald M., Lynne Ling, Jeremy G. Turner, and Larry F. Hughes. 2008. “Inhibitory Neurotransmission, Plasticity and Aging in the Mammalian Central Auditory System.” *Journal of Experimental Biology* 211 (11): 1781–91. doi:10.1242/jeb.013581.

Chandrasekaran, Bharath, Erika Skoe, and Nina Kraus. 2013. “An Integrative Model of Subcortical Auditory Plasticity.” *Brain Topography* 27 (4): 539–52. doi:10.1007/s10548-013-0323-9.

Chase, Steven M, and Eric D Young. 2006. “Spike-Timing Codes Enhance the Representation of Multiple Simultaneous Sound-Localization Cues in the Inferior Colliculus.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 26 (15): 3889–98. doi:10.1523/JNEUROSCI.4986-05.2006.

Cherry, EC. 1953. “Some Experiments on the Recognition of Speech, with One and with Two Ears.” *The Journal of the Acoustical Society of America* 5 (25): 975–79.

Choi, Joon Hwan, Hae Kyung Jung, and Taejeong Kim. 2006a. “A New Action Potential Detector Using the MTEO and Its Effects on Spike Sorting Systems at Low Signal-to-Noise Ratios.” *Biomedical Engineering, IEEE Transactions on* 53 (4): 738–746.

———. 2006b. “A New Action Potential Detector Using the MTEO and Its Effects on Spike Sorting Systems at Low Signal-to-Noise Ratios.” *IEEE Transactions on Bio-Medical Engineering* 53 (4): 738–46. doi:10.1109/TBME.2006.870239.

Coffey, Charles S., Charles S. Ebert Jr., Allen F. Marshall, John D. Skaggs, Stephanie E. Falk, William D. Crocker, James M. Pearson, and Douglas C. Fitzpatrick. 2006. “Detection of Interaural Correlation by Neurons in the Superior Olivary Complex, Inferior Colliculus and Auditory Cortex of the Unanesthetized Rabbit.” *Hearing Research* 221 (1–2): 1–16. doi:10.1016/j.heares.2006.06.005.

Costalupes, J. A., E. D. Young, and D. J. Gibson. 1984. “Effects of Continuous Noise Backgrounds on Rate Response of Auditory Nerve Fibers in Cat.” *Journal of Neurophysiology* 51 (6): 1326–44.

Cranford, J. L. 1975. “Auditory Masking-Level Differences in the Cat.” *Journal of Comparative and Physiological Psychology* 89 (3): 219–23. doi:10.1037/h0076824.

Culling, J. F., and Q. Summerfield. 1995. “Perceptual Separation of Concurrent Speech Sounds: Absence of across-Frequency Grouping by Common Interaural Delay.” *The Journal of the Acoustical Society of America* 98: 785.

Culling, John F., H. Steven Colburn, and Matthew Spurchise. 2001. “Interaural Correlation Sensitivity.” *The Journal of the Acoustical Society of America* 110 (2): 1020–29. doi:10.1121/1.1383296.

Culling, John F., and C. J. Darwin. 1993. “Perceptual Separation of Simultaneous Vowels: Within and Across-formant Grouping by F0.” *The Journal of the Acoustical Society of America* 93 (6): 3454–67. doi:10.1121/1.405675.

Culling, John F., Kathryn I. Hodder, and Chaz Yee Toh. 2003. “Effects of Reverberation on Perceptual Segregation of Competing Voices.” *The Journal of the Acoustical Society of America* 114 (5): 2871–76. doi:10.1121/1.1616922.

Dannenbring, Gary L., and Albert S. Bregman. 1978. “Streaming vs. Fusion of Sinusoidal Components of Complex Tones.” *Perception & Psychophysics* 24 (4): 369–376.

Darwin, C J. 2008. “Listening to Speech in the Presence of Other Sounds.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1493): 1011–21. doi:10.1098/rstb.2007.2156.

Darwin, C. J., and R. W. Hukin. 1997. “Perceptual Segregation of a Harmonic from a Vowel by Interaural Time Difference and Frequency Proximity.” *The Journal of the Acoustical Society of America* 102: 2316.

———. 1998. “Perceptual Segregation of a Harmonic from a Vowel by Interaural Time Difference in Conjunction with Mistuning and Onset Asynchrony.” *The Journal of the Acoustical Society of America* 103: 1080.

———. 1999. “Auditory Objects of Attention: The Role of Interaural Time Differences.” *Journal of Experimental Psychology: Human Perception and Performance* 25 (3): 617.

Darwin, Christopher J., Douglas S. Brungart, and Brian D. Simpson. 2003. “Effects of Fundamental Frequency and Vocal-Tract Length Changes on Attention to One of Two Simultaneous Talkers.” *The Journal of the Acoustical Society of America* 114 (5): 2913–22. doi:10.1121/1.1616924.

Darwin, C.J. 2008. “Listening to Speech in the Presence of Other Sounds.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1493): 1011–21. doi:10.1098/rstb.2007.2156.

Dawes, Piers, Kevin J. Munro, Sridhar Kalluri, and Brent Edwards. 2013. “Unilateral and Bilateral Hearing Aids, Spatial Release from Masking and Auditory Acclimatization.” *The Journal of the Acoustical Society of America* 134 (1): 596–606.

Day, Mitchell L., and Bertrand Delgutte. 2013a. “Decoding Sound Source Location and Separation Using Neural Population Activity Patterns.” *The Journal of Neuroscience* 33 (40): 15837–47. doi:10.1523/JNEUROSCI.2034-13.2013.

Day, Mitchell L, and Bertrand Delgutte. 2013b. “Decoding Sound Source Location and Separation Using Neural Population Activity Patterns.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (40): 15837–47. doi:10.1523/JNEUROSCI.2034-13.2013.

Dillon, Harvey, Gitte Keidser, Anna O’brien, and Heidi Silberstein. 2003. “Sound Quality Comparisons of Advanced Hearing Aids.” *The Hearing Journal* 56 (4): 30–32.

Dirks, Donald D., and Richard H. Wilson. 1969. “The Effect of Spatially Separated Sound Sources on Speech Intelligibility.” *Journal of Speech and Hearing Research* 12 (1): 5–38.

Drennan, W. R., S. Gatehouse, and C. Lever. 2003. “Perceptual Segregation of Competing Speech Sounds: The Role of Spatial Location.” *The Journal of the Acoustical Society of America* 114: 2178.

Drennan, Ward R., Stuart Gatehouse, Patrick Howell, Dianne Van Tasell, and Steven Lund. 2005. “Localization and Speech-Identification Ability of Hearing-Impaired Listeners Using Phase-Preserving Amplification.” *Ear and Hearing* 26 (5): 461–72.

Dubno, Judy R., Donald D. Dirks, and Donald E. Morgan. 1984. “Effects of Age and Mild Hearing Loss on Speech Recognition in Noise.” *The Journal of the Acoustical Society of America* 76 (1): 87–96. doi:10.1121/1.391011.

Durlach, N. I. 1963. “Equalization and Cancellation Theory of Binaural Masking-Level Differences.” *The Journal of the Acoustical Society of America* 35 (8): 1206–18. doi:10.1121/1.1918675.

Durlach, N. I., K. J. Gabriel, H. S. Colburn, and C. Trahiotis. 1986. “Interaural Correlation Discrimination: II. Relation to Binaural Unmasking.” *The Journal of the Acoustical Society of America* 79 (5): 1548–57. doi:10.1121/1.393681.

Durlach, N. I., C. L. Thompson, and H. S. Colburn. 1981. “Binaural Interaction in Impaired Listeners: A Review of Past Research.” *Audiology* 20 (3): 181–211. doi:10.3109/00206098109072694.

Durlach, Nathaniel I. 1972. “Binaural Signal Detection-Equalization and Cancellation Theory.” http://ntrs.nasa.gov/search.jsp?R=19730045482.

Ebert, Charles S, Deidra A Blanks, Mihir R Patel, Charles S Coffey, Allen F Marshall, and Douglas C Fitzpatrick. 2008. “Behavioral Sensitivity to Interaural Time Differences in the Rabbit.” *Hearing Research* 235 (1–2): 134–42. doi:10.1016/j.heares.2007.11.003.

Edmonds, Barrie A., and John F. Culling. 2005a. “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.

———. 2005b. “The Spatial Unmasking of Speech: Evidence for within-Channel Processing of Interaural Time Delay.” *The Journal of the Acoustical Society of America* 117 (5): 3069. doi:10.1121/1.1880752.

Ehret, Günter, and Michael M. Merzenich. 1988. “Complex Sound Analysis (Frequency Resolution, Filtering and Spectral Integration) by Single Units of the Inferior Colliculus of the Cat.” *Brain Research Reviews* 13 (2): 139–63. doi:10.1016/0165-0173(88)90018-5.

Engineer, Crystal T., Claudia A. Perez, YeTing H. Chen, Ryan S. Carraway, Amanda C. Reed, Jai A. Shetake, Vikram Jakkamsetti, Kevin Q. Chang, and Michael P. Kilgard. 2008. “Cortical Activity Patterns Predict Speech Discrimination Ability.” *Nature Neuroscience* 11 (5): 603–8. doi:10.1038/nn.2109.

Fitzpatrick, D C, R Batra, T R Stanford, and S Kuwada. 1997. “A Neuronal Population Code for Sound Localization.” *Nature* 388 (6645): 871–74. doi:10.1038/42246.

Fitzpatrick, D C, S Kuwada, and R Batra. 2000. “Neural Sensitivity to Interaural Time Differences: Beyond the Jeffress Model.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 20 (4): 1605–15.

Fritz, Jonathan B., Mounya Elhilali, Stephen V. David, and Shihab A. Shamma. 2007. “Auditory Attention—focusing the Searchlight on Sound.” *Current Opinion in Neurobiology* 17 (4): 437–455.

Fritz, Jonathan B., Mounya Elhilali, and Shihab A. Shamma. 2007. “Adaptive Changes in Cortical Receptive Fields Induced by Attention to Complex Sounds.” *Journal of Neurophysiology* 98 (4): 2337–2346.

Furman, Adam C., Sharon G. Kujawa, and M. Charles Liberman. 2013. “Noise-Induced Cochlear Neuropathy Is Selective for Fibers with Low Spontaneous Rates.” *Journal of Neurophysiology* 110 (3): 577–86. doi:10.1152/jn.00164.2013.

Furukawa, Shigeto, and John C Middlebrooks. 2002. “Cortical Representation of Auditory Space: Information-Bearing Features of Spike Patterns.” *Journal of Neurophysiology* 87 (4): 1749–62. doi:10.1152/jn.00491.2001.

Garcia-Lazaro, Jose A., Lucile A. C. Belliveau, and Nicholas A. Lesica. 2013a. “Independent Population Coding of Speech with Sub-Millisecond Precision.” *The Journal of Neuroscience* 33 (49): 19362–72. doi:10.1523/JNEUROSCI.3711-13.2013.

Garcia-Lazaro, Jose A, Lucile A C Belliveau, and Nicholas A Lesica. 2013b. “Independent Population Coding of Speech with Sub-Millisecond Precision.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (49): 19362–72. doi:10.1523/JNEUROSCI.3711-13.2013.

Glasberg, Brian R, and Brian C.J Moore. 1990. “Derivation of Auditory Filter Shapes from Notched-Noise Data.” *Hearing Research* 47 (1–2): 103–38. doi:10.1016/0378-5955(90)90170-T.

Goldberg, David H, Jonathan D Victor, Esther P Gardner, and Daniel Gardner. 2009. “Spike Train Analysis Toolkit: Enabling Wider Application of Information-Theoretic

Techniques to Neurophysiology." *Neuroinformatics* 7 (3): 165–78. doi:10.1007/s12021-009-9049-y.

Goldberg, J. M., and P. B. Brown. 1969. "Response of Binaural Neurons of Dog Superior Olivary Complex to Dichotic Tonal Stimuli: Some Physiological Mechanisms of Sound Localization." *Journal of Neurophysiology* 32 (4): 613–36.

Goodman, Dan FM, Victor Benichoux, and Romain Brette. 2013. "Decoding Neural Responses to Temporal Cues for Sound Localization." *eLife* 2 (January): e01312. doi:10.7554/eLife.01312.

Griffin, Sarah J., Leslie R. Bernstein, Neil J. Ingham, and David McAlpine. 2005. "Neural Sensitivity to Interaural Envelope Delays in the Inferior Colliculus of the Guinea Pig." *Journal of Neurophysiology* 93 (6): 3463–78. doi:10.1152/jn.00794.2004.

Groh, J M, A S Trause, A M Underhill, K R Clark, and S Inati. 2001. "Eye Position Influences Auditory Responses in Primate Inferior Colliculus." *Neuron* 29 (2): 509–18.

Groh, Jennifer M, Kristin A Kelly, and Abigail M Underhill. 2003. "A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus." *Journal of Cognitive Neuroscience* 15 (8): 1217–31. doi:10.1162/089892903322598166.

Grothe, Benedikt, Michael Pecka, and D McAlpine. 2010. "Mechanisms of Sound Localization in Mammals." *Physiological Reviews*, 983–1012. doi:10.1152/physrev.00026.2009.

Hall, Joseph W., Richard S. Tyler, and Mariano A. Fernandes. 1983. "Monaural and Binaural Auditory Frequency Resolution Measured Using Bandlimited Noise and Notched-noise Masking." *The Journal of the Acoustical Society of America* 73 (3): 894–98. doi:10.1121/1.389013.

Happel, Max F K, Marcus Jeschke, and Frank W Ohl. 2010. "Spectral Integration in Primary Auditory Cortex Attributable to Temporally Precise Convergence of Thalamocortical and Intracortical Input." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 30 (33): 11114–27. doi:10.1523/JNEUROSCI.0689-10.2010.

Harkins, Judith, and Paula Tucker. 2007. "An Internet Survey of Individuals With Hearing Loss Regarding Assistive Listening Devices." *Trends in Amplification* 11 (2): 91–100. doi:10.1177/1084713807301322.

Harris, David M., Robert V. Shannon, Russell Snyder, and Edward Carney. 1997. "Multi-Unit Mapping of Acoustic Stimuli in Gerbil Inferior Colliculus." *Hearing Research* 108 (1–2): 145–56. doi:10.1016/S0378-5955(97)00047-6.

Hartley, David, Elena Rochtchina, Philip Newall, Maryanne Golding, and Paul Mitchell. 2010. "Use of Hearing Aids and Assistive Listening Devices in an Older Australian Population." *Journal of the American Academy of Audiology* 21 (10): 642–53. doi:10.3766/jaaa.21.10.4.

Häusler, R., S. Colburn, and E. Marr. 1983. "Sound Localization in Subjects with Impaired Hearing." *Acta Oto-Laryngologica* 96 (sup400): 1–62. doi:10.3109/00016488309105590.

Hawkins, David B., and William S. Yacullo. 1984. "Signal-to-Noise Ratio Advantage of Binaural Hearing Aids and Directional Microphones under Different Levels of Reverberation." *Journal of Speech and Hearing Disorders* 49 (3): 278–86. doi:10.1044/jshd.4903.278.

Hazan, Lynn, Michaël Zugaro, and György Buzsáki. 2006a. "Klusters, NeuroScope, NDManager: A Free Software Suite for Neurophysiological Data Processing and Visualization." *Journal of Neuroscience Methods* 155 (2): 207–216.

———. 2006b. “Klusters, NeuroScope, NDManager: A Free Software Suite for Neurophysiological Data Processing and Visualization.” *Journal of Neuroscience Methods* 155 (2): 207–16. doi:10.1016/j.jneumeth.2006.01.017.

Hebrank, Jack, and D. Wright. 1974. “Spectral Cues Used in the Localization of Sound Sources on the Median Plane.” *The Journal of the Acoustical Society of America* 56 (6): 1829–1834.

Heffner, Henry E., Rickye S. Heffner, Christopher Contos, and Tonie Ott. 1994. “Audiogram of the Hooded Norway Rat.” *Hearing Research* 73 (2): 244–247.

Heffner, R S, and H E Heffner. 1988. “Sound Localization and Use of Binaural Cues by the Gerbil (Meriones Unguiculatus).” *Behavioral Neuroscience* 102 (3): 422–28.

Heffner, Rickye S., and Henry E. Heffner. 1985. “Hearing Range of the Domestic Cat.” *Hearing Research* 19 (1): 85–88. doi:10.1016/0378-5955(85)90100-5.

Heijden, Marcel van der, and Constantine Trahiotis. 1998. “Binaural Detection as a Function of Interaural Correlation and Bandwidth of Masking Noise: Implications for Estimates of Spectral Resolution.” *The Journal of the Acoustical Society of America* 103 (3): 1609–14. doi:10.1121/1.421295.

———. 1999. “Masking with Interaurally Delayed Stimuli: The Use of ‘internal’ Delays in Binaural Detection.” *The Journal of the Acoustical Society of America* 105 (1): 388–99. doi:10.1121/1.424628.

Hienz, Robert D., Cynthia M. Aleszczyk, and Bradford J. May. 1996. “Vowel Discrimination in Cats: Acquisition, Effects of Stimulus Level, and Performance in Noise.” *The Journal of the Acoustical Society of America* 99 (6): 3656–68. doi:10.1121/1.414980.

Hienz, Robert D, Phaedra Stiles, and Bradford J May. 1998. “Effects of Bilateral Olivocochlear Lesions on Vowel Formant Discrimination in Cats.” *Hearing Research* 116 (1–2): 10–20. doi:10.1016/S0378-5955(97)00197-4.

Hienz, RobertD., Cynthia M. Aleszczyk, and Bradford J. May. 1996. “Vowel Discrimination in Cats: Thresholds for the Detection of Second Formant Changes in the Vowel /ɛ/.” *The Journal of the Acoustical Society of America* 100 (2): 1052–58. doi:10.1121/1.416291.

Hirsh. 1948a. “The Influence of Interaural Phase on Interaural Summation and Inhibition.” *The Journal of the Acoustical Society of America*, a, 20 (4): 536–44. doi:10.1121/1.1906407.

———. 1948b. “Binaural Summation and Interaural Inhibition as a Function of the Level of Masking Noise.” *The American Journal of Psychology*, b, 61 (2): 205–13. doi:10.2307/1416966.

Hocherman, S., D. A. Benson, M. H. Goldstein Jr., H. E. Heffner, and R. D. Hienz. 1976. “Evoked Unit Activity in Auditory Cortex of Monkeys Performing a Selective Attention Task.” *Brain Research* 117 (1): 51–68. doi:10.1016/0006-8993(76)90555-2.

Hoesel, Richard J. M. van, and Richard S. Tyler. 2003. “Speech Perception, Localization, and Lateralization with Bilateral Cochlear Implants.” *The Journal of the Acoustical Society of America* 113 (3): 1617–30. doi:10.1121/1.1539520.

Holube, Inga, Martin Kinkel, and Birger Kollmeier. 1998. “Binaural and Monaural Auditory Filter Bandwidths and Time Constants in Probe Tone Detection Experiments.” *The Journal of the Acoustical Society of America* 104 (4): 2412–25. doi:10.1121/1.423773.

Hubel, D. H., C. O. Henson, A. Rupert, and R. Galambos. 1959. “Attention Units in the Auditory Cortex.” *Science (New York, N.Y.)* 129 (3358): 1279–80.

Hukin, R. W., and C. J. Darwin. 1995. "Effects of Contralateral Presentation and of Interaural Time Differences in Segregating a Harmonic from a Vowel." *The Journal of the Acoustical Society of America* 98 (3): 1380–87. doi:10.1121/1.414348.

Jacobs, Adam L, Gene Fridman, Robert M Douglas, Nazia M Alam, Peter E Latham, Glen T Prusky, and Sheila Nirenberg. 2009. "Ruling out and Ruling in Neural Codes." *Proceedings of the National Academy of Sciences of the United States of America* 106 (14): 5936–41. doi:10.1073/pnas.0900573106.

Jeffress, Lloyd A. 1948. "A Place Theory of Sound Localization." *Journal of Comparative and Physiological Psychology* 41 (1): 35.

Jenkins, W M, and R B Masterton. 1982. "Sound Localization: Effects of Unilateral Lesions in Central Auditory System." *Journal of Neurophysiology* 47 (6): 987–1016.

Jenkins, W M, and M M Merzenich. 1984. "Role of Cat Primary Auditory Cortex for Sound-Localization Behavior." *Journal of Neurophysiology* 52 (5): 819–47.

Jenkins, William M., and R. Bruce. 1982. "Sound Localization: Effects of Unilateral Lesions in Central Auditory System." *Journal of Neurophysiology* 47 (6): 987–1016.

Jiang, Dan, David McAlpine, and Alan R. Palmer. 1997. "Responses of Neurons in the Inferior Colliculus to Binaural Masking Level Difference Stimuli Measured by Rate-versus-Level Functions." *Journal of Neurophysiology* 77 (6): 3085–3106.

Joris, P. X., and T. C. Yin. 1995. "Envelope Coding in the Lateral Superior Olive. I. Sensitivity to Interaural Time Differences." *Journal of Neurophysiology* 73 (3): 1043–62.

Joris, Philip X. 2003. "Interaural Time Sensitivity Dominated by Cochlea-Induced Envelope Patterns." *The Journal of Neuroscience* 23 (15): 6345–50.

Joris, Philip X., Marcel van der Heijden, Dries H. Louage, Bram Van de Sande, and Cindy Van Kerckhoven. 2005. "Dependence of Binaural and Cochlear 'best Delays' on Characteristic Frequency." In *Auditory Signal Processing*, 477–483. Springer. http://link.springer.com/content/pdf/10.1007/0-387-27045-0_58.pdf.

Joris, Philip, and Tom C.T. Yin. 2007. "A Matter of Time: Internal Delays in Binaural Processing." *Trends in Neurosciences* 30 (2): 70–78. doi:10.1016/j.tins.2006.12.004.

Kavanagh, GL, and JB Kelly. 1987. "Contribution of Auditory Cortex to Sound Localization by the Ferret (Mustela Putorius)." *J Neurophysiol* 57 (6).

Keidser, Gitte, Anna O'Brien, Jens-Uwe Hain, Margot McLelland, and Ingrid Yeend. 2009. "The Effect of Frequency-Dependent Microphone Directionality on Horizontal Localization Performance in Hearing-Aid Users." *International Journal of Audiology* 48 (11): 789–803. doi:10.3109/14992020903036357.

Keidser, Gitte, Kristin Rohrseitz, Harvey Dillon, Volkmar Hamacher, Lyndal Carter, Uwe Rass, and Elizabeth Convery. 2006. "The Effect of Multi-Channel Wide Dynamic Range Compression, Noise Reduction, and the Directional Microphone on Horizontal Localization Performance in Hearing Aid Wearers." *International Journal of Audiology* 45 (10): 563–79. doi:10.1080/14992020600920804.

Kelly, J B, and G L Kavanagh. 1986. "Effects of Auditory Cortical Lesions on Pure-Tone Sound Localization by the Albino Rat." *Behavioral Neuroscience* 100 (4): 569–75.

Kelly, Jack B., Gerard L. Kavanagh, and James C. H. Dalton. 1986. "Hearing in the Ferret (Mustela Putorius): Thresholds for Pure Tone Detection." *Hearing Research* 24 (3): 269–75. doi:10.1016/0378-5955(86)90025-0.

Kidd, G., C. Mason, V. Richards, F. Gallun, and N. Durlach. 2007. "Informational Masking." *Auditory Perception of Sound Sources*, 143–189.

Kidd, Gerald, Christine R. Mason, Andrew Brughera, and William M. Hartmann. 2005. "The Role of Reverberation in Release from Masking Due to Spatial Separation of

Sources for Speech Identification.” *Acta Acustica United with Acustica* 91 (3): 526–36.

Kidd, Gerald, Christine R. Mason, Tanya L. Rohtla, and Phalguni S. Deliwala. 1998. “Release from Masking due to Spatial Separation of Sources in the Identification of Nonspeech Auditory Patterns.” *The Journal of the Acoustical Society of America* 104 (1): 422–31. doi:10.1121/1.423246.

King, A. J., and J. C. Middlebrooks. 2011. “Cortical Representation of Auditory Space.” In *The Auditory Cortex*, edited by J. A. Winer and C. E. Schreiner, 329–40. New York, New York, USA: Springer.

King, Andrew J, and Jan W H Schnupp. 2011. *The Auditory Cortex*. Edited by Jeffery A Winer and Christoph E Schreiner. *Current Biology : CB*. Vol. 17. New York: Springer. doi:10.1016/j.cub.2007.01.046.

King, Andrew J., Jan WH Schnupp, and Timothy P. Doubell. 2001. “The Shape of Ears to Come: Dynamic Coding of Auditory Space.” *Trends in Cognitive Sciences* 5 (6): 261–270.

Kittel, Malte, Eva Wagner, and Georg M Klump. 2002. “An Estimate of the Auditory-Filter Bandwidth in the Mongolian Gerbil.” *Hearing Research* 164 (1–2): 69–76. doi:10.1016/S0378-5955(01)00411-7.

Klasen, Thomas J., Tim Van den Bogaert, Marc Moonen, and Jan Wouters. 2007. “Binaural Noise Reduction Algorithms for Hearing Aids That Preserve Interaural Time Delay Cues.” *IEEE Transactions on Signal Processing* 55 (4): 1579–1585.

Klatt, Dennis H. 1980. “Software for a Cascade/Parallel Formant Synthesizer.” *The Journal of the Acoustical Society of America* 67 (3): 971–95. doi:10.1121/1.383940.

Klumpp, R. G., and H. R. Eady. 1956. “Some Measurements of Interaural Time Difference Thresholds.” *The Journal of the Acoustical Society of America* 28 (5): 859–860.

Kuwada, SHIGEYUKI, TERRENCE R. Stanford, and RANJAN Batra. 1987. “Interaural Phase-Sensitive Units in the Inferior Colliculus of the Unanesthetized Rabbit: Effects of Changing Frequency.” *Journal of Neurophysiology* 57 (5): 1338–1360.

Langford, Ted L., and Lloyd A. Jeffress. 1964. “Effect of Noise Crosscorrelation on Binaural Signal Detection.” *The Journal of the Acoustical Society of America* 36 (8): 1455–58. doi:10.1121/1.1919224.

Larson, V. D., D. W. Williams, W. G. Henderson, L. E. Luethke, L. B. Beck, D. Noffsinger, R. H. Wilson, et al. 2000. “Efficacy of 3 Commonly Used Hearing Aid Circuits: A Crossover Trial. NIDCD/VA Hearing Aid Clinical Trial Group.” *JAMA* 284 (14): 1806–13.

Lee, Chen-Chung, and John C Middlebrooks. 2011a. “Auditory Cortex Spatial Sensitivity Sharpens during Task Performance.” *Nature Neuroscience* 14 (1): 108–14. doi:10.1038/nn.2713.

———. 2011b. “Auditory Cortex Spatial Sensitivity Sharpens during Task Performance.” *Nature Neuroscience* 14 (1). Nature Publishing Group: 108–14. doi:10.1038/nn.2713.

Lee, Chen-Chung, and John C. Middlebrooks. 2013. “Specialization for Sound Localization in Fields A1, DZ, and PAF of Cat Auditory Cortex.” *Journal of the Association for Research in Otolaryngology* 14 (1): 61–82. doi:10.1007/s10162-012-0357-9.

Lesica, N. A., A. Lingner, and B. Grothe. 2010. “Population Coding of Interaural Time Differences in Gerbils and Barn Owls.” *Journal of Neuroscience* 30 (35): 11696–702. doi:10.1523/JNEUROSCI.0846-10.2010.

Lesica, Nicholas A, Andrea Lingner, and Benedikt Grothe. 2010. “Population Coding of Interaural Time Differences in Gerbils and Barn Owls.” *The Journal of*

Neuroscience : The Official Journal of the Society for Neuroscience 30 (35): 11696–702. doi:10.1523/JNEUROSCI.0846-10.2010.

Levitt, H. 1971. “Transformed Up-Down Methods in Psychoacoustics.” *The Journal of the Acoustical Society of America* 49 (2B): 467–77. doi:10.1121/1.1912375.

Levitt, H., and L. R. Rabiner. 1967. “Binaural Release From Masking for Speech and Gain in Intelligibility.” *The Journal of the Acoustical Society of America*, a, 42 (3): 601–8. doi:10.1121/1.1910629.

Li, Ling-yun, Ya-tang Li, Mu Zhou, Huizhong W Tao, and Li I Zhang. 2013. “Intracortical Multiplication of Thalamocortical Signals in Mouse Auditory Cortex.” *Nature Neuroscience* 16 (9): 1179–81. doi:10.1038/nn.3493.

Licklider, J. C. R. 1948. “The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise.” *The Journal of the Acoustical Society of America* 20 (2): 150–59. doi:10.1121/1.1906358.

Lin, Frank R., Kristine Yaffe, Jin Xia, Qian-Li Xue, Tamara B. Harris, Elizabeth Purchase-Helzner, Suzanne Satterfield, Hilsa N. Ayonayon, Luigi Ferrucci, and Eleanor M. Simonsick. 2013. “Hearing Loss and Cognitive Decline Among Older Adults.” *JAMA Internal Medicine* 173 (4). doi:10.1001/jamainternmed.2013.1868.

Lingner, Andrea, Lutz Wiegrebe, and Benedikt Grothe. 2012a. “Sound Localization in Noise by Gerbils and Humans.” *JARO - Journal of the Association for Research in Otolaryngology* 13 (2): 237–48. doi:10.1007/s10162-011-0301-4.

———. 2012b. “Sound Localization in Noise by Gerbils and Humans.” *Journal of the Association for Research in Otolaryngology : JARO* 13 (2): 237–48. doi:10.1007/s10162-011-0301-4.

Litovsky, Ruth Y., H. Steven Colburn, William A. Yost, and Sandra J. Guzman. 1999. “The Precedence Effect.” *The Journal of the Acoustical Society of America* 106 (4): 1633–54. doi:10.1121/1.427914.

Litovsky, Ruth Y., Gary L. Jones, Smita Agrawal, and Richard van Hoesel. 2010. “Effect of Age at Onset of Deafness on Binaural Sensitivity in Electric Hearing in Humans.” *The Journal of the Acoustical Society of America* 127 (1): 400–414. doi:10.1121/1.3257546.

Loizou, Philipos C., Yi Hu, Ruth Litovsky, Gongqiang Yu, Robert Peters, Jennifer Lake, and Peter Roland. 2009. “Speech Recognition by Bilateral Cochlear Implant Users in a Cocktail-Party Setting.” *The Journal of the Acoustical Society of America* 125 (1): 372–83. doi:10.1121/1.3036175.

Lorenzi, C., S. Gatehouse, and C. Lever. 1999b. “Sound Localization in Noise in Hearing-Impaired Listeners.” *The Journal of the Acoustical Society of America* 105 (6): 3454–3463.

Lorenzi, Christian, Stuart Gatehouse, and Catherine Lever. 1999a. “Sound Localization in Noise in Normal-Hearing Listeners.” *The Journal of the Acoustical Society of America* 105 (3): 1810–1820.

Lüling, Hannes, Ida Siveke, Benedikt Grothe, and Christian Leibold. 2011. “Frequency-Invariant Representation of Interaural Time Differences in Mammals.” *PLoS Computational Biology* 7 (3): e1002013. doi:10.1371/journal.pcbi.1002013.

Maddox, Ross K, Cyrus P Billimoria, Ben P Perrone, Barbara G Shinn-Cunningham, and Kamal Sen. 2012. “Competing Sound Sources Reveal Spatial Effects in Cortical Processing.” *PLoS Biology* 10 (5): e1001319. doi:10.1371/journal.pbio.1001319.

Magezi, David A, and Katrin Krumbholz. 2010. “Evidence for Opponent-Channel Coding of Interaural Time Differences in Human Auditory Cortex.” *Journal of Neurophysiology* 104 (4): 1997–2007. doi:10.1152/jn.00424.2009.

Maier, Julia K, Teresa Kindermann, Benedikt Grothe, and Georg M Klump. 2008. “Effects of Omni-Directional Noise-Exposure during Hearing Onset and Age on Auditory Spatial Resolution in the Mongolian Gerbil (Meriones Unguiculatus) -- a Behavioral Approach.” *Brain Research* 1220 (July): 47–57. doi:10.1016/j.brainres.2008.01.083.

Maier, Julia K., and Georg M. Klump. 2006. “Resolution in Azimuth Sound Localization in the Mongolian Gerbil (Meriones Unguiculatus).” *The Journal of the Acoustical Society of America* 119 (2): 1029. doi:10.1121/1.2159429.

Maki, K., and S. Furukawa. 2005. “The Responses of Neurons in the Gerbil Inferior Colliculus to Virtual Acoustic Space Stimuli.” *Acoustical Science and Technology* 26 (1): 82–84.

Maki, Katuhiro, and Shigeto Furukawa. 2005. “Acoustical Cues for Sound Localization by the Mongolian Gerbil, Meriones Unguiculatus.” *The Journal of the Acoustical Society of America* 118 (2): 872. doi:10.1121/1.1944647.

Malhotra, Shveta, Amee J Hall, and Stephen G Lomber. 2004a. “Cortical Control of Sound Localization in the Cat: Unilateral Cooling Deactivation of 19 Cerebral Areas.” *Journal of Neurophysiology* 92 (3): 1625–43. doi:10.1152/jn.01205.2003.

Malhotra, Shveta, Amee J. Hall, and Stephen G. Lomber. 2004b. “Cortical Control of Sound Localization in the Cat: Unilateral Cooling Deactivation of 19 Cerebral Areas.” *Journal of Neurophysiology* 92 (3): 1625–43. doi:10.1152/jn.01205.2003.

Malone, Brian J., Brian H. Scott, and Malcolm N. Semple. 2002. “Context-Dependent Adaptive Coding of Interaural Phase Disparity in the Auditory Cortex of Awake Macaques.” *Journal of Neuroscience* 22 (11): 4625–38.

Marquardt, Torsten, and David McAlpine. 2009. “Masking with Interaurally ‘double-Delayed’ Stimuli: The Range of Internal Delays in the Human Brain.” *The Journal of the Acoustical Society of America* 126 (6): EL177-EL182. doi:10.1121/1.3253689.

May, Bradford J., Glenn S. Le Prell, and Murray B. Sachs. 1998. “Vowel Representations in the Ventral Cochlear Nucleus of the Cat: Effects of Level, Background Noise, and Behavioral State.” *Journal of Neurophysiology* 79 (4): 1755–67.

McAlpine, D, D Jiang, and A R Palmer. 2001. “A Neural Code for Low-Frequency Sound Localization in Mammals.” *Nature Neuroscience* 4 (4): 396–401. doi:10.1038/86049.

McAlpine, David, Dan Jiang, and Alan R. Palmer. 1996. “Binaural Masking Level Differences in the Inferior Colliculus of the Guinea Pig.” *The Journal of the Acoustical Society of America*, b, 100 (1): 490–503. doi:10.1121/1.415862.

———. 2001. “A Neural Code for Low-Frequency Sound Localization in Mammals.” *Nature Neuroscience* 4 (4): 396–401. doi:10.1038/86049.

McCormack, Abby, and Heather Fortnum. 2013. “Why Do People Fitted with Hearing Aids Not Wear Them?” *International Journal of Audiology* 52 (5): 360–68. doi:10.3109/14992027.2013.769066.

McFadden, Dennis, and Edward G. Pasanen. 1976. “Lateralization at High Frequencies Based on Interaural Time Differences.” *The Journal of the Acoustical Society of America* 59 (3): 634–39. doi:10.1121/1.380913.

Metzger, Ryan R., Nathaniel T. Greene, Kristin Kelly Porter, and Jennifer M. Groh. 2006. “Effects of Reward and Behavioral Context on Neural Activity in the Primate Inferior Colliculus.” *The Journal of Neuroscience* 26 (28): 7468–76. doi:10.1523/JNEUROSCI.5401-05.2006.

Mick, Paul, Ichiro Kawachi, and Frank R. Lin. 2014. “The Association between Hearing Loss and Social Isolation in Older Adults.” *Otolaryngology -- Head and Neck Surgery* 150 (3): 378–84. doi:10.1177/0194599813518021.

Mickey, Brian J., and John C. Middlebrooks. 2003. “Representation of Auditory Space by Cortical Neurons in Awake Cats.” *The Journal of Neuroscience* 23 (25): 8649–63.

Middlebrooks, John C., and Peter Bremen. 2013. “Spatial Stream Segregation by Auditory Cortical Neurons.” *The Journal of Neuroscience* 33 (27): 10986–1. doi:10.1523/JNEUROSCI.1065-13.2013.

Middlebrooks, John C, and Peter Bremen. 2013. “Spatial Stream Segregation by Auditory Cortical Neurons.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 33 (27): 10986–1. doi:10.1523/JNEUROSCI.1065-13.2013.

Middlebrooks, JOHN C., and JOHN D. Pettigrew. 1981. “Functional Classes of Neurons in Primary Auditory Cortex of the Cat Distinguished by Sensitivity to Sound Location.” *J Neurosci* 1 (1): 107220.

Miller, J. M., D. Sutton, B. Pfingst, A. Ryan, R. Beaton, and G. Gourevitch. 1972. “Single Cell Activity in the Auditory Cortex of Rhesus Monkeys: Behavioral Dependency.” *Science* 177 (4047): 449–51. doi:10.1126/science.177.4047.449.

Miller, Lee M, and Gregg H Recanzone. 2009. “Populations of Auditory Cortical Neurons Can Accurately Encode Acoustic Space across Stimulus Intensity.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (14): 5931–35. doi:10.1073/pnas.0901023106.

Moore, Brian CJ. 1996. “Perceptual Consequences of Cochlear Hearing Loss and Their Implications for the Design of Hearing Aids.” *Ear and Hearing* 17 (2): 133–161.

Mossop, J E, and J F Culling. 1998. “Lateralization of Large Interaural Delays.” *The Journal of the Acoustical Society of America* 104 (3 Pt 1): 1574–79.

Mulrow, Cynthia D. 1990. “Quality-of-Life Changes and Hearing Impairment: A Randomized Trial.” *Annals of Internal Medicine* 113 (3): 188. doi:10.7326/0003-4819-113-3-188.

Neher, Tobias, Søren Laugesen, Niels Søgaard Jensen, and Louise Kragelund. 2011. “Can Basic Auditory and Cognitive Measures Predict Hearing-Impaired Listeners’ Localization and Spatial Speech Recognition Abilities? A.” *The Journal of the Acoustical Society of America* 130 (3): 1542–1558.

Noble, William, Denis Byrne, and Bernadette Lepage. 1994. “Effects on Sound Localization of Configuration and Type of Hearing Impairment.” *The Journal of the Acoustical Society of America* 95 (2): 992–1005.

Nodal, F. R., V. M. Bajo, C. H. Parsons, J. W. Schnupp, and A. J. King. 2008. “Sound Localization Behavior in Ferrets: Comparison of Acoustic Orientation and Approach-to-Target Responses.” *Neuroscience*, From Cochlea to Cortex: Recent Advances in Auditory Neuroscience, 154 (1): 397–408. doi:10.1016/j.neuroscience.2007.12.022.

Oliver, Douglas L. 2000. “Ascending Efferent Projections of the Superior Olivary Complex.” *Microscopy Research and Technique* 51 (4): 355–363. doi:10.1002/1097-0029(20001115)51:4<355::AID-JEMT5>3.0.CO;2-J.

Oliver, Douglas L., Gretchen E. Beckius, and Amiram Shneiderman. 1995. “Axonal Projections from the Lateral and Medial Superior Olive to the Inferior Colliculus of the Cat: A Study Using Electron Microscopic Autoradiography.” *The Journal of Comparative Neurology* 360 (1): 17–32. doi:10.1002/cne.903600103.

Orton, Llwyd David, and Adrian Rees. 2014. “Intercollicular Commissural Connections Refine the Representation of Sound Frequency and Level in the Auditory Midbrain.” *eLife* 3 (November): e03764. doi:10.7554/eLife.03764.

Palmer, Alan R., Dan Jiang, and David McAlpine. 1999. “Desynchronizing Responses to Correlated Noise: A Mechanism for Binaural Masking Level Differences at the Inferior Colliculus.” *Journal of Neurophysiology* 81 (2): 722–734.

———. 2000. “Neural Responses in the Inferior Colliculus to Binaural Masking Level Differences Created by Inverting the Noise in One Ear.” *Journal of Neurophysiology* 84 (2): 844–52.

Parsons, Carl H., Richard G. Lanyon, Jan W. H. Schnupp, and Andrew J. King. 1999. “Effects of Altering Spectral Cues in Infancy on Horizontal and Vertical Sound Localization by Adult Ferrets.” *Journal of Neurophysiology* 82 (5): 2294–2309.

Patterson, Roy D. 1976. “Auditory Filter Shapes Derived with Noise Stimuli.” *The Journal of the Acoustical Society of America* 59 (3): 640–54. doi:10.1121/1.380914.

Pecka, Michael, Antje Brand, Oliver Behrend, and Benedikt Grothe. 2008. “Interaural Time Difference Processing in the Mammalian Medial Superior Olive: The Role of Glycinergic Inhibition.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (27): 6914–25. doi:10.1523/JNEUROSCI.1660-08.2008.

Perez, Claudia A., Crystal T. Engineer, Vikram Jakkamsetti, Ryan S. Carraway, Matthew S. Perry, and Michael P. Kilgard. 2012. “Different Timescales for the Neural Coding of Consonant and Vowel Sounds.” *Cerebral Cortex*, March. doi:10.1093/cercor/bhs045.

———. 2013. “Different Timescales for the Neural Coding of Consonant and Vowel Sounds.” *Cerebral Cortex* 23 (3): 670–83. doi:10.1093/cercor/bhs045.

Peterson, Gordon E., and Harold L. Barney. 1952. “Control Methods Used in a Study of the Vowels.” *The Journal of the Acoustical Society of America* 24 (2): 175–84. doi:10.1121/1.1906875.

Pfingst, B. E., and T. A. O’Connor. 1981. “Characteristics of Neurons in Auditory Cortex of Monkeys Performing a Simple Auditory Task.” *Journal of Neurophysiology* 45 (1): 16–34.

Pfingst, B. E., T. A. O’Connor, and J. M. Miller. 1977. “Response Plasticity of Neurons in Auditory Cortex of the Rhesus Monkey.” *Experimental Brain Research* 29 (3–4): 393–404. doi:10.1007/BF00236178.

Picou, Erin M., Elizabeth Aspell, and Todd A. Ricketts. 2014. “Potential Benefits and Limitations of Three Types of Directional Processing in Hearing Aids.” *Ear and Hearing* 35 (3): 339–52. doi:10.1097/AUD.0000000000000004.

Plomp, R., and H. J. M. Steeneken. 1969. “Effect of Phase on the Timbre of Complex Tones.” *The Journal of the Acoustical Society of America* 46 (2B): 409–21. doi:10.1121/1.1911705.

Pollack, Irwin, and W. J. Trittipoe. 1959a. “Binaural Listening and Interaural Noise Cross Correlation.” *The Journal of the Acoustical Society of America*, a, 31 (9): 1250–52. doi:10.1121/1.1907852.

Pollack, Irwin, and William Trittipoe. 1959b. “Interaural Noise Correlations: Examination of Variables.” *The Journal of the Acoustical Society of America*, b, 31 (12): 1616–18. doi:10.1121/1.1907669.

Quaranta, Nicola, Francesco Coppola, Mara Casulli, Maria Rosaria Barulli, Francesco Panza, Rosanna Tortelli, Vincenzo Solfrizzi, Carlo Sabbà, and Giancarlo Logroscino. 2015. “Epidemiology of Age Related Hearing Loss: A Review.” *Hearing, Balance and Communication* 13 (2): 77–81. doi:10.3109/21695717.2014.994869.

Rakerd, Brad, and W. M. Hartmann. 1986. “Localization of Sound in Rooms, III: Onset and Duration Effects.” *The Journal of the Acoustical Society of America* 80 (6): 1695–1706. doi:10.1121/1.394282.

Reale, RA, and JF Brugge. 1990. “Auditory Cortical Neurons Are Sensitive to Static and Continuously Changing Interaural Phase Cues.” *J Neurophysiol* 103: 313–38.

Rébillat, Marc, Victor Benichoux, Makoto Otani, Renaud Keriven, and Romain Brette. 2014. “Estimation of the Low-Frequency Components of the Head-Related Transfer Functions of Animals from Photographs.” *The Journal of the Acoustical Society of America* 135 (5).

Recanzone, Gregg H., Darren C. Guard, and Mimi L. Phan. 2000. “Frequency and Intensity Response Properties of Single Neurons in the Auditory Cortex of the Behaving Macaque Monkey.” *Journal of Neurophysiology* 83 (4): 2315–31.

Rose, J. E., N. B. Gross, C. D. Geisler, and J. E. Hind. 1966. “Some Neural Mechanisms in the Inferior Colliculus of the Cat Which May Be Relevant to Localization of a Sound Source.” *Journal of Neurophysiology* 29 (2): 288–314.

Ryan, A., and J. Miller. 1978. “Single Unit Responses in the Inferior Colliculus of the Awake and Performing Rhesus Monkey.” *Experimental Brain Research* 32 (3): 389–407.

Ryan, Allen. 1976. “Hearing Sensitivity of the Mongolian Gerbil, Merionesunguiculatis.” *The Journal of the Acoustical Society of America* 59 (5): 1222–26. doi:10.1121/1.380961.

Ryan, ALLEN F., JOSEF M. Miller, BRYAN E. Pfingst, and G. K. Martin. 1984. “Effects of Reaction Time Performance on Single-Unit Activity in the Central Auditory Pathway of the Rhesus Macaque.” *J Neurosci* 4: 298–308.

Sachs, Murray B., Ian C. Bruce, Roger L. Miller, and Eric D. Young. 2002. “Biological Basis of Hearing-Aid Design.” *Annals of Biomedical Engineering* 30 (2): 157–68. doi:10.1114/1.1458592.

Salminen, Nelli H, Hannu Tiitinen, Santeri Yrttiaho, and Patrick J C May. 2010. “The Neural Code for Interaural Time Difference in Human Auditory Cortex.” *The Journal of the Acoustical Society of America* 127 (2): EL60-5. doi:10.1121/1.3290744.

Schebesch, Gabriele, Andrea Lingner, Uwe Firzlaff, Lutz Wiegrebe, and Benedikt Grothe. 2010. “Perception and Neural Representation of Size-Variant Human Vowels in the Mongolian Gerbil (Meriones Unguiculatus).” *Hearing Research* 261 (1–2): 1–8. doi:10.1016/j.heares.2009.12.016.

Schilling, John R., Roger L. Miller, Murray B. Sachs, and Eric D. Young. 1998. “Frequency-Shaped Amplification Changes the Neural Representation of Speech with Noise-Induced Hearing Loss.” *Hearing Research* 117 (1–2): 57–70. doi:10.1016/S0378-5955(98)00003-3.

Schmitzer-Torbert, N., J. Jackson, D. Henze, K. Harris, and A. D. Redish. 2005a. “Quantitative Measures of Cluster Quality for Use in Extracellular Recordings.” *Neuroscience* 131 (1): 1–11.

Schmitzer-Torbert, N, J Jackson, D Henze, K Harris, and A D Redish. 2005b. “Quantitative Measures of Cluster Quality for Use in Extracellular Recordings.” *Neuroscience* 131 (1): 1–11. doi:10.1016/j.neuroscience.2004.09.066.

Scott, Brian H., Brian J. Malone, and Malcolm N. Semple. 2007. “Effect of Behavioral Context on Representation of a Spatial Cue in Core Auditory Cortex of Awake Macaques.” *Journal of Neuroscience* 27 (24): 6489–99. doi:10.1523/JNEUROSCI.0016-07.2007.

Scott, Brian H, Brian J Malone, and Malcolm N Semple. 2009. “Representation of Dynamic Interaural Phase Difference in Auditory Cortex of Awake Rhesus Macaques.” *Journal of Neurophysiology* 101 (4): 1781–99. doi:10.1152/jn.00678.2007.

Shackleton, Trevor M., Robert H. Arnott, and Alan R. Palmer. 2005. “Sensitivity to Interaural Correlation of Single Neurons in the Inferior Colliculusof Guinea Pigs.”

Journal of the Association for Research in Otolaryngology 6 (3): 244–59. doi:10.1007/s10162-005-0005-8.

Shackleton, Trevor M., Ray Meddis, and Michael J. Hewitt. 1992. “Across Frequency Integration in a Model of Lateralization.” *The Journal of the Acoustical Society of America* 91 (4): 2276–79. doi:10.1121/1.403663.

Shamma, Shihab A., Naiming Shen, and Preetham Gopalaswamy. 1989. “Stereausis: Binaural Processing without Neural Delays.” *The Journal of the Acoustical Society of America* 86 (3): 989–1006. doi:10.1121/1.398734.

Shera, Christopher A., John J. Guinan, and Andrew J. Oxenham. 2002. “Revised Estimates of Human Cochlear Tuning from Otoacoustic and Behavioral Measurements.” *Proceedings of the National Academy of Sciences* 99 (5): 3318–23. doi:10.1073/pnas.032675099.

Shore, S. E., C. J. Sumner, S. C. Bledsoe, and J. Lu. 2003. “Effects of Contralateral Sound Stimulation on Unit Activity of Ventral Cochlear Nucleus Neurons.” *Experimental Brain Research* 153 (4): 427–435.

Simpson, Andrea. 2009. “Frequency-Lowering Devices for Managing High-Frequency Hearing Loss: A Review.” *Trends in Amplification* 13 (2): 87–106. doi:10.1177/1084713809336421.

Sinnott, Joan M., Stacy L. Street, Kelly W. Mosteller, and Trina L. Williamson. 1997. “Behavioral Measures of Vowel Sensitivity in Mongolian Gerbils (Meriones Unguiculatus) : Effects of Age and Genetic Origin.” *Hearing Research* 112 (1–2): 235–46. doi:10.1016/S0378-5955(97)00125-1.

Siveke, Ida, Christian Leibold, and Benedikt Grothe. 2007. “Spectral Composition of Concurrent Noise Affects Neuronal Sensitivity to Interaural Time Differences of Tones in the Dorsal Nucleus of the Lateral Lemniscus.” *Journal of Neurophysiology* 98 (5): 2705–15. doi:10.1152/jn.00275.2007.

Siveke, Ida, Michael Pecka, Armin H Seidl, Sylvie Baudoux, and Benedikt Grothe. 2006. “Binaural Response Properties of Low-Frequency Neurons in the Gerbil Dorsal Nucleus of the Lateral Lemniscus.” *Journal of Neurophysiology* 96 (3): 1425–40. doi:10.1152/jn.00713.2005.

Sivian, L. J., and S. D. White. 1933. “ON MINIMUM AUDIBLE SOUND FIELDS.” *The Journal of the Acoustical Society of America* 4 (4): 288–321. doi:10.1121/1.1915608.

Soede, Wim, Frans A. Bilsen, and Augustinus J. Berkhout. 1993. “Assessment of a Directional Microphone Array for Hearing-impaired Listeners.” *The Journal of the Acoustical Society of America* 94 (2): 799–808. doi:10.1121/1.408181.

Sondhi, Man Mohan, and Newman Guttman. 1966. “Width of the Spectrum Effective in the Binaural Release of Masking.” *The Journal of the Acoustical Society of America* 40 (3): 600–606. doi:10.1121/1.1910124.

Spitzer, M W, and M N Semple. 1995. “Neurons Sensitive to Interaural Phase Disparity in Gerbil Superior Olive: Diverse Monaural and Temporal Response Properties.” *Journal of Neurophysiology* 73 (4): 1668–90.

Stecker, G. Christopher, Ian A. Harrington, and John C. Middlebrooks. 2005. “Location Coding by Opponent Neural Populations in the Auditory Cortex.” *PLOS Biology* 3 (3): e78. doi:10.1371/journal.pbio.0030078.

Stern, Richard M., Andrew S. Zeiberg, and Constantine Trahiotis. 1988. “Lateralization of Complex Binaural Stimuli: A Weighted-image Model.” *The Journal of the Acoustical Society of America* 84 (1): 156–65. doi:10.1121/1.396982.

Syka, Josef, Jiří Popelář, Eugen Kvašňák, and Jaromír Astl. 2000. “Response Properties of Neurons in the Central Nucleus and External and Dorsal Cortices of the Inferior

Colliculus in Guinea Pig." *Experimental Brain Research* 133 (2): 254–66. doi:10.1007/s002210000426.

Tanaka, H, K Hachisuka, and H Ogata. 1999. "Sound Lateralisation in Patients with Left or Right Cerebral Hemispheric Lesions: Relation with Unilateral Visuospatial Neglect." *Journal of Neurology, Neurosurgery, and Psychiatry* 67 (4): 481–86.

Ter-Mikaelian, Maria, Dan H. Sanes, and Malcolm N. Semple. 2007a. "Transformation of Temporal Properties between Auditory Midbrain and Cortex in the Awake Mongolian Gerbil." *The Journal of Neuroscience* 27 (23): 6091–6102. doi:10.1523/JNEUROSCI.4848-06.2007.

Ter-Mikaelian, Maria, Dan H Sanes, and Malcolm N Semple. 2007b. "Transformation of Temporal Properties between Auditory Midbrain and Cortex in the Awake Mongolian Gerbil." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 27 (23): 6091–6102. doi:10.1523/JNEUROSCI.4848-06.2007.

Thomas, Hardy, Jochen Tillein, Peter Heil, and Henning Scheich. 1993. "Functional Organization of Auditory Cortex in the Mongolian Gerbil (Meriones Unguiculatus). I. Electrophysiological Mapping of Frequency Representation and Distinction of Fields." *The European Journal of Neuroscience* 5 (7): 882–97.

Tollin, Daniel J., and Tom C. T. Yin. 2005. "Interaural Phase and Level Difference Sensitivity in Low-Frequency Neurons in the Lateral Superior Olive." *The Journal of Neuroscience* 25 (46): 10648–57. doi:10.1523/JNEUROSCI.1609-05.2005.

Van den Bogaert, Tim, Simon Doclo, Jan Wouters, and Marc Moonen. 2008. "The Effect of Multimicrophone Noise Reduction Systems on Sound Source Localization by Users of Binaural Hearing Aids." *The Journal of the Acoustical Society of America* 124 (1): 484–497.

Van den Bogaert, Tim, Thomas J. Klasen, Marc Moonen, Lieselot Van Deun, and Jan Wouters. 2006. "Horizontal Localization with Bilateral Hearing Aids: Without Is Better than with." *The Journal of the Acoustical Society of America* 119 (1): 515–526.

Vanderweele, Dennis A., and Robert M. Abelson. 1973. "Selected Schedules of Reinforcement in the Mongolian Gerbil." *Psychological Reports* 33 (1): 99–104. doi:10.2466/pr0.1973.33.1.99.

Vickers, Deborah A., Brian CJ Moore, and Thomas Baer. 2001. "Effects of Low-Pass Filtering on the Intelligibility of Speech in Quiet for People with and without Dead Regions at High Frequencies." *The Journal of the Acoustical Society of America* 110 (2): 1164–1175.

Victor, J D, and K P Purpura. 1996. "Nature and Precision of Temporal Coding in Visual Cortex: A Metric-Space Analysis." *Journal of Neurophysiology* 76 (2): 1310–26.

Victor, Jonathan D., and Keith P. Purpura. 1997. "Metric-Space Analysis of Spike Trains: Theory, Algorithms and Application." *Network: Computation in Neural Systems* 8 (2): 127–164.

Vonderschen, Katrin, and Hermann Wagner. 2009. "Tuning to Interaural Time Difference and Frequency Differs between the Auditory Arcopallium and the External Nucleus of the Inferior Colliculus." *Journal of Neurophysiology* 101 (5): 2348–61. doi:10.1152/jn.91196.2008.

———. 2012. "Transformation from a Pure Time Delay to a Mixed Time and Phase Delay Representation in the Auditory Forebrain Pathway." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 32 (17): 5911–23. doi:10.1523/JNEUROSCI.5429-11.2012.

———. 2014. “Detecting Interaural Time Differences and Remodeling Their Representation.” *Trends in Neurosciences* 37 (5). Elsevier Ltd: 289–300. doi:10.1016/j.tins.2014.03.002.

Vuorialho, Arja, Petri Karinen, and Martti Sorri. 2006. “Counselling of Hearing Aid Users Is Highly Cost-Effective.” *European Archives of Oto-Rhino-Laryngology and Head & Neck* 263 (11): 988–95. doi:10.1007/s00405-006-0104-0.

Wakeford, Onslow S., and Donald E. Robinson. 1974. “Detection of Binaurally Masked Tones by the Cat.” *The Journal of the Acoustical Society of America* 56 (3): 952–56. doi:10.1121/1.1903354.

Walton, Joseph P. 2010. “Timing Is Everything: Temporal Processing Deficits in the Aged Auditory Brainstem.” *Hearing Research*, Special Issue: Aging and Hearing, 264 (1–2): 63–69. doi:10.1016/j.heares.2010.03.002.

Warren, Edus H, and M. Charles Liberman. 1989. “Effects of Contralateral Sound on Auditory-Nerve Responses. I. Contributions of Cochlear Efferents.” *Hearing Research* 37 (2): 89–104. doi:10.1016/0378-5955(89)90032-4.

Whitton, Jonathon P., Kenneth E. Hancock, and Daniel B. Polley. 2014. “Immersive Audiomotor Game Play Enhances Neural and Perceptual Salience of Weak Signals in Noise.” *Proceedings of the National Academy of Sciences* 111 (25): E2606–E2615.

Wightman, F L, and D J Kistler. 1992. “The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization.” *The Journal of the Acoustical Society of America* 91 (3): 1648–61.

Wilbanks, W. A., and John K. Whitmore. 1968. “Detection of Monaural Signals as a Function of Interaural Noise Correlation and Signal Frequency.” *The Journal of the Acoustical Society of America* 43 (4): 785–97. doi:10.1121/1.1910897.

Winer, Jeffery A., and Christoph E. Schreiner. 2005a. *The Inferior Colliculus*. Springer. http://books.google.com/books?hl=en&lr=&id=6Ce3PMzcRqsC&oi=fnd&pg=PR7&dq=winer+schreiner+Inferior+colliculus+book&ots=OXBOZlwRrE&sig=F5NfdGumWM4h0b1XQhVsQRzShQM.

Winer, Jeffery A, and Christoph E Schreiner, eds. 2005b. *The Inferior Colliculus*. New York: Springer.

Yamada, K, K Kaga, A Uno, and M Shindo. 1996. “Sound Lateralization in Patients with Lesions Including the Auditory Cortex: Comparison of Interaural Time Difference (ITD) Discrimination and Interaural Intensity Difference (IID) Discrimination.” *Hearing Research* 101 (1–2): 173–80.

Yin, T C, and J C Chan. 1990a. “Interaural Time Sensitivity in Medial Superior Olive of Cat.” *Journal of Neurophysiology* 64 (2): 465–88.

Yin, T. C., and J. C. Chan. 1990b. “Interaural Time Sensitivity in Medial Superior Olive of Cat.” *Journal of Neurophysiology* 64 (2): 465–88.

Yin, T. C., J. C. Chan, and L. H. Carney. 1987. “Effects of Interaural Time Delays of Noise Stimuli on Low-Frequency Cells in the Cat’s Inferior Colliculus. III. Evidence for Cross-Correlation.” *Journal of Neurophysiology* 58 (3): 562–83.

Yin, T. C., J. C. Chan, and D. R. Irvine. 1986. “Effects of Interaural Time Delays of Noise Stimuli on Low-Frequency Cells in the Cat’s Inferior Colliculus. I. Responses to Wideband Noise.” *Journal of Neurophysiology* 55 (2): 280–300.

Yin, T. C., and S. Kuwada. 1983. “Binaural Interaction in Low-Frequency Neurons in Inferior Colliculus of the Cat. III. Effects of Changing Frequency.” *Journal of Neurophysiology* 50 (4): 1020–42.

Zurek, Patrick M., Richard L. Freyman, and Uma Balakrishnan. 2004. “Auditory Target Detection in Reverberation.” *The Journal of the Acoustical Society of America* 115 (4): 1609–20. doi:10.1121/1.1650333.

Zurek, Paul M. 1993. “Binaural Advantages and Directional Effects in Speech Intelligibility.” *Acoustical Factors Affecting Hearing Aid Performance* 2: 255–275.

Zwiers, Marcel P, Huib Versnel, and A John Van Opstal. 2004. “Involvement of Monkey Inferior Colliculus in Spatial Hearing.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 24 (17): 4145–56. doi:10.1523/JNEUROSCI.0199-04.2004.

Zwislocki, J., and R. S. Feldman. 1956. “Just Noticeable Differences in Dichotic Phase.” *The Journal of the Acoustical Society of America* 28: 860.

It's over,
now
let's
dance