Complexity, the auditory system, and perceptual learning in naïve users of a visual-to-auditory sensory substitution device.

Brown, David J

Complexity, the auditory system, and perceptual learning in naïve users of a visual-to-auditory sensory substitution device.

By

David J Brown

A Thesis.

Submitted in partial fulfilment of the requirements

of the Degree of Doctor of Philosophy.

I, David John Brown, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: [can be digital signature]

Date:

Details of collaboration and publications:

Collaboration with Simpson, A.J. in Chapter 2. Journal Article: Brown, D.J., Simpson, A.J., & Proulx, M.J. (in press). Visual Objects in the Auditory System in Sensory Substitution: How much information do we need?, *Multisensory Research.*

Abstract

Sensory substitution devices are a non-invasive visual prostheses that use sound or touch to aid functioning in the blind. Algorithms informed by natural crossmodal correspondences convert and transmit sensory information attributed to an impaired modality back to the user via an unimpaired modality and utilise multisensory networks to activate visual areas of cortex. While behavioural success has been demonstrated in non-visual tasks suing SSDs how they utilise a metamodal brain, organised for function is still a question in research.

While imaging studies have shown activation of visual cortex in trained users it is likely that naïve users rely on auditory characteristics of the output signal for functionality and that it is perceptual learning that facilitates crossmodal plasticity.

In this thesis I investigated visual-to-auditory sensory substitution in naïve sighted users to assess whether signal complexity and processing in the auditory system facilitates and limits simple recognition tasks. In four experiments evaluating; signal complexity, object resolution, harmonic interference and information load I demonstrate above chance performance in naïve users in all tasks, an increase in generalized learning, limitations in recognition due to principles of auditory scene analysis and capacity limits that hinder performance.

Results are looked at from both theoretical and applied perspectives with solutions designed to further inform theory on a multisensory perceptual brain and provide effective training to aid visual rehabilitation.

Acknowledgements.

I am much indebted to the people who have supported me. My Mum and Dad obviously, I love you both. Elissa for the kick start, Amy for the early mornings, and Amon Tobin for the late nights.

Obviously, I would like to thank my supervisor Michael Proulx for his invaluable help and humour, Queen Mary University of London for the funding, and the fine folks I've had the pleasure of sharing office and bar space with. Memories I won't forget…for a while.

Table of contents

List of Figures.

List of Tables.

List of Acronyms.

A1	primary auditory cortex
AMD	age related macular degeneration.
AV	audio-visual
BL	Blindness
BOLD	blood-oxygen level dependent.
BS	blindfold sighted
CB	congenitally blind
CVP	cortical vision prosthesis
EB	early blind
FM	frequency modulated
fMRI	functional magnetic resonance imaging
HCI	human-computer interaction
LOC	lateral occipital cortex
PET	positron electron tomography
PSVA	Prosthesis for Substitution of Vision with Audition
RHT	reverse hierarchy theory
RP	retinitis pigmentosa
RT	reaction time
S1	primary somatosensory cortex

SC	speeded classification
SIM	simultaneous presentation
SSD	sensory substitution device
SUCC	successive presentation
TDU	tongue display unit
TMT	tactile matching task
TVSS	tactile visual substitution system
V1	primary visual cortex
VA	visual-to-auditory sensory substitution
VIm	visual impairment
VMT	visual matching task
VP	visual prosthesis
VT	visual-to-tactile sensory substitution
WHO	World Health Organisation

1.0 Introduction

If asked the question ‘What would you miss most if you lost your sight?’ it would be unsurprising to hear responses based on aesthetics like a ‘beautiful sunset’ or ‘the glorious hues of autumn’. On a personal level one may miss the look on a child’s face at Christmas when a snowstorm of wrapping paper reveals delights. Yet often it seems we undervalue vision, partly because vision, or the task of ‘seeing’, is easy. We open our eyes and we see. It is seemingly that simple! If a similar question was posed regarding what functional difficulties you would encounter in vision loss then the value of vision is exemplified. Simple tasks such as access to information, locating objects, navigating in the environment, recognising faces, avoiding potential danger, and many more become problematic. For people with severe visual impairment and blindness these are everyday problems in an environment tailored for visual beings.

In this opening chapter I will first evaluate the prevalence and causes of ‘the problem’ before describing invasive and non-invasive methods of visual rehabilitation. My main focus of non-invasive sensory substitution will look not only at the effectiveness of sensory substitution devices but also the crossmodal correspondences and organization of the brain that facilitate their use. The final section will assess perceptual learning and how this impacts on increasing functioning with sensory substitution.

1.1 Prevalence of visual impairment and blindness.

The 2010 World Health Organisation (WHO) estimation on the prevalence of visual impairment (VIm) and blindness (BL) states that of a global population of 6737.5 million people, 285.39 million (4.24%) suffer from VIm (low vision + blindness) of which 39.37 million (0.58%) are legally classified as blind (visual acuity in the better eye of 20/200 or less, or a reduction in visual field to $<10^0$) (Pascolini & Mariotti, 2012). This incidence is geographical, with only the high income regions of Europe (total population (TP)=13.2%, VIm=9.9% , BL=7%) and the Americas (TP 13.6%, VIm 9.3%, B 8%) proportionately lower than the general population (VIm and BL %'s of visually impaired community, not overall population). Incidence can also be distributed based on age with the WHO subgrouping into 3 categories: 0-14 years old (TP=27.44%, VIm=6.64%, BL= 3.60%), 15-49 years old (TP=52.66%, VIm=28.12%, BL=14.70%), and ≥ 50 years old (TP=19.90%, VIm=65.24%, BL=81.70%). As clearly demonstrated by the figures VIm and BL are mainly associated with ageing, implying a gradual degenerative disorder, and with areas of low GDP, indicating that VIm and BL at least partially correlated with health care provision. The prevalence of VIm and BL based on age and particularly regions, is salient in the distribution of the causes of VIm and BL and subsequent availability and provision of the complementary techniques described throughout this thesis.

1.2 Causes of blindness.

The etiologies and symptoms of VIm and BL are wide-ranging. They share commonality in that they degrade vision and are mainly ocular disorders. The wide-ranging causes and symptoms require differing methods of rehabilitation. Table 1.1 gives a brief overview of 8 leading causes of Vim and BL worldwide giving relative prevalence, symptoms, eye damage, treatments, and suitability for implant technology or sensory substitution.

Table 1.1. Major causes of blindness including symptoms, treatment and suitability for sensory substitution.

Disorder	**Prevalence**	**Symptoms**	**Damage to eye**	**Treatments**	**IM**	**SSD**
Cataracts	BL 51%, VI 33%	Gradual reduction of light passing through lens. If bilateral can cause blindness	Opacification of the lens via clumping of proteins.	Contact and intraocular lens. Surgery	x	✓
Glaucoma	BL 8%, VI 2%	Chronic – vision loss form periphery inwards. Acute – pain, redeye, rapid loss of vision Raised IOP	Some glaucomas can lead to atrophy of the optic nerve. Damage to trabecular network. Retinal ganglian cell loss	Medication Surgery • Canaloplastry • Trabeculectomy • Drainage implants	x	✓
Age Related Macular Degeneration	BL 5% VI 1%	Gradual degradation of central visual field from drusen on macula lutea, contrast sensitivity	Atrophy of retinal pigment epithelium layer (DRY AMD). Retinal scaring (WET AMD)	Medication Surgery • Photodynamic therapy • photocoagulation • Macular translocation Lens implantation	✓	✓
Corneal opacities	BL 4% VI 1%	Blurred vision, photophobia, swelling of eye tissue, vision loss.	Corneal damage from chemical, physical strike	Medication Surgery • Phototherapeutic keratectomy Cornea transplant	x	✓
Trachoma	BL 3% VI 1%	Infectious, conjunctivitis Entropian eyelid, scar cornea, Vision loss.	Corneal opacity from scarring	Medication – antibiotics Surgery • Cornea transplant Lifestyle measures	x	✓
Uncorrected refractive errors	BL 5% VI 42%	Short, long sighted, Blurring of near/far objects	Curvature of cornea, astigmatism.	Glasses/contact lens Refractive surgery.	x	✓
Diabetic retinopathy	BL 1% VI 1%	Dots on retina,	Macula edema	Medication Surgery • Photocoagulation • Vitrectomy	x	✓
Retinitis pigmentosa	1 in 5000 worldwide Genetic inheritance	Peripheral vision loss, photophobia, blindness	Retinal pigment epithelium distrophy	Medication	✓	✓

Cortical Blindness.

All disorders described in Table 1.1. are ocular visual impairments (i.e. involve the eye). They are thus suitable for sensory substitution as while the technique is not dependent on intact retinal cells it does require a functioning cortex. Cortical blindness in which vision loss is due to damage to the cortex, even though the eyes may be physiologically intact, exemplifies the magnitude of visual rehabilitation. i.e. using techniques such as retinal implants and sensory substitution to bypass ocular disorders is not an all-encompassing solution to blindness, even theoretically. It also shines a light on Paul Bach-y-Rita's quote that underpins the theory of the field of sensory substitution research.

"We see with our minds, not our eyes."

Cortical blindness is a bilateral dysfunction of the striate cortex, or its immediate afferents, causing loss on conscious vision in the contralateral hemifield (Zrenner et al., 2011). The eye may be fully functioning (i.e. responsive to light) but damage to primary visual cortex (V1) disrupts processing of information. This is particularly debilitating as V1 is not only thought to be critical in the processing of visual features such as orientation, colour and localisation, but also a conduit of visual information into higher-order extrastriate areas. Cortical blindness can be congenital or acquired with the primary causes being loss of blood flow to cortex due to posterior cerebral artery blockage (ischemic stroke), cerebral haemorrhage from cardiac surgery, cerebral angiography, head trauma to the occipital lobe and bilateral lesions of V1 (Stingl, Greppmaier, Wilhelm, & Zrenner, 2010) Most acquired cortical blindness is transient with spontaneous visual improvements in the first month after the trauma, although with little improvement after three months (Prokofyeva & Zrenner, 2012). Total vision loss is rare, with a general retention of degraded residual vision, or permanent, although this is dependent on the etiology. For example, bi-lateral lesions of visual cortex have a limited

prognosis. Prevalence of cortical blindness as a factor of total blindness is low, unsurprising as most blindness is an age related disorder. It is however, the leading cause of bilateral vision loss in children in Western countries (Zrenner et al., 2011) and frequently occurs with other ocular disorders (Martinez-Conde, Macknik, & Hubel, 2004). Due to being non-ocular interventions differ from ones described in the rest of the thesis, as invasive implant technology and sensory substitution require a functioning occipital cortex. Present treatments for cortical blindness include saccadic training to improve the patient's occulomotor strategies and repetitive training to facilitate perceptual learning in the damaged visual system. For a full review of visual rehabilitation for cortical blindness see (Das & Huxlin, 2010).

1.3 Treatments.

The varied and wide ranging aetiologies of VIm and BL illustrate not only the complexity of the problem faced in visual rehabilitation but emphasises the necessity of utilising a variety of compensatory techniques. The following section looks at how vision rehabilitation research has utilised both invasive and non-invasive methods and technologies. Firstly however we must consider what is meant by vision.

1.3.1 What is 'vision' as a restoration goal?

Visual ability is often measured optically as visual acuity. This is measured using Snellen charts consisting of lines of alphanumeric characters which are required to be identified by the patient. Symbols representing normal acuity subtend an angle of five minutes of arc with line thickness and spaces subtended 1 minute of arc. This line is defined as 20/20, that is, the

smallest that can be read from 20 feet away by a person with normal visual acuity. The legal blindness definition is 20/200. Here characters that can be read at 20 feet from the chart are equivalent to what could be read from 200 feet away with normal acuity. Normal acuity does not necessarily equate to normal vision however, as colour blindness, reduced contrast, inability to accurately detect motion, reduced size of the field of view, can cause visual impairment.

With the acuity and resolution of the eye being fine grained restoration of 'normal' vision by means of implant or sensory substitution is way beyond the present technology and understanding of brain function. Levels of acuity have been measured using these techniques but the primary goal is a drive towards ‘useful’ functional vision, that is, a working acuity that facilitates functioning in everyday tasks usually mediated by vision; localisation and recognition of objects, navigation within the environment, and acquisition of data/information. There are alternate definitions of what constitutes ‘vision’ from a phenomenological viewpoint, and in sensory substitution (Auvray & Myin, 2009) but for the present thesis functional vision is the primary concern.

1.3.2 Invasive - Visual Prostheses.

Visual prostheses (VP) are technological devices used to elicit phosphenes via electrical stimulation of any part of the visual system. A phosphene is a sensation of light caused by mechanical or electrical stimulation rather than by light itself. While a number of devices have been developed, from simulated prototype models to devices in clinical testing, they can be categorised by the anatomical point of stimulation. Retinal prosthesis (epi and sub) are targeted to eye pathologies, such as age related macular degeneration (AMD) and retinitis pigmentosa (RP), where retinal damage is not complete and the optic nerve is intact.

Pathologies with almost total atrophy of optic nerve or severe eye disfigurement require non-retinal prostheses targeted at cortical or optic nerve levels.

Retinal implants.

As over half of the cases of blindness are caused by retinal damage (Zrenner, 2002b) biomedical implant technology in this anatomical structure are a logical area of research. Two main types of retinal implants have been developed based on the area of implantation: epi retinal, and sub-retinal. In all cases the purpose of implants is to electrically stimulate surviving retinal cells to elicit a basic visual percept. Figure 1.1., adapted from Zrenner et al (2011) shows the set up for a subretinal implant system. Similar systems are used for epi-retinal implants but with the array mounted on the inner retinal layer.

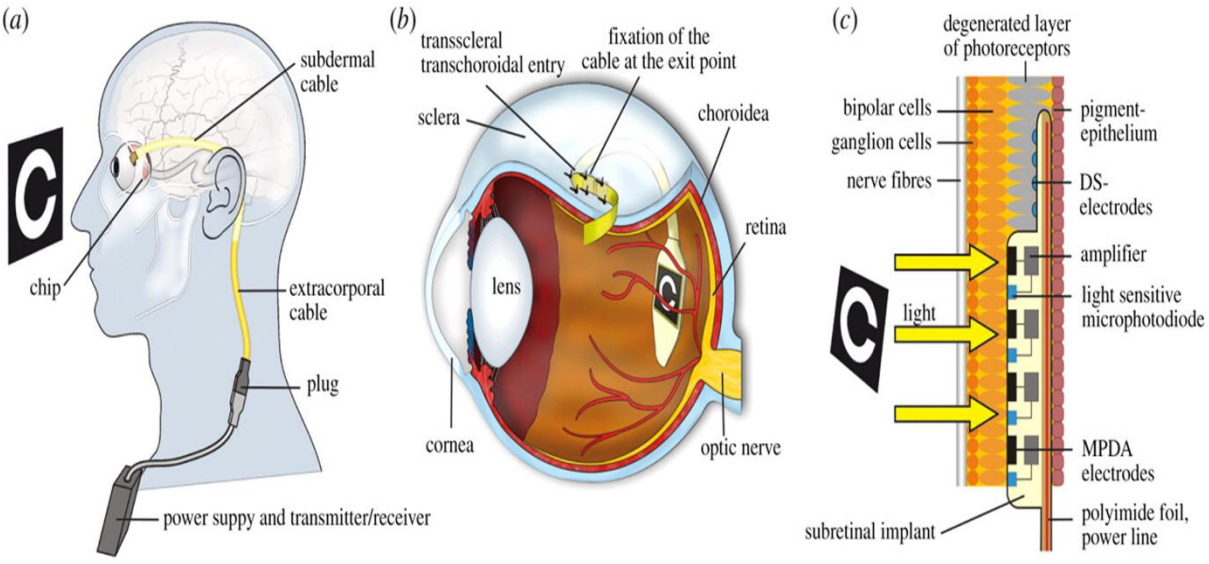

Figure 1.1: Representation of a subretinal implant system, adapted from (E. Zrenner et al., 2011). The cable from the implanted chip in the eye leads to the exit behind the ear, and connects to wirelessly operated power control unit. b) Position of the implant under the transparent retina. c) MPDA photodiodes, amplifiers and electrodes in relation to retinal neurons and pigment epithelium.

Epi-retinal implants.

The basic epi-retinal implant technology consists of three components: an external sensor (video camera) to record sensory information, an external processor to reduce image resolution and convert information into spatial and temporal patterns, with a wireless RF to pass information to an implanted silicon platinum array mounted on the inner retinal layer. The use of external components conveys a number of advantages: implants are smaller as they don't require internal power , are easy to upgrade (Wentai et al., 2000), the physician has external control of the system allowing tailoring of image processing to individuals (Weiland & Humayun, 2005; Weiland, Liu, & Humayun, 2005), and the vitreus humour of the eye acts as an effective heat sink for the implant (Piyathaisere et al., 2003).

Primarily, epi-retinal implants are a feasible intervention for sufferers of RP and AMD, signified by damage to photoreceptor cells on the outer retina but intact ganglion cells in the inner and middle retinal layers. For AMD, symptomised by degeneration of the central visual field, epi-retinal implants supplement remaining peripheral vision with central vision information (Chader, Weiland, & Humayun, 2009). This focus on electrical stimulation of the middle and inner retinal layers allows implants to be suitable for patients with retinal diseases beyond the photoreceptor level as large portions of the retina (external layer) are bypassed. Aside from an intact ganglion cell layer, suitable candidates for implant technology must also display good general health and a commitment to rehab as learning to use the technology is arduous and physically uncomfortable.

At present retinal implants offer low acuity and technological problems. The use of an external camera requires head rather than eye movements for a 'shift in gaze', while stimulation affects not just ganglion bodies but also other axons associated with retinotopic areas. Adjustment of the subsequent distorted patterns requires computational manipulation

at the image processing stage (Zrenner, 2002b) and further image processing is required to compensate for processing normally found in bypassed retinal layers (Weiland et al., 2005).

The first tested device, the ARGUS used a silicon platform array of 16 electrodes with the follow up ARGUS II having a 60 electrode array, although the functional acuity is likely to be less as not all electrodes will contact target cells (Weiland et al., 2005). While a 200 electrode model is under development simulations have indicated this to be insufficient for tasks such as reading, face recognition and navigation (Weiland et al., 2005) highlighting the challenge of raising acuity while keeping the physical properties of the device small. Research has implied that this hypothetical acuity may be overstated as a 60 channel output has been deemed sufficient for reading of enlarged text (Fornos, Sommerhalder, & Pelizzone, 2011), and room navigation in experienced users (Dagnelie et al., 2007). In clinical trials with the 25 electrode EPI-RET, which requires the crystalline lens to be replaced with a receiver chip, all patients were able to distinguish spatial and temporal patterns of stimulation (Klauke et al., 2011). However, as we evaluate in Chapter 3 with sensory substitution, high levels of acuity may not be a requisite in successful object recognition.

Sub-retinal implants.

Sub-retinal implants are passive devices (internal power source - although some sub-retinal implants require external power to enhance signal) implanted on the outer surface of the retina between the photoreceptor layer and retinal pigment epithelium. Stimulation is directly to retinal cells and, unlike epi-retinal, relies on normal processing of the inner and middle retinal layers (Weiland & Humayun, 2005). The standard implant consists of a silicon wafer of light sensitive microphotodiodes implanted directly adjacent to the damaged photoreceptors (Chader et al., 2009; Zrenner, 2002a) which generate signals from incoming light. Sub-retinal implantation conveys a number of advantages over epi-retinal implants:

fixation is less problematic as it is constrained by distance between retina and retinal pigment epithelium, normal eye movements can be used to fix gaze, and retinotopic stimulation is more accurate as the pattern of light is a direct reflection of the stimulus image (Weiland et al., 2005). However implant size, and subsequent acuity, is restricted by the sub-retinal space, while heat generated by the implant may damage the retina. Sub-retinal implants are also not suitable for retinal diseases that go beyond the photoreceptor layer.

Clinical studies have demonstrated the efficacy of sub-retinal implants. A study using 10 implanted patients demonstrated improvement in perception of visual contrast, shape and movement (Zrenner, 2002a, 2002b)while Stingl using the Retina Implant AG, a 1500 microphotodiode implant, reported 5 of 8 patients experiencing visual perceptions, although external power was required and 6 of 9 implants failed within 9 months of implantation (Stingl et al., 2013). Patients using the same implant system however were able to read letters and combine them into words (Zrenner et al., 2011) and recognise Braille characters (Lauritzen et al., 2012). Further work done by the Boston Subretinal Implant Project focused on analysis of implant function, with all patients reporting phosphene production and some successfully completing shape recognition and motion detection tasks (Rizzo, Wyatt, Loewenstein, Kelly, & Shire, 2003a, 2003b).

If the physiology of the eye, due to the type of impairment, is not suitable for retinal implant technology then it is logical to look further up the visual hierarchy for an area of stimulation. Two implant technologies have been developed for this, although only in proof of concept rather than marketable devices.

1.3.3 Invasive - Optic Nerve Implants.

The optic nerve consists of retinal ganglion cell axons and support cells which run from the retina, via visual nuclei, to primary visual cortex. As all visual information transducted in the retina is transmitted to cortex via the optic nerve, it provides a good theoretical candidate for implant stimulation, as a minimal number of stimulators should provide phosphene sensitivity over a large area of the visual field (Capelle, Trullemans, Arno, & Veraart, 1998). Initial work from Belgium used a spiral cuff biocompatible four electrode array implanted on the surface of the optic nerve of a sufferer of RP. During stimulation perceived phosphenes were described as 2-60 dot clusters, of various colours, arranged in rows or arrays (Veraart et al., 1998). Subsequent research demonstrated simple pattern recognition, localisation and discrimination of objects, and mobility assessments (Brelen, Duret, Gerard, Delbeke, & Veraart, 2005; Delbeke, Oozeer, & Veraart, 2003; Delbeke et al., 2001; Delbeke et al., 2002; Veraart, Wanet-Defalque, Gérard, Vanlierde, & Delbeke, 2003) with pattern recognition orientation at high levels (63% and 100% respectively) after training

Considering the optic nerve is a highly dense bundle of between 0.7 and 1.2 million nerve fibres accurate focal stimulation and detailed perception is problematic using surface mounted electrodes. C-Sight lab developed a penetrating electrode in an attempt to increase spatial resolution and lower the threshold of electrical stimulation. Captured sensory information was coded into specific spatiotemporal signals and delivered as impulses via an embedded array. Action potentials generated and transmitted to visual cortex give a visual perception (Chai, Li, et al., 2008)allowing Chinese character recognition in behavioural experiments (Chai et al., 2007; Chai, Zhang, et al., 2008).

1.3.4 Invasive - Cortical implants.

Foerster's creation of phosphenes via electrical stimulation of the occipital cortex (Foerster, 1929) demonstrated the theoretical possibility of cortical vision prostheses (CVP). Although this initial work produced single phosphenes, Krieg later postulated that as the visual cortex is 'roughly' retinotopic, concurrent stimulation of multiple cortical sites could produce a coherent image (Krieg, 1953). Technological advances by the late 1960's resulted in the first device for permanent cortical stimulation. Brindley and Lewin's 'surface' device sat on top of visual cortex with electrodes sited beneath the pericranium (Brindley & Lewin, 1968b). Position of elicited phosphenes in the visual field, from a 25V current, were found to roughly correspond with electrode position on the cortical surface allowing the user to recognise patterns produced by the array. However, aside from flickering phosphenes, spatial resolution was poor with electrodes less than 3mm apart producing a 'strip' of light rather than individual phosphenes and flickering.

Dobelle and Mladejovsky, showed that constant stimulation produced phosphenes that dimmed over a 10-15 second time frame, postulating that phosphene brightness was a logarithmic function of the current value of the stimulating device. This could be counteracted by frequently refreshing the stimulation to avoid adaptation (Dobelle & Mladejovsky, 1974). Interestingly this also found in vision. Martinez-Conde and colleagues (2004) found that when a small screen was attached to the eye to negate formation of a new image from eye movements, the image phases as a result of not being able to refresh, i.e. showing rapid adaptation (Dobelle, Mladejovsky, & Girvin, 1974; Martinez-Conde et al., 2004). Dobelle also demonstrated that the spatial organisation of phosphenes in the visual field was interlinked with eye movements requiring prospective implants to have eye movement detection to centralise phosphenes in the visual field (Dobelle et al., 1974).

With rapid technological advances the development of future CVP's may be constrained by retinotopy. Aside from the 'rough' mapping being poor for precise measurements, the topographic representation of the visual field is not linear with two adjacent rows of neurons not necessarily mapping two adjacent areas of the visual field. What is stimulated on the array may not be represented as such. Considering visual neurons also code for colour, orientation, direction and depth it is no surprise recent CVP research has focused on developing an accurate retinotopic map, with technological advances such as larger arrays and external power a secondary consideration (Rush & Troyk, 2012; Srivastava, Troyk, & Dagnelie, 2009; Srivastava et al., 2007; Troyk et al., 2003).

Interim Summary.

So far the prevalence of visual impairment and blindness and the wide ranging aetiologies have emphasised the difficulty and diverse approaches required in visual rehabilitation. Invasive technologies are limited to eye pathologies where there is remaining retinal layers, such as AMD and RP, and provide a very low 'functional' acuity.

Implant technology is also limited by cost and availability. For example, the ARGUS II presently costs $150,000 and while Vaidya and colleagues make a case for cost effectiveness in RP relative to conventional interventions (Vaidya et al., 2014) this is still beyond the financial reach of most sufferers. Availability is also limited by the number of clinical trials presently being conducted by the various labs, and the suitability for implantation as stated above.

1.4 Non-invasive – sensory substitution.

Alternative methods and technologies have been developed that use non-invasive methods to aid visual rehabilitation and provide some of the functional aspects of vision, in a more cost effective, and widely available manner. This field of non-invasive visual rehabilitation is called sensory substitution.

1.4.1 What is sensory substitution?

At a basic definitional level sensory substitution transmits information usually attributed to an impaired sensory modality, via an unimpaired sensory system. This is facilitated using a combined hardware/software technological prosthesis, the sensory substitution device (SSD), which extracts sensory information from the environment, subjects it to a conversion process, and then transmits it back to the user. The processing of the transmitted information is carried out in the cortical areas of the brain to elicit the final percept. Sensory substitution can therefore be seen as brain or human-computer interaction (HCI) with success being dependent not only in the technology (particularly the conversion algorithm) but also in an understanding of how the brain processes multimodal sensory information. Naturally these feedback with the latter informing the principles of the former and experimental research using SSDs providing a better understanding of the brains processing of multisensory perception. SSDs therefore can be viewed not only as an aid to visual rehabilitation in the blind but as a valuable tool in perceptual research in the general population.

In the next section I will evaluate the crossmodal correspondences that ultimately inform SSD algorithms. Secondly I will discuss the organization and functioning of the brain, particularly in visually impaired populations, that facilitates the use of SSDs. In doing this I argue against the idea held less than 100 years ago of a static brain with cortical areas

‘hardwired’ to process modality specific sensory information, in favour of a dynamic brain model based on the integration of sensory information from multiple modalities and non-modality specific processing based on function

1.4.2 Crossmodal correspondences

Perception is not unimodal. At a basic level this is epitomised in language. For example, ‘bright’ and ‘loud’ are applied to both visual colour and auditory sound characteristics demonstrating that we don’t ‘think’ of vision and audition as completely independent. It is unquestionable that we have sensory organs specialized to extract particular forms of sensory information from the environment i.e. eyes for light information, ears for pressure/sound, tongue for chemical/taste etc but it also widely accepted that the brain integrates sensory information from multiple modalities to provide a cohesive ‘view’ of the world. Indeed the integration or binding of information from multiple modalities has been shown to facilitate superior performance compared to a unimodal counterpart. While crossmodal correspondences have been studied across multiple modalities this review will focus on those between vision and audition, as these are the substituted and substituting modalities in the visual-to-auditory sensory substitution paradigm.

In an early demonstration of crossmodal correspondences, Kohler found that when presented with two visual shapes, one spiky and one rounded, and required to match them with two nonsense words ‘Takete’ and ‘Baluma’ most people paired the spiky shape with ‘Takete’ emphasising an arbitrary object sound association(Kohler, 1947) . This sound, or phonetic symbolism made famous in Ramachandran and Hubbards ‘Bouba/Kiki’ paradigm has been replicated using different words (Ramachandran & Hubbard, 2003), across cultures (Davis, 1961; Hinton, Nichols, & Ohala, 2006), in pre-lingual infants (Maurer, Pathman, &

Mondloch, 2006), and across other modalities such as taste (Bremner et al., 2013), and touch (Fryer, Freeman, & Pring, 2014).

While this demonstrates a robust natural crossmodal association, two seminal studies go further in illustrating that concurrently presented audio-visual (AV) information can not only influence efficiency of information processing but also bidirectionallly modulate percepts. In the McGurk illusion synchronously paired visual and auditory components elicit a perception of a non-presented intermediate phoneme. For example the auditory phoneme 'ba-ba' is overdubbed on a video of a person saying 'ga-ga' and the perceived phoneme is that of 'da-da'. As the multimodal information is incongruent, visual information is weighted more heavily than auditory providing a 'best guess' of the true percept (McGurk & MacDonald, 1976). Conversely Shams double flash illusion demonstrates a heavier weighting to auditory information. Participants are presented with a brief visual flash on screen accompanied by one or two rapid auditory 'beeps' and required to indicate how many flashes they see. Providing the two 'beep's are within a temporal window then two flashes are perceived. If one beep is presented then one flash is perceived. As there is objectively only one flash the concurrent auditory information is modulating the final 'visual' percept (Shams, Kamitani, & Shimojo, 2000) .Both of these crossmodal illusions are robust to feedback implying low-level perceptual integration.

We live in multisensory environments therefore it seems ecologically valid that we develop to process integrated information and that it should convey behavioural advantages over unimodal perception. Crossmodal audio-visual information has been shown to enhance visual perception (Frassinetti, Bolognini, & Ladavas, 2002) visual search (Iordanescu, Guzman-Martinez, Grabowecky, & Suzuki, 2008) and increase performance in spatial and temporal tasks. In speeded classification (SC) paradigms in which participants have to rapidly

discriminate visual targets while presented with task irrelevant auditory stimuli, response times and accuracy decrease if the auditory stimulus is incongruous i.e. high visual elevation paired with low pitch tone (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Marks, 1974) with developmental studies showing this even in 3-4 month old infants (Walker et al., 2010). Interestingly, in a SC paradigm, replacement of the auditory stimuli with vocalised words 'high' and 'low' elicited the same correspondence suggesting a high-level semantic, linguistic modulation (Gallace & Spence, 2006). High pitch also maps to other dimensions such as brighter and lighter stimuli (Marks, 1987; Martino & Marks, 1999; Odgaard, Arieh, & Marks, 2003) higher spatial frequency (Evans & Treisman, 2010), upward movement (H. H. Clark & Brownell, 1976), smaller object size (Evans & Treisman, 2010; Gallace & Spence, 2006; Marks, 2004), and more angular shapes (Marks, 1987).

The robust mapping of pitch to elevation is important in the conversion algorithm of The vOICe SSD, coding for y-axis spatial position. If we are trying to convey functional information via an alternate modality it makes sense to represent it in a dimension that is naturally understood by the brain. A second spectral factor in The vOICe algorithm, the mapping of visual brightness to auditory loudness is also demonstrated in the crossmodal research. For example, both adults and children are found to make consistent crossmodal mappings between the brightness of a visual object and the loudness of a concurrently presented auditory stimulus (Marks, 1987; Stevens & Marks, 1965).

As shown by the McGurk and double flash illusions, weighting of specific unimodal information is influential in the percept. This weighting of AV information has also been demonstrated to effect multisensory integration. Visual information has been shown to dominate over concurrent audio information in bimodal spatial perception (Alais & Burr, 2004a; Bertelson & Aschersleben, 1998; Driver & Spence, 1998) and motion (Kitagawa &

Ichihara, 2002; Lewis, Beauchamp, & DeYoe, 2000; Soto-Faraco, Spence, & Kingstone, 2004b),while in temporal tasks the opposite is found with auditory dominance for interval duration (Burr, Banks, & Morrone, 2009; Grondin, 1993; Ortega, Guzman-Martinez, Grabowecky, & Suzuki, 2014; Romei, De Haas, Mok, & Driver, 2011), synchronization of auditory and visual flicker (Shipley, 1964) and rate perception (Recanzone, 2003). As will be seen, this visual preference for spatial information and auditory preference for temporal information will be important in multisensory learning.

Crucial in multisensory integration is the binding of the unimodal stimuli into one perceived event based on: low-level spatial and temporal synchrony, (Spence, 2011) temporal correlation (Radeau & Bertelson, 1987; Recanzone, 2003), or top down cognitive factors such as semantic congruency (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004).Spatiotemporal intermodal incongruence's or conflicts elicit both immediate and after effects. Incongruent audio-visual spatial information will show a localisation bias towards the visual information, in the ventriloquist effect, even when cued to the auditory stimulus (Bermant & Welch, 1976; Bertelson & Radeau, 1981) - this is evaluated in Chapter 4 when incongruent audio-visual stimuli are used to counteract harmonic distortion. A temporal ventriloquist effect has been shown for time perception. Separation of asynchronous audio-visual stimuli was perceived as shorter if presented in congruent rather than incongruent spatial locations (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002; Vroomen & de Gelder, 2003) with the auditory information appearing to dominate (Fendrich & Corballis, 2001; Soto-Faraco et al., 2004b).

The evidence from the AV crossmodal literature demonstrates that natural perception involves the integration of sensory information from multiple modalities and that this can be behaviourally advantageous if congruent in low or high-level dimensions. It also provides

solid reasoning for the specific principles of the algorithms used in SSDs. As already stated the technology is one two major components of sensory substitution. The other is the brain, and how this is structurally and functionally organized in a way that promotes effective SSD use.

1.4.3 Crossmodal plasticity, and the metamodal theory of brain organization.

A misconception in neuroscience, held until recent decades, of a static model of brain organisation involving 'hardwired' cortices specific to a sensory input, has been challenged in recent years. The integration of crossmodal information in perception, as described previously, demonstrates a level of brain flexibility while seminal work by the likes of Hubel and Wiesel demonstrate that, far from being static, the human and non-human brain undergoes rapid massive changes during development and following injury and is therefore dynamic, flexible and plastic. While initially thought to be restricted to early childhood (Cohen et al., 1999; Hubel & Wiesel, 1970) recent discoveries have implied that this neural reorganization occurs well into adulthood. Neuroplasticity may be developmental, in response to brain injury, or through perceptual learning and has been shown to occur intra-modally in the case of peripheral damage or, crucially for our field of interest, crossmodally after sensory deprivation.

Intramodal plasticity refers to changes in the cortical representation within a sensory system, through an increase or decrease in use, due to peripheral damage, injury or expertise. Imaging studies have demonstrated the expansion and reconfiguration of cortical maps after injury with a reduction of map size in the injured modality/limb and map expansion for the limb or modality experiencing additional use (Rauschecker, 2008). Expert musicians, contrasted to controls, show expansion of cortical maps in primary auditory cortex (Pantev et al., 1998)

whilst this auditory map expansion, through ‘injury’, is also found in blind populations (Elbert et al., 2002). In the tactile modality, primates trained on frequency discrimination show a topographic reorganization of hand representation (Recanzone, Schreiner, & Merzenich, 1993), with primate and human studies demonstrating rapid massive cortical reorganisation for areas of primary somatosensory (S1) and primary motor cortex after finger amputation (Kaas, 2000; Weiss et al., 2000) or nerve blockage (Weiss, Miltner, Liepert, Meissner, & Taub, 2004). In blind populations, anecdotally attributed to having supra-normal abilities in their unimpaired senses, Braille readers show superior performance on tasks involving two-dimensional stimuli presented on a finger pad (Foulke & Warm, 1967), spatial offset in an embossed dot pattern (Grant, Thiagarajah, & Sathian, 2000), and grating or bar stimuli orientation discrimination (Stevens, Foulke, & Patterson, 1996; Van Boven, Hamilton, Kauffman, Keenan, & Pascual-Leone, 2000) implying a reconfiguration of somatosensory maps through perceptual training. Imaging studies on blind Braille readers show expanded maps for the ‘reading finger’ compared to other fingers, and sighted controls (Pascual-Leone & Torres, 1993; M. Wong, Gnanakumaran, & Goldreich, 2011), with TMS to motor cortical areas confirming this (Pascual-Leone et al., 1993). Longitudinal studies imply this reconfiguration takes place in two phases; a rapid transient enlargement due to unmasking of prior connections or synaptic efficacy, followed by a slower less dynamic expansion signifying structural plasticity (Pascual-Leone, Hamilton, Tormos, Keenan, & Catala, 1999). The expansion of cortical maps intramodally may result in a magnification of the haptic abilities in this population explaining superior tactile discrimination. However, TMS to sensorimotor cortex in pattern discrimination tasks showed superior performance for blind Braille readers, compared to sighted, and non-Braille reading controls (Pascual-Leone et al., 1999) suggesting more than intramodal plasticity. Interestingly, imaging studies on Braille readers also show recruitment of cortical areas not associated with the tactile

modality, e.g. the occipital cortex (Sadato et al., 1998; Sadato et al., 1996) implying that plastic changes occur across modalities as well.

Crossmodal plasticity refers to the reassignment of functional processing to an alternate modality, for example the processing of auditory information by the visual cortex. Invasive studies on animals have demonstrated this reassignment. Rauschecker's (1995) sound localization task in binocularly deprived cats showed crossmodal compensation at neuronal and behavioural levels. An increase in density and sharpening of auditory filters was found in the sulcus of the anterior ectosylvian cortex of the visually deprived cats but not sighted controls. Visual areas of the anterior ectosylvian decreased dramatically in size implying expansion of the auditory field into the adjoining visual areas (Rauschecker, 1995). Previously Rauschecker (Rauschecker, Tian, Korte, & Egert, 1992) demonstrated visuotactile plasticity in that facial vibrissae of BD cats showed crossmodal compensatory hypertrophy and subsequent expansion of the sensory field into visual areas.

In a fascinating animal study by Sur et al (1990) the retinal nerves of young ferrets were directed to project into the auditory thalamus, in essence creating a 'new' visual cortex in a brain region predestined to be auditory in nature. Not only did this new cortex show a topographical organization with neurons selective for orientated visual stimuli, as would be found in the 'normal' visual cortex, but in a further study by von Melchner et al (2000) response to light stimuli demonstrated functionality (Sur, Pallas, & Roe, 1990; von Melchner, Pallas, & Sur, 2000). These demonstrations of neural rewiring, expansion of receptive fields, and crossmodal functioning in sensory deprived animals suggest similar should be found in the visually impaired.

Imaging studies have shown recruitment of visual cortex in blind participants in a multitude of non-visual paradigms. Veraart, using Positron Electron Tomography (PET), demonstrated

abnormal occipital activity in blind subjects at rest compared to sighted controls (De Volder et al., 1997; Wanet-Defalque et al., 1988) while task independent activity of occipital cortex has been shown in auditory tasks (Kujala et al., 1995; Roder et al., 1999; Sadato et al., 1996). Sound localisation is an important factor for the blind in navigation and numerous studies have looked at auditory localisation in these populations. Compared to controls, congenitally blind showed higher activation in posterior regions implying occipital activation (Leclerc, Saint-Amour, Lavoie, Lassonde, & Lepore, 2000) with a later study showing enhanced audio/visual area connectivity (Leclerc, Segalowitz, Desjardins, Lassonde, & Lepore, 2005). In another PET sound localisation study Weeks (2000) showed strong activation in right dorsal occipital cortex in blind but not sighted populations (Weeks et al., 2000). As this region is implicated in visuo-spatial processing there is reason to suggest that functional and neuronal coding abilities are retained in these areas enabling processing of information from an alternate modality (Collignon, Voss, Lassonde, & Lepore, 2009). This idea is corroborated by studies that demonstrate similarities in processing of visual and auditory motion (Poirier et al., 2005; Poirier, Collignon, et al., 2006; Saenz, Lewis, Huth, Fine, & Koch, 2008).

Crossmodal plasticity is also demonstrated in visuo-tactile domains, especially illustrated by imaging studies on Braille readers. As Braille can be considered a basic but highly effective form of direct sensory substitution this is salient. Braille reading is a complex cognitive task involving a number of processes: control of finger movements, perception of raised dots, pattern recognition, and semantic and lexical processing. Unsurprisingly, activation is found in typical somatosensory regions concerned with each process (Battaglia-Mayer et al., 2001; Lloyd, Morrison, & Roberts, 2006; Marconi et al., 2001; Nachev, Kennard, & Husain, 2008), however, activation is also shown in ‘visual’ areas in Braille readers. PET studies have shown V1 activation for a word/non word discrimination task in early blind (EB) and congenitally blind (CB) Braille readers, and also in non-Braille tasks constructed from Braille fields

(Sadato et al., 1996). Functional magnetic resonance imaging (fMRI) in a non-Braille task showed activation of ventral occipital cortex and deactivation of secondary somatosensory cortex in blind Braille readers but not controls (Sadato et al., 1998) with similar activation found using a passive, rather than active, Braille hand finger (Sadato, Okada, Honda, & Yonekura, 2002). Imaging studies that control for linguistic processes have demonstrated almost identical activation of occipital and occipito-temporal visual areas for verb generation in Braille and audio presentations for blind and sighted participants respectively (Burton, Snyder, Conturo, et al., 2002; Burton, Snyder, Diamond, & Raichle, 2002) with TMS to medial-occipital areas interfering in this generation in congenitally blind (Amedi, Floel, Knecht, Zohary, & Cohen, 2004). Increased visual cortical recruitment in blind compared to sighted has also been shown for non-Braille tasks such as tactile motion (Ricciardi et al., 2007) and, orientation discrimination (Ptito, Moesgaard, Gjedde, & Kupers, 2005).

While the behavioural and imaging studies illustrate crossmodal plasticity in blind populations the speed of this can be evaluated using short-term visual deprivation in sighted populations. Blindfolding studies have demonstrated behavioural advantages for auditory perception, reducing localisation inaccuracies (Lewald, 2007), improving loudness and pitch discrimination (Gibby, Gibby, & Townsend, 1970) and increasing perception of harmonicity, an idea we discuss in Chapter 4 (Landry, Shiller, & Champoux, 2013). Five days of blindfolding is enough to improve discrimination of Braille characters and show crossmodal activation in occipital cortex (Kauffman, Theoret, & Pascual-Leone, 2002; Merabet et al., 2008) with as little as 90 minutes visual deprivation enough to elicit superior tactile perception (Facchini & Aglioti, 2003) with similar duration deprivation enough to excite visual cortex as shown by fMRI (Boroojerdi et al., 2000).

These results illustrate three important points; Firstly, intramodal and crossmodal plasticity is found in the general population and not just those with sensory impairment. Secondly, plasticity can be extremely rapid. Too rapid to be consistent with the formation of new neural connections implying that in many cases crossmodal plasticity involves the unmasking of previously established, but redundant, connections (Pascual-Leone & Hamilton, 2001). If we consider the unmasking of connections then we must also assume there are already established connections between unimodal cortices. This has been demonstrated in that sensory areas in primates brains have afferent connections from multiple modalities (Falchier, Clavagnier, Barone, & Kennedy, 2002; Murata, Cramer, & Bach-y-Rita, 1965; Rockland & Ojima, 2003). Thirdly, we must take care when discussing sensory substitution to understand that improvements in performance can be down to intra- or crossmodal neural unmasking, likely in combination. For example, in blind participants superior frequency discrimination and sound localisation in near and far space has been demonstrated (Doucet et al., 2005; Gougoux et al., 2004; Voss et al., 2004) with imaging showing a correlation between localisation ability and occipital cortex activation (Gougoux, Zatorre, Lassonde, Voss, & Lepore, 2005), implying crossmodal plasticity between visual and auditory areas. However, in a higher-order auditory task (complex versus non-complex sounds) fMRI showed increased activation in voice areas of the auditory cortex implying intramodal plasticity (Gougoux et al., 2009).

Contrary to the idea of a static brain with ‘hardwired’ sensory cortices, the evidence presented thus far illustrates a highly flexible plastic brain with structural, and functional connectivity between sensory areas. This supports an idea of a meta-modal organization of the brain, a theory formulated to explain results that show that brain areas usually associated

with visual stimuli maintain selectivity in the absence of visual input. In this, computations are posited to be based on function rather than being modality specific (Pascual-Leone & Hamilton, 2001) with cortical areas responsive to specific forms of stimuli. Furthermore these specific areas retain function even when visual input is missing. For example, the lateral occipital area is activated by visual and tactile information and codes for object shape irrespective of modality of input (Pietrini et al., 2004). While the LOC is non-responsive to auditory signals, we generally don't perceive shape via auditory information, it is engaged when the auditory signal is from a VA SSD (Amedi et al., 2007) further demonstrating the supramodal characteristics of the cortical region.

In the metamodal theory, local neural networks compete for sensory information that is functionally relevant to the region e.g. spatial information in the visual cortex or temporal information in the auditory cortex. Over time this region becomes the operator for that type of task. While the brain may appear to be organized by modality it may just be representative of the dominance of that modality for a specific task e.g. visual cortex for spatial tasks (Pasqualotto, Spiller, Jansari, & Proulx, 2013). We will further assess this regional specificity and the metamodal model of brain organisation in the section on perceptual learning. For a review of further evidence supporting a metamodal theory of the brain see (Ricciardi & Pietrini, 2011).

Interim summary.

In this section non-invasive methods of visual rehabilitation have been introduced. Evidence has shown that multisensory stimuli are integrated to form a coherent percept with natural crossmodal mappings for spatial and temporal dimensions. Imaging studies have shown extensive crossmodal plasticity from visual deprivation and rapid behavioural and plastic

changes in sighted populations. The evidence thus points towards a meta-modal organization of the brain based on function with task specificity for cortical areas

This alternative crossmodal model of a flexible, plastic, dynamic task-based machine is the framework that allows sensory substitution to work. The next section will look at how theory has been put into practice by introducing some of the landmark SSDs, explain how they work, and more pertinently, how effective they are.

1.4.4. Visual-to-tactile sensory substitution (VT).

In VT sensory substitution the tactile modality is used to substitute for the impaired visual modality i.e. characteristics of visual information are mapped to a tactile representation for stimulation. While the late American neuroscientist Paul Bach-y-Rita (1934-2006) is often considered the forefather of sensory substitution, with his landmark Tactile Vision Substitution System (TVSS), a 'substitution' system for the visually impaired pre-dated this by 140 years. It can be argued as to whether the tactile writing system Braille is a true example of sensory substitution as it is viewed in the modern day, however at a basic definitional level it performs an equitable function with research into Braille providing valuable information in crossmodal correspondences.

Braille

"we, the blind, are as indebted to Louis Braille as mankind is to Gutenberg".

Helen Keller

If classified as such, Braille is easily the most historically prevalent and arguably most successful sensory substitution system in that is has given millions of visually impaired people widespread access to recorded information, an analogue to the development of the printing press. The importance of this is exemplified by the above statement from deaf blind author, educator and activist Helen Keller and in statistics that show higher levels of employment in visually impaired Braille readers compared to non-readers (Goertz, van Lierop, Houkes, & Nijhuis, 2010).

The system developed in 1829 by Louis Braille as a simplification of Charles Barbier de La Serre's "Ecriture Nocturne" or night writing consisted of raised dots in a 2x3 polybius square each representing an alphanumeric character (Foulke, 1982)**.** Characters are read, using a single 'reading' fingertip in a left-to-right manner as in visual reading. Mean Braille reading speeds have been estimated at 100 words per minute with some reports of up to 225 words a minute (Crandell & Wallace, 1974) which compares favourably with that of sighted reading. The latter reading speed can be partly attributed to the use of contractions in Grade 2 Braille compared to the single character Grade 1 Braille.

While Braille has been undoubtedly a valuable system for both users and researchers it has been in decline since the advent of screen readers which convert visual text into synthesised speech output. Of course it is, and always has been, limited in what sensory information it can provide i.e. access to the written word. The findings from research into Braille however were important in the development of modern day VT SSDs.

Visual-to-tactile sensory substitution devices.

Inspired by the success of the Braille substitution system and increasing understanding of the crossmodal nature of perceptual processing the first ‘modern day’ SSD was developed in the late 1960’s by the American neuroscientist Paul Bach-y-Rita (Bach-y-Rita, 1968; Bach-y-Rita, Collins, Saunders, White, & Scadden, 1969; Bach-y-Rita, Collins, White, et al., 1969). The TVSS was an exemplar of an SSD in that it comprised of a three component hardware/software system to convert visual characteristics into a tactile counterpart. The basic components of the TVSS, and nearly all following devices, were: a sensor to extract visual information from the environment, a computer running an algorithm to convert information from the substituted modality into a format that can be delivered via another modality, and a transmitter to convey the converted information back to the user. Figure 1.2. shows the hardware setup of the TVSS (Bach-y-Rita, Collins, White, et al., 1969). The information collected by the video camera, converted in the computer is relayed to a 20x20 vibrotactile solenoid array set in the dentist style chair. Excitation of these solenoids, corresponding to light and dark pixels, elicits vibratory and pressure stimulation of mechanoreceptors in the dermis of the skin on the subjects back. This stimulation, when exceeding receptor thresholds, propagates action potentials to cortical areas subsequently building a ‘rough’ spatial crossmodal percept. As proof of concept the TVSS demonstrated the feasibility of VT sensory substitution but was limited in application.

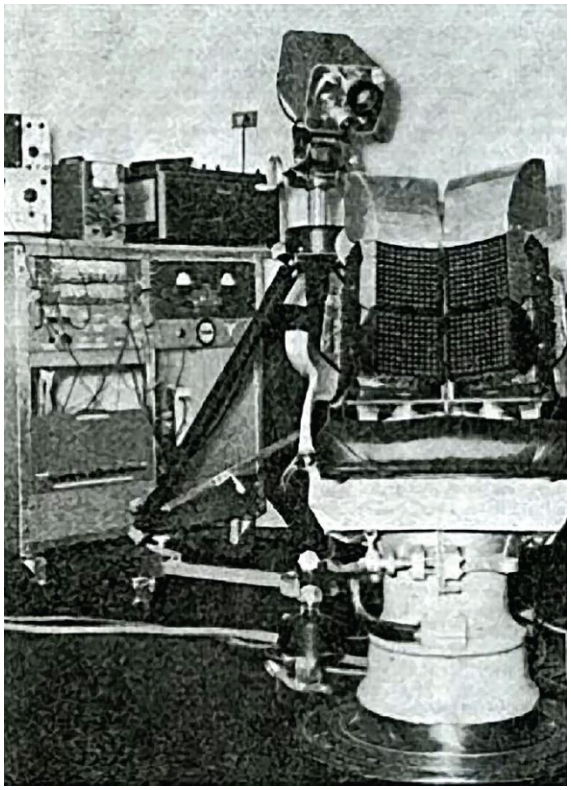

Figure 1.2: Paul Bach-y-Rita's original TVSS from (Bach-y-Rita, Collins, White, et al., 1969)

Primarily this was technological regarding the size of the components. As can be seen in Figure 1.2. the original TVSS was lab based only due to the choice of stimulation point. The low density, sparse distribution, and rapid adaptation of mechanoreceptors in the back, required a large array to convey simple information (Lenay, Gapenne, Hanneton, Marque, & Genouelle, 2003) thus impractical for everyday use. To account for this research in VT sensory substitution was twofold: miniaturization of components, and stimulation of areas with a high density of touch receptors.

With the largest somatosensory cortical representation being for the hands and around the mouth more recent incarnations of VT devices have utilised the fingers (Frisken-Gibson, Bach-y-Rita, Tompkins, & Webster, 1987) or tongue (Bach-y-Rita & Tyler, 2000) as stimulation points. The latter is especially practical as a contact point in the human-machine interface due to high mobility, protected situation and sensitivity (electrotactile stimulation of the tongue requires 3% of the voltage required to stimulate the finger). Saliva in the mouth also acts as an electrolytic solution providing a good electrical contact (Bach-y-Rita, Kaczmarek, Tyler, & Garcia-Lara, 1998).

The transmitter in VT SSDs that use the tongue as a contact point is referred to as the tongue display unit (TDU) as shown in Figure 1.3. for Wicab's Brainport. Converted visual information is transmitted from the PC via a ribbon cable to the 144 pin electrode array which delivers spatially relevant electrotactile information to the dorsum of the tongue. While receptor density of the tongue allows for smaller more portable arrays acuity is restricted by the direct spatial representation.

Figure 1.3: The hardware components of the Brainport TDU. The sensor is mounted central on the glasses with the transmitter as 20x20 array of vibro-tactile solenoids. Image taken from (Y. Danilov & Tyler, 2005)

1.4.5. Visual-to-auditory sensory substitution devices.

Why audition? There are both practical and neurological reasons for using audition rather than touch as the substituting modality. From a practical perspective the components of these devices are commonplace, low cost, energy economic, portable, and technologically advanced. The three component SSD can comprise of a webcam (sensor), netbook PC or smartphone (run the software) and stereo headphones (transmitter) all of which can be attained cheaply. For example, The vOICe system (see below) can be set up for about £250.

Using audition also negates the irritation at transmitter connecting points found with tactile devices and the natural hearing systems large frequency range is capable of processing large amounts of complex sensory information, a requisite when processing visual information.

Echolocation

VA SSDs can be categorized based on conversion principles. Echolocation devices use principles similar to that used by some microbats (Jones & Teeling, 2006; Simmons & Stein, 1980), odontocetes (Li, Wang, Wang, & Akamatsu, 2005), cave swiftlets (Griffin & Thompson, 1982)**,** shrews (Gould, Negus, & Novick, 1964)**,** and technological systems such as Sonar, echo sounding, and medical ultrasonography (Altes, 1976; Kane, Grassi, Sturrock, & Balint, 2004; Knott & Hersey, 1958). A frequency modulated (FM) signal is transmitted from the device and telemetry used to ascertain distance and target object shape from the temporal and intensity characteristics of the returning signal. For example, pitch can be mapped to object distance and horizontal localisation to inter-aural disparity as found in the Sonic Glasses or Sonic Pathfinder (Heyes, 1984; Kay, 1964, 1985). While effectiveness of echolocation devices has been demonstrated for wayfinding and provision of spatial information in three-dimensional scenes (Hughes, 2001) the majority of studies have employed behavioural and psychophysical paradigms and given little consideration to the neural correlates of echolocation.

Intriguingly echolocation is also found in humans. Self-taught CB echolocator Daniel Kish uses his tongue to emit short, spectrally broad ‘palate click’ signals, analogous to the FM signal in Sonar. Kish’s technique is both effective in accomplishing tasks such as bike riding, navigation, and ball games without traditional aids, and also transferrable in that Kish has taught numerous sighted and blind people to use the technique. fMRI studies have evaluated the neural correlates of echolocation in Kish and other echolocators. Exposure to recordings of echolocated clicks within natural background noise elicited blood-oxygen level dependent (BOLD) activity in primary auditory cortex in all participants. In echolocators, but not controls, robust activation was also found in the calcarine complex, however, this V1 activity

was not apparent when the echoes were ‘scrubbed’ from the soundscape implying that the presence of low-amplitude echoes facilitated visual cortex activation in blind participants (Thaler, Arnott, & Goodale, 2011). Intriguingly, V1 activation in a CB participant demonstrated a contralateral preference similar to what is found in light related (visual) activity. Experts and naïve users have demonstrated discrimination of materials using echolocation with activation in parahippocampal cortex in sighted and blind echolocators, although not in comparable controls (Milne, Arnott, Kish, Goodale, & Thaler, 2014) with head movements appearing to facilitate superior performance in 2-D object recognition (Milne, Goodale, & Thaler, 2014). The ability to echolocate has been shown to correlate with levels of visual imagery implying that visual imagery is a strategy for echolocation or that both tasks, not reliant on retinal input, depend on recruitment of calcarine cortex (Thaler, Wilson, & Gee, 2014).

The relative success of natural echolocation is one explanation for the lack of recent research in echolocation based SSDs. For example, when contrasting a number of different vocalised ‘clicks’ with the particular palette click employed by Kish, the human click gave the most detailed feedback of the immediate surroundings (Rojas, Hermosilla, Montero, & Espí, 2009). All participants acquired basic echolocation skills within a few days of training, with performance in the blind participants’ best, possibly due not having to overcome visual bias. This supports the idea that, at present, visual restoration using echolocation is more user and cost effective using ‘natural’ techniques rather than technology.

Image-to-sound sensory substitution.

Non-telemetrical VA SSDs use direct mapping of visual to auditory characteristics to facilitate image-to-sound conversion. Typically a captured image is segmented into an array of pixels, dependent on the particular device’s acuity/resolution, and each pixel subjected to

the conversion principles. The summation of pixel auditory output, again dependent on device, creates the image soundscape. Processing is carried out in ‘real time’ although by definition this is broad as, for example, ‘real time’ using The vOICe at default (P. Meijer, 1992) settings refers to a series of one second duration snapshots presented simultaneously, variable as the sensor moves across the scene.

There have been a number of visual-to-auditory SSDs developed since the early 1990’s of which two have dominated the research literature. The device utilised in the studies in this thesis, The vOICe – the middle 3 letters spell out ‘Oh I See’, is a hardware/software device that uses non-specialized hardware (webcam, PC or smartphone, stereo headphones) combined with a freely available software algorithm to convert visual to auditory information. Visual sensory information is extracted from the environment by a sensor (webcam), and compartmentalised into 4076 greyscale pixels by the software. The three conversion principles are applied to each of these pixels before being summated to give the final soundscape.

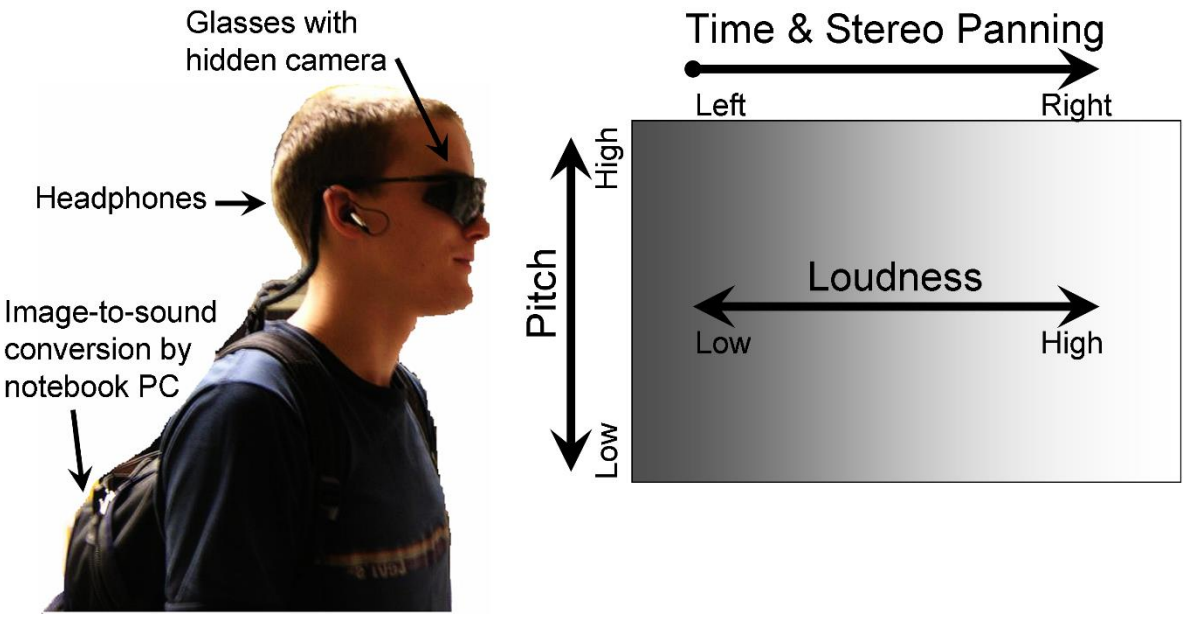

Figure 1.4: Hardware setup (left) and diagram of conversion algorithm (right) for The vOICe visual-to-auditory sensory substitution device. Image taken from (Proulx, Stoerig, Ludowig, & Knoll, 2008b)

The three conversion principles provide three pieces of information, of which two are spatial. Pixel brightness is mapped to auditory volume with white pixels loud and black silent, with 16 degradations of grey/volume between. Pixel spatial position require two coordinates (x & y) cross referenced for location. Vertical location (y) is coded to sine wave frequency. Pixels high up in the cameras visual field produce a high pitched sine wave, relative to lower pixels within the 500Hz- 5000Hz frequency range of the device. Both this and the luminance mappings are informed by the natural crossmodal correspondences described in *Section 1.4.2.* Horizontal pixel location (x) is mapped to both a temporal and stereo scan position. The device scans left-to-right across the image in a set time frame (1 second at default). Pixels to the left of the image are heard early in the soundscape relative to ones to the right. If using stereo headphones left-hand side pixels will be head to the left in the stereo field (in the left headphone) with pixels further to the right being heard later and to the right. The device can be used in mono (1 headphone only) with x position relying only on temporal information - this is assessed in Chapter 2 with evaluation of performance differences comparing mono and stereo device output. All sonified pixels in a column are played concurrently with each of these raster lines presented simultaneously (from right to left) over the duration of the scan. The resultant soundscape is the sonification of the particular frame. As the sensor is moved across the visual scene the soundscape updates in real time, or 1 second blocks/frames dependent on the visual characteristics of the scene.

While the frequency/elevation mapping and left-to-right scan direction is fixed, toggles in the software allow for manipulation of the other conversion principles. For example, contrast can be reversed so that black pixels are coded to maximum volume and white to silence, while x-axis scan rate can doubled or reduced by magnitudes of x2, x4, x8. A final within algorithm

manipulation is the edge enhancement setting. Here a Sobel operator is used to detect changes in luminance. Object edges then appear brighter (louder) to distinguish them from the solid 'fill' pixels, although The vOICe blends in a low intensity version of the original image to retain some of the original surface shading. Manipulations can also be carried out using the hardware components. As already mentioned the device can be used in mono by inserting only one headphone, an advantage if requiring ambient noise from the environment, and sensor position is also variable. Brown and colleagues (2011) evaluated various device settings in 'real time/device' object localisation/recognition paradigm showing that device setting advantages were task dependent. For example, sensor position was salient in that a head mounted sensor facilitated superior object localisation but inferior recognition while reverse contrast outperformed normal contrast although again task dependent. In concluding the authors posited that at the early stages of learning a reduction in the signal-to-noise ratio and understanding of the auditory characteristics was as important as the crossmodality in device use (D. J. Brown, T. Macpherson, & J. Ward, 2011). This idea of the importance of auditory characteristics is a focus of this thesis. In Chapters 3 & 4 I extend these findings by assessing the magnitude of information required to successfully recognise objects from their soundscapes and also evaluating in training phases simulated versions of device settings (time scan and edge enhancement).

The smartphone version of The vOICe incorporates a number of features not found on the PC version to enhance the experience. A basic colour recognition feature gives vocalised colour information and interlinking with Google Googles provides basic object recognition, again vocalised. Navigation in outdoor environments is enhanced by links to Google Maps via GPS to give both directional and map information.

Alongside The vOICe, The Prosthesis for Substitution of Vision by Audition (PSVA) developed by Capelle and colleagues in 1998, is the most prevalent device in VA SSD research. The PSVA camera captures the image and represents if on a two-resolution artificial retina. The global image is a 8x8 pixel matrix with each of the four central pixels of the matrix are replaced by a 'foveal' area pixel consisting of a 8x8 smaller pixels. This is analogous to the high resolution fovea in the eye. The resultant visual area comprises of a 60 pixel periphery and a 64 pixel central fovea. Each pixel is then subjected to conversion principles to create the soundscape. Like The vOICe pixel luminance is mapped to volume. Pitch is also used as a representation of spatial position, however in the PSVA algorithm sinusoidal tone frequency is associated with horizontal, rather than vertical, spatial position with a slow pitch increase from left-to-right. Unlike The vOICe use of the PSVA is active in that the user must manually scan the image with the camera contrasted to the passive self-scanning vOICe (Capelle et al., 1998).

While both The vOICe and PSVA process images into greyscale a relatively new VA SSD, EyeMusic, has introduced a mapping for basic colour into the algorithm (Abboud, Hanassy, Levy-Tzedek, Maidenbaum, & Amedi, 2014). Images are recorded via a camera and resized to a 40 (x) x 24 (y) pixel image. A colour-clustering algorithm strips the image down to a six colour representation. The conversion to an auditory signal is representative of The vOICe algorithm in that pixel luminance is mapped to auditory volume and x-axis position coded to a left-to-right time scan. y axis pixel position is again mapped to pitch but with two major differences. The frequency range is restricted to a ceiling of 1568hz (compared to 5000hz for The vOICe) to restrict the unpleasantness of higher pitched sounds in the soundscape (Kumar, Forster, Bailey, & Griffiths, 2008; T. Wright & Ward, 2013). Secondly, while the frequency/elevation mapping of The vOICe uses pure sine waves in its coding the EyeMusic utilises musical notes on a pentatonic scale to represent elevation. This should further

increase the pleasantness of the soundscape by reducing auditory dissonance – in Chapter 4 I assess dissonance and, more saliently, consonance as a potential negating factor is using The vOICe. Colour in the EyeMusic algorithm is represented by 5 different musical instruments: Red-Reggae Organ, Green – Rapmans Reed, White – Choir, Blue – Brass Instruments, Yellow – String Instruments. Black is again mapped to silence. Using a combination of these principles therefore gives you the xy position of the pixel (time scan x musical note), luminance of the pixel (volume), and the colour (musical instrument).

While the utilisation of colour can primarily be thought of as serving aesthetic purposes in the sighted population, and thus of limited use in the visually impaired, this negates the importance of colour as a factor in object discrimination within scenes (Goffaux et al., 2005; Yip & Sinha, 2002). For example, successful discrimination of one object from a number of others of similar dimensions is increased if the pop out object is of a different colour.

While the three VA devices described above are presently being used in various research paradigms a number of other devices have been developed and tested in the past. SmartSight, which employed similar conversion principles to The vOICe and PVSA but was limited with regards to its frequency range (Cronly-Dillon, Persaud, & Gregory, 1999), The VIBE, an open source Sourceforge hosted project, which generates sinusoidal sounds from virtual sources, with uniformly receptive fields coding for loudness and stereo/binaural panning and frequency coding for horizontal and vertical spatial position respectively. An advantage of The VIBE visual-to-sound algorithm is that all aspects are configurable by the user (Auvray, Hanneton, Lenay, & O'Regan, 2005; Hanneton, Auvray, & Durette, 2010). The Kromophone was the first SSD to try and represent colour in its output. Information recorded from the sensor was averaged around the pixel in centre of visual field and then pixel colour mapped to soundscapes. Representations of different colours are mapped to spatial positions: red –

high pitch/right ear, blue – low pitch/left ear with green – middle pitch/both ears. Other colours are presented as a combination of sounds relative to mixing colours on a palette (Capalbo & Glenney, 2009). Colour representation is also a goal of another recent device See colOr (J. D. Gomez, Bologna, & Pun, 2014).

1.4.6. How effective are visual-to-auditory sensory substitution devices?

Effectiveness of SSDs must be assessed within context. Are we trying to directly replicate the visual experience, or to provide methods to function in tasks that are predominantly modulated by vision, or view these as not mutually exclusive? Regarding the former we can test the effectiveness of SSDs using standard ophthalmological tests such as the Snellen tumbling E. This measure of acuity has been demonstrated in a number of SSDs with results adapted for the resolution and visual field width of the specific device. Acuity using The vOICe in both trained (100 hours) and naïve users have demonstrated levels of up to 20/200 and 20/408 respectively, approaching the minimal legal definition of blindness (Haigh, Brown, Meijer, & Proulx, 2013; Striem-Amit, Guendelman, & Amedi, 2012). This compares favourably with acuity levels found in VT SSDs (Chebat, Rainville, Kupers, & Ptito, 2007; Chuang, Margo, & Greenberg, 2014; Sampaio, Maris, & Bach-y-Rita, 2001) and is significantly better than found in invasive techniques (Fernandes, Diniz, Ribeiro, & Humayun, 2012). These illustrate that sensory substitution can certainly increase acuity but how does this translate to other tasks normally modulated by vision?

Numerous studies using the PVSA has demonstrated successful form and pattern recognition in both early blind and sighted users (Arno, Capelle, Wanet-Defalque, Catalan-Ahumada, & Veraart, 1999; Arno, Vanlierde, et al., 2001; Collignon, Lassonde, Lepore, Bastien, & Veraart, 2007; Poirier, De Volder, Tranduy, & Scheiber, 2007; Poirier, De Volder, Tranduy,

& Scheiber, 2006) with similar success found in users of EyeMusic (Abboud et al., 2014), See ColOr (J. D. Gomez et al., 2014) and VT devices using tactile solenoids on the back (Bach-y-Rita, Collins, Saunders, et al., 1969; Bach-y-Rita, Collins, White, et al., 1969; Bach-y-Rita et al., 1998; Sampaio et al., 2001)and tongue (Kaczmarek, 2011; Nau, Pintar, Arnoldussen, & Fisher, 2015; Vincent, Tang, Zhu, & Ro, 2014). Curiously, considering SSDs are representing two dimensional scenes, depth perception and 3D trajectory has been demonstrated in both VA (Renier, Collignon, et al., 2005; Renier & De Volder, 2010) and VT devices (Bach-y-Rita, Collins, White, et al., 1969; Chekhchoukh, Goumidi, Vuillerme, Payan, & Glade, 2013; Epstein, 1985; Epstein, Hughes, Schneider, & Bach-y-Rita, 1986). Motion detection, including motion parallax, is available to SSD users of both formats (Bach-y-Rita, Collins, Saunders, et al., 1969; Bach-y-Rita, Collins, White, et al., 1969; Bach-y-Rita & Kercel, 2003; Poirier, Collignon, et al., 2006) and interestingly 'visual' illusions are effective in VA sensory substitution. Renier, using the PSVA, found both the vertical-horizontal and Ponzo illusions recreated in the crossmodal system (Renier, Bruyer, & De Volder, 2006; Renier, Laloyaux, et al., 2005) although only for BS and not EB subjects. The implications being that while the SSD provides enough low-level information to recreate the illusion an understanding of visual percepts such as size-distance (Ponzo) is a requisite to elicit the effect.

The true value of an SSD however is in how it can be applied to facilitate everyday functioning. Provision of information and data to the visually impaired is generally in the form of Braille and screen readers but has also been demonstrated in VA devices including the PSVA and The vOICe (Bliss, Katcher, Rogers, & Shepard, 1970; Craig, 1981, 1983; Loomis, 1974, 1981a, 1981b; Reich, Szwed, Cohen, & Amedi, 2011; Striem-Amit, Cohen, Dehaene, & Amedi, 2012).

The ‘what’ and ‘where’ dual process of vision can be evaluated using recognition and localisation paradigms. Object localisation and recognition has been shown to be effective in both VT devices (Williams, Ray, Griffith, & De l’Aune, 2011) and VA devices such as EyeMusic (Levy-Tzedek, Hanassy, Abboud, Maidenbaum, & Amedi, 2012), The VIBE (Hanneton et al., 2010), and The vOICe whether trained (Amedi et al., 2007; Auvray, Hanneton, & O'Regan, 2007; Merabet et al., 2009), in naïve users given an explanation of the algorithm (Proulx, Stoerig, Ludowig, & Knoll, 2008) and even in BS users who were given no information aside from that the device uses an algorithm to convert visual features to sounds (D. J. Brown et al., 2011; Kim & Zatorre, 2008).

Navigation is another important facet in everyday life relying heavily on vision. Traditional navigation aids such as the white cane and guide dog facilitate object avoidance and detection respectively. SSDs have been used to evaluate spatial perception and navigation in both real world and virtual environments. Devices such as the EyeCane have been used to extend the range of the white cane to up to 5 metres using sonifications of depth information (Maidenbaum et al., 2012) while successful navigation and virtual route recognition has been demonstrated in the VA device The VIBE (Durette, Louveton, Alleysson, & Hérault, 2008)(and VT TDU’s (Chebat, Schneider, Kupers, & Ptito, 2011; Kupers, Chebat, Madsen, Paulson, & Ptito, 2010) Considering devices such as The vOICe and EyeMusic now have algorithms that run on smartphones, with GPS and mapping capabilities integrated, the potential for SSDs to provide both global (route) and local (obstacle avoidance) navigation aids is an exciting area of research offering scope for sonified and vocalised maps.

Outside visual impairment SSDs have been used to aid in a number of disorders such as the sexual rehabilitation of men with chronic spine injuries (Borisoff, Elliott, Hocaloski, & Birch, 2010), and vestibular rehabilitation (Danilov, Tyler, Skinner, & Bach-y-Rita, 2006; Tyler,

Danilov, & Bach, 2003; Uneri & Polat, 2009)and recently a lower leg mounted device to provide distal information when walking (Lobo, Travieso, Barrientos, & Jacobs, 2014).

1.4.7. Neural correlates of sensory substitution.

As shown in the behavioural literature, SSDs can be effective 'tools' for accomplishing 'visual' tasks. Considering the basis of sensory substitution, imaging studies should be expected to show activation in 'visual' area of the cortex.. In a task to identify geometric shapes by touch, 7 blind participants who reported visual qualia showed activation in occipital cortex, not found in sighted controls (Ortiz et al., 2011). Localisation tasks using The vOICe showed bilateral activation of V3 (BA19) for a visual dot in left field and left V3 activation when dot was to right (Stiles, Chib, & Shimojo, 2012) with other studies using this device demonstrating activation in left pre-central sulcus, bilateral lateral occipital cortex, occipital parietal and posterior occipital cortices (Striem-Amit, Dakwar, Reich, & Amedi, 2012). Amedi's seminal shape recognition study, for example, showed vOICe activation in lateral occipital cortex with TMS to this area disrupting the task (Amedi et al., 2007; Merabet et al., 2009). In another vOICe study, EEG showed early interactions between visual and auditory cortex in a shape matching task (Graulty, Papaioannou, Bauer, Pitts, & Canseco-Gonzalez, 2014) while activation of the visual word form area is found in trained blind users (Striem-Amit, Cohen, et al., 2012) and crossmodal activation shown in visual cortex in sighted users of the PSVA (Poirier et al., 2007). In the tactile modality, using TDU's in short shape discrimination training in EB activated occipital cortex (Ptito et al., 2005).

While the above demonstrates the recruitment of visual areas, primarily occipital cortex, in both VA and VT sensory substitution there is still robust activation in associated unimodal areas. For example, in echolocation strong activation is found in auditory cortex (Thaler et al., 2011; Thaler, Milne, Arnott, Kish, & Goodale, 2014) while various tasks using VA SSDs

demonstrate high peaks of activity in primary auditory cortex (Hertz & Amedi, 2014; Kim & Zatorre, 2011; Ortiz et al., 2011; Striem-Amit & Amedi, 2014), A1 and Heschl's gyrus, (Striem-Amit, Cohen, et al., 2012). Similarly in VT substitution activation is shown in somatosensory cortex, parietal and pre-frontal motor cortices, bilateral and dorsal premotor cortex (Gagnon et al., 2012; Kupers et al., 2010; Poirier, De Volder, et al., 2006; Ptito et al., 2005). Interestingly, for this thesis, studies that report unimodal activation state it is generally before -training with the crossmodal activation apparent only after a period of learning. This implies that in naïve users the input signal is processed in unimodal areas and subjected to potential limitations of the unimodal system. This was hypothesized in a previous study on object recognition and localisation as a function of SSD device settings (D. J. Brown et al., 2011). As this thesis is concerned with naïve users there is a definite logic to applying unimodal rules to signal processing. This is assessed in Chapters 3 and 4 and touched on in Chapter 2 when auditory object formation is evaluated in both the resolution of objects and potential frequency based confounds.

1.4.8. What is the phenomenological experience?

While the literature indicates recruitment of the occipital cortex in SSD use are users actually seeing? Is the phenomenological experience in the substituted (visual) or substituting (audition, tactile) modality? Dominance theory proposes the latter in that there is no perceptual modality change while deference theory posits the opposite, that the perception is the substituted modality (Hurley & Noë, 2003).

Naturally we must first look at the qualitative experience. To which modality do users assign the subjective experience to? Ward and Meijer (Ward & Meijer, 2010b) looked at the qualitative experience of two long term users of The vOICe , PF and CC.

“Just sound?… No, it is by far more, it is sight! There IS true light perception generated by The vOICe.” From interview with PF in Ward and Meijer (2010).

Both users describe the perception of ‘visual’ experience such as depth, motion, and with PF colours. Acuity and fine detail is described as poor, similar to 3D line drawings, and especially for modern unfamiliar objects. This last point is salient. As both PF and CC are late blind this infers prior visual experience a requisite to add fine detail to the percept i.e. PF can’t perceive fine details about her computer as the device was invented after she lost her sight but can ‘add’ colour to the perception of familiar objects even though The vOICe doesn’t code colour in the algorithm.

Interestingly Ward et al report that some congenitally blind also describe SSD use as ‘visual’ rather than auditory or tactile (Ward & Meijer, 2010b). If they are ‘seeing’ then it follows that the perception must be experientially different to both a sighted persons definition off seeing and also the blind SSD users ‘normal’ sensory experiences e.g. sound, touch, taste, smell etc. The perceptual experience is therefore novel or amodal. This amodality is corroborated in other self-report. Trained vOICe users reported a ‘visual’ or amodal feel for localization tasks and auditory for recognition, emphasizing that whilst able to perceive visual qualities they did not have a true visual or auditory experience (Auvray et al., 2007). The experience is ‘something different’. How we best qualify the perceptual experience of sensory substitution is then constrained on how we define what constitutes a sensory modality.

Is it defined by the sensory organ used, properties of the stimulus, behavioral response, qualitative experience, sensory motor equivalence? (Grice, 1962; Morgan, 1977; O'Regan & Noe, 2001). If considering sensory organ used the energy format of the extracted information then SSD use is in the substituting modality. However, behaviorally, users of the TVSS show avoidance behaviour when an object is moved rapidly towards the camera lens (Bach-y-Rita,

2002), in much the same way as a sighted person would to an object moving towards their eye, implying visual characteristics. The perception may also change depending on task, especially the sensory motor invariants specific to the modalities. For example, as we move towards an object it expands on the retina and also increases in aural intensity. O'Regan et al posited that the more perception with an SSD shares sensorimotor invariants to the substituted modality the more it resembles that modality(O'Regan, Myin, & Noe, 2005). That SSDs have to be actively manipulated by the user appears to confirm sensorimotor equivalences effect on perception. Auvray (2007) suggested that active manipulation of the sensor establishes links between action and sensorimotor changes in stimulation and these links are pre-conditional in perception (Auvray et al., 2007).

Considering the lack of concrete evidence for dominance or deference theories, Auvray (Auvray & Myin, 2009), suggested that the elicited perception is in fact amodal, as per qualitative report and that SSDs should be seen as 'mind enhancing tools'(A. Clark, 2004) that extend perception in novel ways rather than substitution systems.

Interim Summary.

In the previous section a number of SSD's have been described in both the technological components and behavioural results. Imaging studies show recruitment of the occipital cortex after training on these devices and the phenomenological experience has been briefly touched on. It is interesting to note that in many of the cited studies users can function way above chance levels with minimal or even no training, emphasising the inherent understanding of cross-modal correspondences. The often rapid improvement over short durations of training however emphasises the importance of training and perceptual learning.

1.5 Perceptual learning.

This section is an adaptation of Proulx, Brown, Pasqualotto, and Meijer (2014) Multisensory Perceptual Learning and Sensory Substitution in *Neuroscience and Biobehavioral Reviews.*

"The key to ultimate success is the determination to progress day by day"

Edmar Mednis.

The quote from the American International Grandmaster of Chess, Edmar Mednis, although about chess, illustrates the importance of determined learning. In chess the basic rules and movements of each piece are simple and yet the game requires dedication, persistence and thorough analysis to master or play at a high level. This can be viewed as analogous to sensory substitution. Instinctive understanding of cross-modal correspondences facilitate encouraging results in naïve device users in simple non-visual tasks. Short durations of training increase levels of task performance until a performance threshold is reached, either through increased task complexity or conversely when learning on the task has rendered it too easy. However, mastery of the device requires analysis of mappings and the dedication to learn what is a difficult endeavour. The structure of training is therefore vital in providing non-redundant information and a manageable and rewarding experience. The research described next looks at perceptual learning in the substituted and substituting modalities of VA sensory substitution prior to looking how this impacts on multisensory learning.

Perceptual learning can be classed as specific, in which improvement on a task is restricted to the specific stimuli used in training. However, for learning to be optimal it needs to generalise to alternate stimuli not encountered in training. The breadth of perceptual generalization informs both the structure and efficacy of training paradigms. If the range of

generalization is narrow then this needs to be reflected in training using a limited set of alternate stimuli. As this range broadens the variation and number of generalizable stimuli utilised in training can be increased to increase the pace of learning.

1.5.1 Visual perceptual learning.

Research in visual perceptual learning has shown that training on a task restricts learning to the specific stimuli used in training, implying neuroplastic changes at low cortical levels. This psycho-anatomy approach indicates the specificity of improved performance is at low levels as neurons at that level have field properties for learning the particular visual features (Stromeyer & Julesz, 1972). Improved performance, i.e. effective learning, due to training has been demonstrated over a number of stimulus features such as vernier acuity (Fahle, Edelman, & Poggio, 1995; McKee & Westheimer, 1978; Saarinen & Levi, 1995), motion (K. Ball & Sekuler, 1987), orientation and texture (Karni & Sagi, 1991), and spatial frequency (Fiorentini & Berardi, 1980, 1981). Spatially, learning can be specific in a non-transfer across visual fields or feature specific in which one orientation does not generalize to another (Schoups, Vogels, & Orban, 1995). However research in visual generalization seemingly contradicts these results. For example, contrary to the specific learning found by Karni et al (1991) texture is found to generalize across eyes (Schoups et al., 1995). Theoretical explanations for these inconsistent results can be found in the Reverse Hierarchy Theory (RHT) of visual perceptual learning (Ahissar & Hochstein, 2004). This theory posits that the difficulty and characteristics of the task influence the cortical level in which learning is represented. Primary to the theory is that difficult tasks, where more specific discrimination is required, focuses attentional resources to primary sensory areas such as V1 with associated small receptive fields. Conversely less complex tasks, in which the utilisation of more general

features facilitates successful discrimination, drives attention to higher cortical association areas such as the intraparietal area signified buy large receptive fields. As theorised, perceptual learning can occur at all levels of processing: firstly high-level areas would be recruited with feedback to low-level sensory areas only if required for discrimination. The level of processing therefore influences the type of learning that can occur. If processing remains at low-levels then the perceptual learning will be specific to the features and spatial arrangement of the trained stimulus. Generalization to spatial feature information in other stimuli will only be apparent for high-levels of processing (Proulx, Brown, Pasqualotto, & Meijer, 2014).

The literature has confirmed the role of high-level cortical areas in perceptual learning. In a double training task in which one retinal location was exposed to a relevant task and another to an irrelevant task a transfer of learning to the irrelevant retinal location was shown implying high-order, non-retinotopic brain areas involved in the promotion of location generalization. The RHT, and associated use of feedback loops has been applied to both audition (P. C. M. Wong, Skoe, Russo, Dees, & Kraus, 2007) and as I illustrate in chapter 2 auditory/multisensory paradigms using output from a VA SSD.

1.5.2 Auditory perceptual learning.

Audition is the substituting modality in The vOICe and thus the research on learning in this domain is particularly important as it allows for comparison for performance in 'auditory' SSD paradigms. Furthermore, the understanding of what drives generalization in audition should aid with learning through training with a VA SSD.

A prototypical auditory perceptual learning task is that of temporal discrimination. In this the requirement is for the listener to choose a reference tone (generally the shorter one) from two

tones played successively. The difference in temporal duration of the comparison tone is varied, based on correct responses, until a perceptual threshold is reached in which the listener cannot recognise the reference tone above chance levels. Using this threshold as a dependent variable, repeated training on this discrimination task should lower the threshold indicating perceptual learning. As all other features (frequency, volume) of the stimuli are kept consistent the improvement in learning is specific to the temporal features. Furthermore, using multiple training sessions over a number of days allows plotting of a time course illustrating the speed of learning. Paradigms such as these in the auditory domain have demonstrated that training facilitates a rapid improvement on discrimination to the specific stimulus (B. A. Wright, Buonomano, Mahncke, & Merzenich, 1997; B. A. Wright & Fitzgerald, 2005) within 100's of trials. This paradigm however also allows us to test for generalized learning by provision of pre- and post-test stimuli that vary in spectral or temporal features to the trained stimuli, for example a different frequency or duration. If training on the specific stimuli facilitates an increase in performance at post-test for the alternate stimuli then the learning has generalised. Generalization to spectral features, especially frequency, has been demonstrated in a number of studies (B. A. Wright & Fitzgerald, 2005; B. A. Wright, Wilson, & Sabin, 2010) however this generalization is not usually found for temporal features, although one study by Lapid did imply generalization from an interval to an untrained duration, i.e. a filled interval (Lapid, Ulrich, & Rammsayer, 2009).

Contrasted to the lack of generalization for temporal discrimination, learning on frequency discrimination has shown partial generalization on untrained stimulus durations and across ears (Delhommeau, Micheyl, Jouvent, & Collet, 2002) with Micheyl (Micheyl, Bernstein, & Oxenham, 2006; Micheyl, Delhommeau, Perrot, & Oxenham, 2006) demonstrating full generalization from the trained to untrained ear. Further evidence shows that frequency

discrimination generalizes across conditions in which the pure tone frequency is fixed or ‘roves’ across frequencies with both wide and narrow frequency bands generalizing to the fixed frequency, and from the fixed to the narrow in poor listeners (Amitay, Hawkey, & Moore, 2005).

For frequency and amplitude, listeners trained on pure tones generalized to complex tones containing harmonics of the fundamental frequency that could be resolved by the peripheral auditory system but not to tones with unresolved harmonics. Furthermore as there was no generalization to noise bands modulated at the fundamental frequency the implications being that the auditory system uses two processes to encode pitch dependent on the resolution of the harmonics (Demany & Semal, 2002; Grimault, Micheyl, Carlyon, Bacon, & Collet, 2003). This is evaluated in Chapter 4 where the effect of harmonic relations on object formation is evaluated.

Amitay illustrated an intriguing result in frequency discrimination in that improvement was found in threshold differentials of 0hz and thus the trained stimuli were perceptually impossible to discriminate. Results for these ‘impossible’ stimuli compared positively to easy (400Hz) and difficult (7Hz) with the authors positing that training may improve the ability to attend to low-level, task specific, representations of the stimulus, rather than adaptation of the comparison mechanism (Amitay et al., 2005). Micheyl offered further explanation for these counterintuitive results suggesting that the random variability of neural responses to auditory stimuli render the identical stimuli as qualitatively different enough to fine tune the comparison mechanism required for learning (Micheyl, Bernstein, et al., 2006; Micheyl, Delhommeau, et al., 2006).

While the above studies imply rapid learning to the trained stimuli and generalization primarily to spectral but not temporal features a critical point is that most studies, particularly

in temporal discrimination, utilise simple unimodal stimuli (i.e. pure tones). In chapter 2 I evaluate whether the simplicity of the stimuli hinders the breadth of generalized learning, by using complex soundscapes from a sonified image in a temporal interval discrimination task.

1.5.3. Multisensory perceptual learning.

Is there generalized and specific multisensory learning? Shams and Seitz (2008), proposed that multisensory learning is more ecologically valid than unisensory models as we develop in multisensory environments and thus learning is likely to reflect this (Shams & Seitz, 2008).

In speech perception, phonetic perpetual learning is found to be not only specific but generalizes in both infants and adults (Hervais-Adelman, Davis, Johnsrude, Taylor, & Carlyon, 2011; Kraljic & Samuel, 2006; Maye, Weiss, & Aslin, 2008)with interestingly, sleep promoting generalization in phonological categories (Fenn, Gallo, Margoliash, Roediger, & Nusbaum, 2009). Nagarajan , in a somatosensory interval discrimination task found that whilst discrimination was temporally specific generalization was found intramodally to other skin locations and crossmodally to the auditory modality (Nagarajan, Blake, Wright, Byl, & Merzenich, 1998). In another tactile discrimination task Planetta et al (2008) found generalization to motor interval production with the same temporal features, and Bartolo et al (2009) found generalization to vision in an auditory interval reproduction task (Bartolo & Merchant, 2009; Planetta & Servos, 2008). There is not generalization in all multisensory learning however. For example, Lapid (2009) when investigating whether training on auditory durations generalized to visual intervals found no crossmodal learning (Lapid et al., 2009). Crossmodal generalization has been shown in paradigms using SSDs. Kim (2008) using a VA SSD showed rapid generalization to novel stimuli with the same authors later showing auditory generalization of shape and abstraction to an untrained modality (Kim & Zatorre, 2008, 2010, 2011).

Are there candidate areas of the brain where this cross-modal generalization occurs? Firstly it appears crucial that for crossmodal generalization the stimuli must share some spatiotemporal features such as duration or location (Bartolo & Merchant, 2009), indeed spatiotemporal congruency appears to be at the core of multisensory integration and processing (Lewald, Ehrenstein, & Guski, 2001; Macaluso & Driver, 2005). Providing there is this form of implicit association even task irrelevant paradigms will facilitate cross-modal generalization (Seitz & Watanabe, 2009). Secondly, the features that are more salient to a task are likely to generalize across modalities (Jain, Fuller, & Backus, 2010).

The metamodal model of brain organization explains the neural basis of information process, while perceptual learning theories explain the cognitive basis in information processing. In Proulx et al (2014) the authors propose a unified model to provide a more explanatory basis for multisensory perceptual learning, with emphasis on processing required for the task rather than source of input. Figure 1.5a depicts this model for unisensory and multisensory perceptual learning and 1.5b how this is influenced by visual deprivation. There are two factors key to this model. Firstly, brain areas for specific modalities are functionally optimal for particular computations; auditory areas for temporal features/tasks and visual for spatial. This is posited in the metamodal theory. Secondly, multisensory stimuli are complex in that they provide a richer source of information (Proulx et al., 2014).

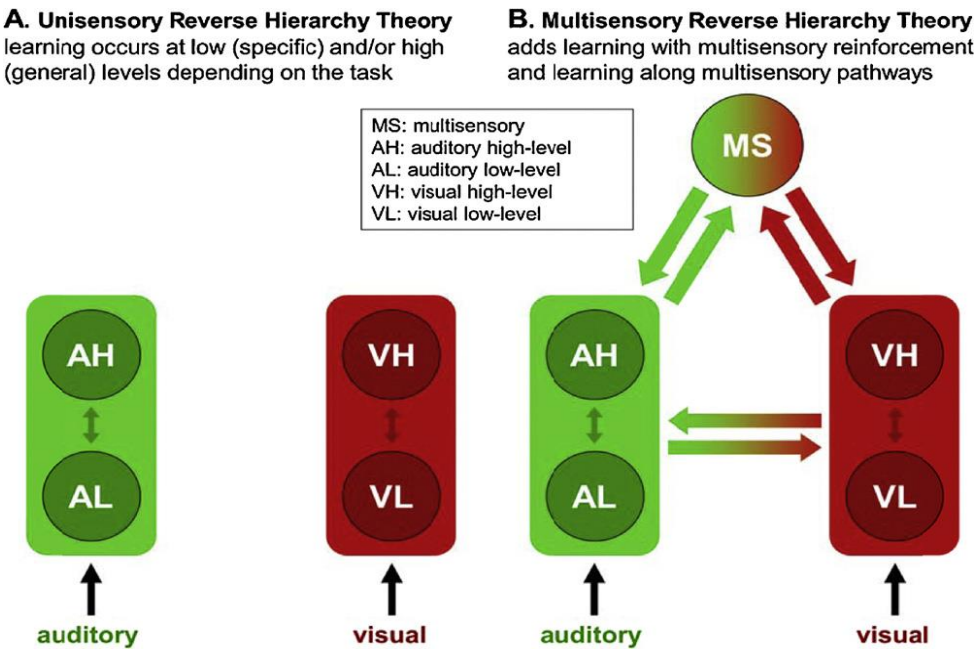

Figure 1.5a: Figure depicting a unisensory and a multisensory reverse hierarchy theory of perceptual learning. Taken from (Proulx, Brown, Pasqualotto, & Meijer, 2014)

In Figure 1.5a, Unisensory learning is modality specific, auditory (green) or visual (red), with low-level primary cortical areas for specific learning and high-level association areas for generalization as specified in the RHT (Ahissar & Hochstein, 2004). The mechanism for multisensory learning is shown for two separate conditions: firstly, learning under multisensory stimulation can result in activity in higher-level multisensory associative areas as posited by Sham and colleagues (Shams & Seitz, 2008) : secondly, learning can progress from low-level primary sensory areas to higher-level multisensory areas under complex stimulation, as multisensory tasks are less likely to be defined by simple low-level features. Activity may then cascade back down the hierarchy allowing generalization across modalities if the high-level multisensory areas are implicated in learning multisensory or unisensory tasks. The impact of blindness is shown in 5b by removing the visual input (red). In the metamodal model the 'visual' cortex is responsible for spatial processing. Therefore if presented with a task that would require spatial processing, usually attributed to vision, auditory stimulation (green) can activate and induce perceptual learning in 'visual' areas.

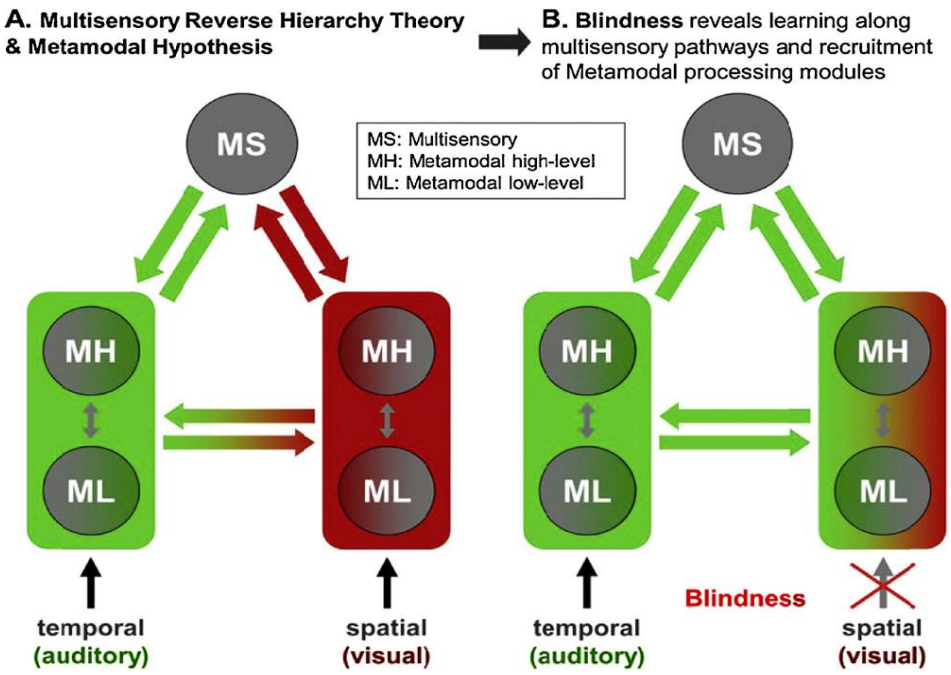

Figure 1.5b: Depiction of the implications of a metamodal brain organization for perceptual learning including showing the impact of blindness. Taken from (Proulx et al., 2014)

As stated, the key to this model is complexity. Complex learning or multisensory tasks are typically richer than unisensory tasks and are thus less likely to be defined by single low-level features. For example, while Braille reading is shown to activate low-level visual cortex in the blind, the semantic and linguistic requirements must be mediated at higher-order cortical levels. Learning therefore takes place beyond the primary cortices in higher-order areas allowing for generalized learning..

A multisensory reverse hierarchy theory could be used to evaluate whether a more informationally rich stimulus would facilitate superior learning in a typically uni-sensory task by driving processing to higher levels. For example, in temporal interval discrimination long periods of training result in generalization to novel frequencies but not durations (B. A. Wright et al., 2010). Would a complex stimulus comprised of multiple frequency bands modulate the breadth of generalization to new durations? In Chapter 2 I test this assertion by

adapting a low-level perceptual learning task by utilising a complex stimulus created from sonified image. I evaluate both the speed of specific learning and the breadth of generalization to untrained spectral and temporal features. This is interesting since Ahissar and colleagues (2009), in applying the RHT to auditory learning, posited that if complex stimuli facilitate perpetual learning then benefits could be seen at both lower and higher level representations. That is, low-level processing would be sufficient for specific frequencies and temporal resolution but the integration of spectral and temporal features at higher levels would be a requisite of generalization (Ahissar, Nahum, Nelken, & Hochstein, 2009).

1.6.0. Rationale.

In this introductory chapter I first described the 'problem' of blindness and visual impairment regarding worldwide prevalence and the large range of etiologies described in Table 1. The range of different causes of blindness requires a range of methodologies in visual rehabilitation, from medication and surgery to invasive implant technology at various levels of the 'visual' system; eye, optic nerve, cortex, and the focus of this thesis, sensory substitution.

Prior to describing a number of the different sensory substitution devices that have been developed it was important to give an insight in the theory behind why they work. The algorithms rely on crossmodal correspondences to inform the conversion principles and an interconnected metamodal organization of the brain to transmit and process sensory information unusually attributed to the impaired modality.

As well as describing some of the most popular SSDs in both substituting modalities, I also looked at how successful they are in a variety of 'visual' tasks, a brief evaluation of the

phenomenological experience, and areas of activation demonstrated by brain imaging studies in sensory substitution. It was important to stress that the cross-modal activity appears to not be instantaneous but manifests after periods of training, and activation prior to this is found in unimodal areas attributed to the modality of input (i.e. in VA substitution primary auditory cortex). Indeed intramodal connectivity within the auditory cortex may be influential in early learning. It is here that the focus of this thesis lies, in how naïve users of SSDs learn to effectively use the device. As early stage use appears to be a process of discrimination of the auditory characteristics it seems logical to question whether the principles of auditory perception limit the formation of object-based representation from sensory features.

The final section of this chapter looked at theories of perceptual learning in both the substituted (vision) and substituting (audition) modalities and also multisensory perceptual learning and how this can be applied to sensory substitution.

While there are the two main goals of the thesis the choice and design of experiments was not purely from a theoretical perspective. Three of the four experimental chapters were formulated to address a problem in device use, for example, how much sonified information is necessary to create object-based representations and does reducing the sensory load facilitate superior performance? This creates a dual purpose for the thesis; to advance theory on sensory substitution and the crossmodal nature of the brain, and how techniques can be applied to training. This is deemed of great importance considering the underuse, in the visually impaired community, of devices such as The vOICe. This may well be for a number of reasons; reluctance to use technology in an aged population, over blown expectations (in late blind especially, the ‘substitution’ is a poor analogue to the ‘real thing’), after initial positive results learning is slow, this is not an easy endeavour. Some of these issues can be

managed with education, such as managing expectations, while others can be aided by the provision of effective training protocols to enhance learning in quick and effective ways.

In Chapter 2 I look at learning in a low-level temporal discrimination task using a complex stimulus from a sonified image. The results are contrasted to what is found in the unimodal auditory literature to assess whether the use of the complex stimuli at low-levels drives processing to higher association levels, resulting in the broadening of generalized learning on spectral and temporal dimensions. The representation of information on the x-axis of the algorithm is based on both a left-to-right stereo scan and a temporal component to give a dual coding of horizontal visual information. While this should provide more accurate information the use of the stereo signal necessitates a headphone in each ear thus reducing perception of ambient sound in the environment. In applied use this could be problematic to a user generally reliant on sound to function in the environment. I therefore contrasted performance in full stereo mode with a single ear, time-scan only, condition to measure the magnitude of performance difference in the two modes. The results of this experiment should demonstrate whether the signal is being processed as 'purely' auditory, in which it would elect similar results as auditory paradigms, or whether it is promotes differential based on cross-modality or stimulus complexity.

In Chapter 3 I extend the idea, from a theoretical standpoint, of an auditory signal by assessing how auditory object representations may be created, and from an applied perspective, how much information is required for successful object recognition? One focus in sensory substitution and invasive visual rehabilitation research has been levels of acuity. in Chapter 3, I assess the importance of heightened acuity in simple object recognition in naïve users. In a forced choice object recognition procedure I manipulated the amount of information in the stimuli by varying the pixel resolution of the visual image prior to

sonification. Listeners were then required to match the variously degraded soundscapes with the high resolution visual and tactile objects. The objective was to demonstrate an information threshold where successful object discrimination breaks down.

The focus on the potential limitations of the auditory system on sensory substitution is continued in Chapter 4. While I have discussed the crossmodality of sensory substitution in some depth, in naïve users the characteristics of the auditory signal are important in device use. In learning, it may well be that unisensory information is integrated in multisensory areas, but this is still filtered by the auditory system prior to input into higher order association areas. I looked at a practical applied problem in sensory substitution, the discrimination of fine object features consistent in temporal duration, to evaluate why this breaks down at thresholds way above found in auditory perception. In a secondary consideration, evaluation was made on performance if categorically congruent and incongruent information from another modality was provided synchronously alongside the auditory information.

The final experimental Chapter 5 relates to Chapter 3. In Chapter 3 information is degraded to evaluate how little information is needed for object recognition. Chapter 5 looks at complexity at the other end of the spectrum to ask whether there are capacity limits in object recognition and if a division of features, based on successive or simultaneous presentation type, can elicit increased performance.

The overall goals of the thesis are to assess whether in naïve users, the output signal from the device is being processed primarily as an auditory signal, as is shown in imaging studies, and if object recognition is therefore potentially limited by auditory perceptual principles. Secondly, how the complexity of the signal influences performance and how this can impact

on perceptual learning. These goals are not mutually exclusive of course and should hopefully feedback into ways to develop effective training protocols.

Chapter 2

In Chapter 2 I evaluate the speed of learning and the breadth of generalisation in duration discrimination using a complex stimulus created from a sonified image. Results are compared to the literature in a similar unimodal task to ask the question; does an increase in signal complexity facilitate superior temporal discrimination? As a secondary measure a device setting is assessed to see if it elicits superior performance in this specific task.

This Chapter is an adaptation of:

Brown, D.J. & Proulx, M.J. (2013). Increased Signal Complexity Improves the Breadth of Generalization in Auditory Perceptual Learning. *Neural Plasticity*,

Increased signal complexity improves the breadth of generalization in auditory perceptual learning.

David J. Brown [1]

[1] Biological and Experimental Psychology Group, School of Biological and Chemical Sciences, Queen Mary University of London.

[2] Department of Psychology, University of Bath.

Abstract

Perceptual learning can be specific to a trained stimulus or optimally generalise to novel stimuli with the breadth of generalization being important for how we structure perceptual training programs. Adapting an established auditory interval discrimination paradigm to utilise complex signals I trained human adults on a standard interval for 2, 4, or 10 days. I then tested on the standard, alternate frequency, interval and stereo input conditions to evaluate the rapidity of specific learning and breadth of generalization over the time course. In comparison to previous research using simple stimuli, the speed of perceptual learning and breadth of generalization was more rapid and greater in magnitude, including novel generalization to an alternate temporal interval within stimulus type. I also investigated the long-term maintenance of learning, and found that specific and generalized learning was maintained over 3 and 6 months. I discuss these findings regarding stimulus complexity in perceptual learning and how they can inform the development of effective training protocols.

2.1.0 Introduction

Animals improve in the extraction and encoding of sensory information from the environment through perceptual learning. Psychophysical studies have established that practicing a task leads to specific improvements that are often restricted to stimuli used during training (Fiorentini & Berardi, 1980; McKee & Westheimer, 1978)**.** While these paradigms typically utilise simple unisensory stimuli, the Reverse Hierarchy Theory of perceptual learning is consistent with evidence that the 'default' setting in perception is one of higher order complex objects. For example, ecologically it is unusual to be presented with simple pure tones in isolation, but rather the complex frequency changes present in vocal communication such as birdsong and human speech (G. F. Ball & Hulse, 1998; Doupe & Kuhl, 1999; Fitch, Miller, & Tallal, 1997).

Auditory research shows that while specific learning is found in most tasks, generalization to novel stimuli is generally restricted to spectral features of the stimuli (Amitay et al., 2005; Demany & Semal, 2002; Fitzgerald & Wright, 2005; Irvine, Martin, Klimkeit, & Smith, 2000; Micheyl, Bernstein, et al., 2006; B. A. Wright et al., 2010).

In contrast, generalization to temporal stimulus features appears very limited although it has been found for transferral from interval to duration within the same stimulus length, and onset/offset asynchrony respectively (Karmarkar & Buonomano, 2003; Mossbridge, Scissors, & Wright, 2008). With regards to generalization to new intervals/durations - although Lapid and colleagues reported such generalization (Lapid et al., 2009), this is in contrast to the majority of studies in which no such transfer of learning is found (Karmarkar & Buonomano, 2003; B. A. Wright et al., 1997; B. A. Wright et al., 2010) and concerns a transfer across stimulus type (Empty-Filled). This limitation of generalization appears to be consistent even after extensive training (B. A. Wright et al., 1997; B. A. Wright & Sabin, 2007; B. A. Wright

et al., 2010) ,and with spectral feature processing and specific learning attributed to initial regions in the auditory cortex there is no anatomical limitation to this neural plasticity. However, temporal generalization may be sited in secondary auditory and multisensory areas utilising top-down processes to facilitate this learning. One key might be the use of simple versus complex stimuli during training (Ahissar & Hochstein, 2004; Ahissar et al., 2009).

Here I investigated the perceptual learning of complex auditory stimuli. Utilising an established temporal interval discrimination paradigm (B. A. Wright et al., 2010), I tested the specificity of learning to complex stimuli and the generalization to untrained durations within the same stimulus type. Using the data from Wright et al (2010) as a comparison for simple-stimulus based perceptual learning I tested whether the use of complex stimuli would speed perceptual learning and increase the breadth of generalization. I adapted the classic auditory learning paradigm in two ways (B. A. Wright et al., 1997; B. A. Wright & Sabin, 2007; B. A. Wright et al., 2010). First, the stimuli were complex, created by sonifying an image using a visual-to-auditory sensory substitution device (SSD) called the vOICe.(P. Meijer, 1992) This device uses crossmodal correspondences to transmit sensory information usually associated with an impaired modality (vision) via an unimpaired modality (audition). From an applied perspective it aims to give a basic visual percept to the visually impaired whilst theoretically acting as a valuable tool to evaluate multisensory processes in perception. The stimuli were created using The vOICe two reasons. Primarily the transformation algorithm of this device ensures that the auditory output signal is necessarily complex as over 4000 sonified ‘visual’ pixels create a soundscape comprising of multiple frequency and temporal components. Not only has this device been used to investigate the neural basis of auditory object recognition (Amedi et al., 2007; D. J. Brown et al., 2011), but results from this experiment could be

extrapolated to help formulate effective training paradigms for sensory substitution device usage. The second adaptation to the paradigm used by Wright and colleagues was the use of filled durations rather than empty intervals in both the training and test phases to evaluate whether the use of within-type complex stimuli would facilitate a learning advantage over simple stimuli. The literature has shown that while discrimination differences have been shown for empty intervals and filled durations, the methodology (2AFC) and durations (90-220ms) utilised in the present experiment show no significant differences and therefore comparisons with empty interval paradigms are valid (Rammsayer & Leutner, 1996).

Based on applying RHT to auditory stimuli, I hypothesized that complex stimuli can be learned specifically, but also increase the breadth of generalization. I also predicted that if signal complexity facilitates generalization then the use of The vOICe's stereo mode, with its two factor principle for horizontal spatial localisation, would outperform the monaural setting. If perceptual learning occurs at a higher, central neural level and results in generalization due to stimulus complexity , it is possible that such neural plasticity should be long lasting (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999)(Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999). While maintenance of perceptual learning has been demonstrated over 4 and 8 weeks respectively (Fitzgerald & Wright, 2005; Mossbridge, Fitzgerald, O'Connor, & Wright, 2006), I extended this time frame by conducting a follow-up experiment after 3 and 6 month periods signified by an absence of further training.

2.2.0 Materials and Methods

Listeners.

Twenty-four paid listeners (15 female) were recruited. Listener age range was between 19 and 35 (M=23.50, SD=4.9). All listeners reported normal hearing, normal or corrected eyesight, a formal education to undergraduate level or above, a good understanding of the English language, and provided written informed consent. Twenty-one of the listeners self-reported as right handed. Listeners were assigned to experimental groups in a pseudorandom manner aside from the gender split where 5 females were in each group. Each group completed the same task but was differentiated on the number of training days undertaken (2, 4 or 10).

Materials.

Stimuli were designed using The vOICe (Meijer, 1992) and Adobe Audition 3 - see 'Stimulus Design' below. The script was run in Matlab and Psychtoolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) on a Windows PC with a Creative Labs Soundblaster Titanium ASIO soundcard to ensure low latency. All auditory signals were transmitted to the listener through Sennheiser HD555 over ear headphones. The blindfold used was the Mindfold Inc. (Tucson, AZ).

Stimulus Design.

A plain white triangle (apex upwards) on a black background was sonified using The vOICe's image sonification feature. Prior to sonification, the device scan rate was set at 'x8' to reduce the temporal length of the stimulus to 125 milliseconds. This was then trimmed to remove the soundscape representing the black areas at each side of the triangle base resulting in an auditory stimulus of 90ms. Adobe Audition 3 was used to apply a 10ms cosine ramp

fade in and out to the stimulus onset and offset. This was done to avoid any distortions, or spectral splatter to the start and end of the soundscape, thus providing a clear signal. Frequency was measured as a range as the experimental aim was to create a complex signal comprised of a range of frequencies (each of the 4096 pixels has its own frequency, amplitude and temporal feature). For the standard stimulus the fundamental frequency was centered at 1 kilohertz (kHz), temporal duration of 90ms and amplitude of -85dB. The alternate 'test' stimuli were created by manipulating the standard stimulus in either Adobe Audition 3 (frequency) or The vOICe (interval). The frequency range was increased using a 0.60 ratio that raised the frequency range to one centered at 4kHz whilst retaining the 90ms temporal duration. The alternate duration was generated using the same visual stimulus but sonified using The vOICe at a 250ms scan rate. After the trim and ramp were applied the resultant stimuli was at 1khz frequency with a temporal duration of 220ms. For the stereo condition the frequency and duration values were the same as the standard (90ms, 1kHz) but the signal was conveyed through both headphones binaurally.

Figure 1.1. shows how The vOICe sonifies visual images in real time converting visual features (brightness and spatial position) to auditory features (amplitude, frequency, time and stereo panning). Each of the 4096 pixels in the recorded greyscale image is subjected to 3 conversion principles. Visual brightness is coded to auditory amplitude with brighter pixels eliciting louder tones. Spatial position uses two principles to code for vertical and horizontal localisation. On the y-axis pixel position corresponds to frequency with higher frequencies representing pixels higher up in the recorded image. A one second left-to-right time scan across the image provides a temporal cue to position on the x-axis with pixels to the left of the image being heard earlier in the time scan. If used in stereo mode a left-to-right pan across the stereo field provides, in conjunction with the time scan, a more accurate and complex coding feature for horizontal localisation with left orientated pixels being heard in

the left headphone. To give the final ‘soundscape’ all pixel sounds in a column are played concurrently (64 pure tones imposed over each other) with these 174 columns, or raster lines, then played sequentially over the duration of the time scan. The resulting ‘soundscape’ is a complex signal comprising of a large number of frequencies and amplitudes, played back to the user either monaurally or binaurally via headphones.

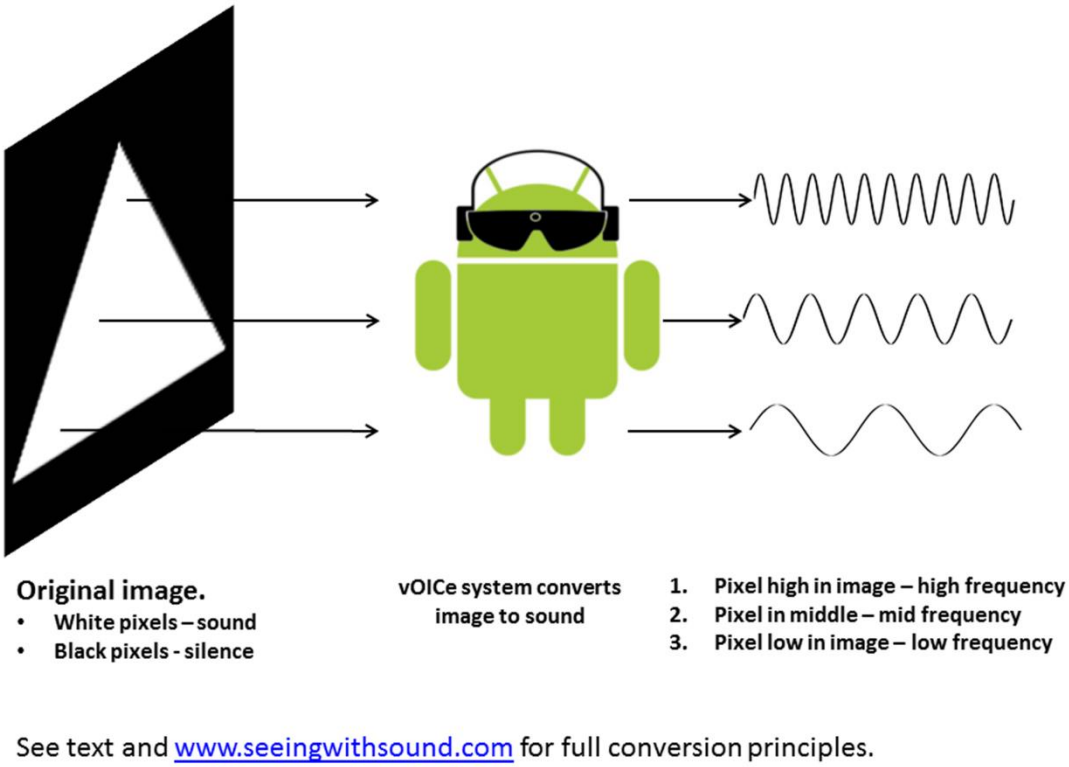

Figure 2.1: Conversion of image to sound using The vOICe algorithm. White pixels in the image are represented by a sound with black pixels silent. The elevation of each pixel is coded to frequency with pixels higher in the image having a higher frequency sine wave. All pixels in a vertical raster line are played simultaneously with a 1 second left-to-right horizontal scan across the image resulting in the soundscape for the image. (Image created by author)

Procedure

Listeners were assigned a work station, the procedure explained to them both verbally and via an information sheet, and written consent obtained. The blindfold and headphones were then put on and each listener guided to the '1' and '2' keys on the number pad on the PC keyboard. Listeners were then instructed to press the spacebar twice to start the first block of 60 trials. This double press of the spacebar was used to start all blocks in the condition (9 on training days and 5 for test days).

Figure 2.2. displays a sample trial for the standard condition. For each trial the listeners were presented with a pair of tones, separated by 970ms, in the left headphone. One of these tones was the 'reference' tone (t) which was temporally consistent throughout all trials in the particular condition. The comparison' tone ($t + \Delta t$) varied in duration dependent on previous answers and the 3 up/1 down psychophysical staircase procedure.

Three correct consecutive responses reduced the Δt by 1 unit whilst one incorrect response increased the Δt by one unit. The trial where the direction changed – from decrease to increase or vice versa – was classed as a reversal. For the first three reversals the unit change was 5ms with a 1ms change for subsequent reversals in each block.

While Figure 2.2. illustrates a trial in the standard condition this could also be represented for the other conditions. For example, in the alternate duration condition the reference cut-off point is at the same point on the downslope of the triangle because the signal duration was set using a slower scan speed. The triangle retains its proportionality to the background. The alternate frequency condition kept the same scan speed as the standard and with the temporal cuts being made to the auditory waveform post-sonification, only the spectrograph would differ (the triangle image is not showing specific frequency, just duration).

Listeners were required to indicate using the number keys whether the reference tone was presented first or second in the pairing. After the keystroke was made, feedback was provided by a 'pure tone' in the right headphone for an incorrect answer followed by the onset of the next trial. Correct responses resulted in the next trial starting with no prior auditory feedback.

After a 60 trial block was completed, the next block was initiated by the listener by a double depression of the space bar. This allowed the listener to take a short break at their own discretion. 'Official' breaks were also offered between the 5^{th} and 6^{th} blocks on a training day. During this intermission, the listeners were allowed to remove the headphones but not the blindfold. On the test days short breaks were taken between the conditions whilst the next conditions script was loaded into Matlab and an official break was offered after the first two conditions (10 blocks). The average time duration per block was four minutes.

The pre-test consisted of 5 blocks of each of the 4 conditions; standard, alternate duration, alternate frequency, stereo (1200 trials in total). The presentation of the conditions was varied amongst groups but was kept consistent within group concerning the pre- and post-tests. The standard condition was presented first for all groups in the pre-test phase.

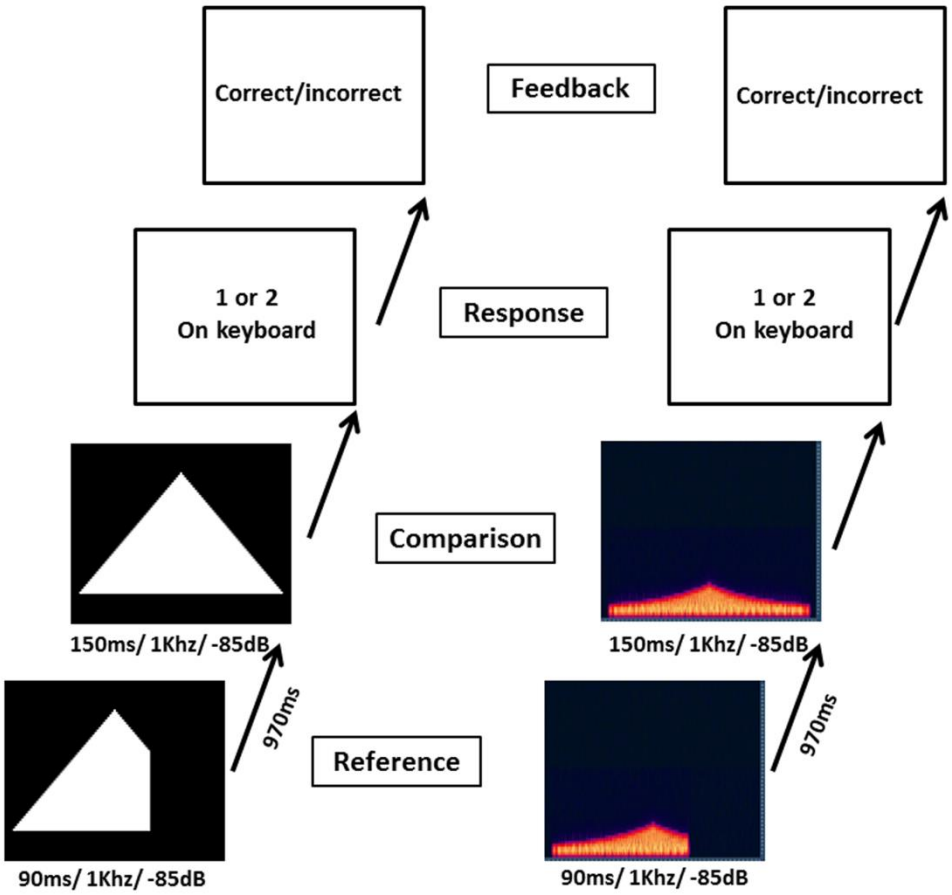

Figure 2.2: Representation of a sample trial. Listeners are presented with a reference soundscape followed by a 970ms inter-stimulus gap. They are then presented with a comparison tone and required to indicate whether the reference tone was presented 1st or 2nd. In the standard condition the reference tone is always of the same duration with reference and comparison tones presented in a random order. Feedback is given after response. The duration of the reference tone is stable with the comparison tone adapted on a 3 up/1 down staircase procedure. The left hand column of the figure shows the image that was sonified , with the right hand column showing the spectrogram for the resultant soundscape.

Calculation of thresholds.

Thresholds were obtained by first removing the first 3 or 4 reversals in each block to ensure an even number of reversals. If this resulted in there being less than 6 reversals in the block then the block was disregarded. For the accepted blocks the Δt for each of the reversals was noted and averaged across the block to give a block threshold. On the proviso that there were at least 3 (pre and post-test) or 6 (training) thresholds mean scores were calculated for

individual listeners and experimental groups for each session. Weber fractions were computed by dividing the total Δt by t and then entered for analysis.

2.3.0 Results

2.3.1 Learning on trained stimulus (specific learning)

Figure 2.3. summarises the results for specific learning of the standard duration (90ms, 1kHz, -85dB) over time. At pre-test there was no significant difference in the baseline scores for the three groups ($F(2,23)=0.147$, $p=0.864$, $\eta p^2=0.031$) and so levels of improvement from pre to post-test can be attributed to task duration. All three groups improved over time from pre-test to post-test (mean as a Weber fraction $\Delta t/t = 0.076$) as would be expected. A 2 time (pre and post-test) x 3 group (2d, 4d, 10d) ANOVA with *time* as a repeated measure showed this to be highly significant ($F(1,21)=52.392$,$p<0.0001$, $\eta p^2=0.714$). However, the amount of time training had little effect with no 'time x group' interaction ($F(2,21)=0.485$,$p=0.623$, $\eta p^2=0.044$) as all groups improved with equal magnitude. Improvement over the first 3 sessions (pre-test to training day 2) displayed a similar trend in that all groups improved over this time ($F(2,42)=43.663$,$p<0.0001$,$\eta p^2=0.675$) and again at a similar magnitude ($F(4,42)=0.508$,$p=0.730$, $\eta p^2=0.046$). Due to the possibility of a disparate number of blocks in the pre-test (5) versus training days (9) influencing the means, a 2 time (training days 1&2) x 3 group (2d, 4d, 10d) ANOVA with repeated measures on time was conducted. All groups improved over these 2 days ($F(1,21)=11.296$,$p=0.003$, $\eta p^2=0.350$) with no 'time x group' interaction ($F(2,21)=0.580$, $p=0.569$, $\eta p^2=0.051$). A final comparison in specific learning was to evaluate whether this improvement continued after the second day of training. A 5 time (pre-test, training days 1 to 4) x 2 group (4d,10d) ANOVA with time as repeated measures showed that this specific learning continued over time ($F(4,56)=19.256$,$p<0.0001$,

ηp^2=0.579) with equal amounts of learning for both groups (F(4,56)=0.459,p=0.766, ηp^2=0.032). Again to account for different block numbers the same analysis was conducted for these two groups from training days 1 to 4 with an improvement over time, albeit smaller than from pre-test (F(3,42)=2.868,p=0.048, ηp^2=0.170) with no group interaction (F(3,42)=1.225,p=0.312, ηp^2=0.080)

The results from the specific learning aspect of the experiment indicate that all groups improved over time, demonstrated by a lowering in discrimination thresholds from pre to post-test. Subdivision into the experimental groups was used to show whether this learning over time was consistent. As there was no significant difference between the three groups the implications are that the rate of specific learning is not dependent on the total amount of training and that the magnitude is equal across groups. Temporally, the largest amount of improvement was displayed over the first 3 or 4 sessions with any further learning over time at a lower magnitude (for all groups). This suggests that whilst initial specific learning is rapid, further learning can be viewed as fine tuning.

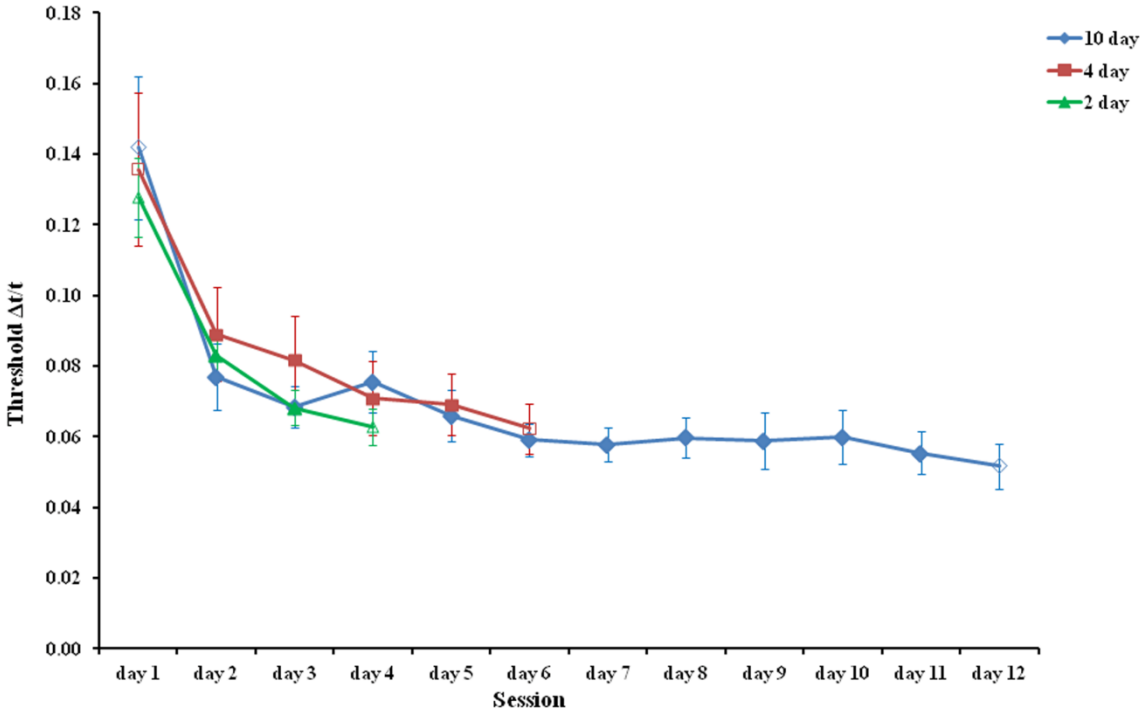

Figure 2.3. Learning curves showing mean temporal-duration discrimination (Δt/t for 79% correct performance) on the trained standard interval (90ms/1kHz, -85dB). Shapes on the lines represent experimental groups defined by number of days training (♦=10d, ■=4d, ▲=2d) with empty symbols showing pre and post-test sessions. All other 'days' are training days. Error bars indicate ± 1SEM.

2.3.2 Generalization

Figure 2.4. summarises the results for pre and post test scores for the trained standard duration (90ms, 1kHz, -85dB), and untrained frequency (90ms, 4kHz, -85dB), stereo (90ms, 1kHz, -85dB stereo), and duration (220ms, 1kHz, -85dB) conditions. Whilst the former tests for the specific learning described in the above section, the latter three indicate generalized learning. At baseline there were no group differences for frequency ($F(2,23)=0.150, p=0.861, \eta p^2=0.013$), stereo ($F(2,23)=1.638, p=0.218, \eta p^2=0.125$), or duration ($F(2,23)=0.204, p=0.817, \eta p^2=0.017$) again implying that group differences in improvement from pre to post-test was resultant of duration of training on the standard duration.

Frequency. The alternate frequency condition tested generalization to a spectral feature of the algorithm but retained the same temporal features as the trained standard. A 2 time (pre and post-test) x 3 group (2d,4d,10d) ANOVA, with time as a repeated measure, showed that all groups improved over time from pre to post-test ($F(1,21)=29.712, p<.0001, \eta p^2=0.586$) with a total mean reduction in discrimination threshold of (M=0.034). However, this was not dependent on group as there was no significant time x group interaction ($F(2,21)=0.089, p=0.915, \eta p^2=0.008$). This suggests that generalization to the untrained frequency occurred very early in the time course (2 days) in comparison to the 'simple' auditory paradigm (4-10days). It would be interesting to evaluate different frequency ranges for generalization. The upper limit of 4000Hz is just below the upper range of the device but far below the normal hearing range of humans. Theoretically, generalization should be found to frequencies above 4000Hz but more interestingly, considering the reduction of top-end frequency ranges in SSDs such as EyeMusic, it would be informative to generalize downwards to lower ranges.

Duration. In contrast to the frequency condition the alternate duration condition tested the temporal features of the multi-modal signal while retaining the spectral features of the trained stimulus. A 2 time (pre and post-test) x 3 group (2d,4d,10,d) ANOVA, with repeated measures on *time,* displayed a highly significant main effect of time ($F(1,21)=55.668, p<0.0001, \eta p^2=0.726$) with a mean reduction in discrimination thresholds across the full data set (M=0.022). In this condition the number of training days on the standard duration did have a significant difference on the amount of learning transfer with a significant time x group interaction ($F(2,21)=5.240, p=0.014, \eta p^2=0.333$). Contrasting the three groups to show where on the time course this generalization occurred showed that there was no difference between 2 and 4 day groups ($F(1,14)=0.600, p=0.452, \eta p^2=0.041$) but a highly significant difference between 10 and 2 days training ($F(1,14)=8.028, p=0.013,$

ηp^2=0.364). It appeared therefore that generalization occurred after 2 days of training. It seems highly likely however that this generalization occurred later in the time course as comparison of the 10 and 4 day groups was borderline significant (F(1,14)=4.424,p=0.054 ηp^2=0.240). To test if group composition influenced this contrast both listener age and gender were entered into a 2 x 3 ANCOVA to account for possible individual differences. Whilst age showed no influence (F(1,14)=4.466,p=0.054, ηp^2=0.242) there was a significant time x group interaction with gender as the covariate (F(1,13)=5.250,p=0.039, ηp^2=0.288). Thus generalization to the alternate temporal duration condition likely occurred somewhere between 4 and 10 days training on the standard. This is in contrast to training with simple stimuli where no generalization to the untrained duration was found after 10 days of training. Again it would be interesting to assess the limits of this duration generalization. At what baseline duration does generalization break down? With the task requiring the listener to pick the shortest stimulus (reference tone) performance should vary if the duration of the reference is manipulated. As Vierordt's law posits that short intervals tend to be overestimated and long intervals underestimated (Fortin & Rousseau, 1998), theoretically this would imply a reduction in performance either side of an ideal duration. Further research would be required to evaluate this.

Stereo. In the stereo condition a comparison was made between hearing the signal in stereo, where the x-axis is represented by both a time scan and stereo pan, and the monaural condition of the trained duration where the horizontal axis is represented by just the time scan. This was done simply by panning the output signal to the left channel in the stimulus design, this is analogous to using the device with only one headphone. I hypothesised that the combination of both time and stereo pan would result in a more complex signal than the time scan alone as it requires the processing of two bits of information to elicit the same result. While there were no group differences at baseline there was a significant difference between

the stereo and mono (standard) conditions that contained the same frequency and temporal features ($t(23)=6.188, p<0.0001, d=1.37$). Although this could convey an advantage for the stereo input over the mono input it must be taken into consideration that due to presentation order at pre-test each listener will have partaken in at least 300 trials at the standard duration (mono) before the stereo condition. With regards to group differences in the generalization to stereo stimuli a 2 time (pre and post-test) x 3 group (2d,4d,10d) ANOVA, with time as a repeated measure was conducted. As with all the other conditions there was a main effect of time ($M=0.032$) ($F(1,21)=22.841, p<0.0001, \eta p^2=0.521$) in that all listeners improved discrimination thresholds from pre to post-tests irrespective of number of days training on the standard stimulus. The number of days training didn't have a significant effect on the magnitude of improvement ($F(2,21)=1.740, p=0.200, \eta p^2=0.142$).

The results from the generalization section of the paradigm show that all groups improved on all conditions from pre to post-test. Group comparisons however showed that, unlike Wright et al (2010), training on the specific duration significantly increased the magnitude of learning on the alternate temporal duration. In this condition generalization occurred in the latter stages of the time course with only the 10 day training group showing this significant improvement. Improvement on the frequency condition was rapid within the first few sessions of training whilst the use of the stereo input conveyed an advantage over the monaural input at both pre-test and post-test. Indeed the post-test means for this condition were lower than for the trained standard condition implying an overall benefit of utilising the stereo input.

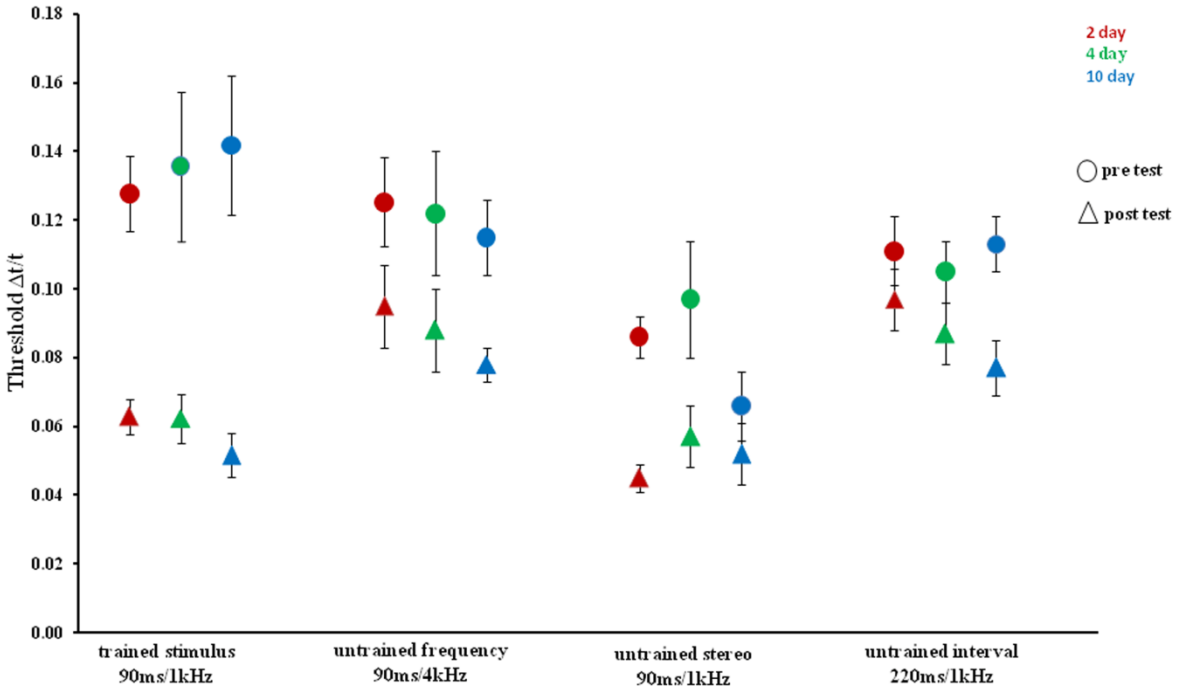

Figure 2.4 Mean temporal duration discrimination thresholds (Δt/t for 79% correct performance) for the trained interval (90ms/1kHz), untrained frequency (90ms/4kHz), untrained stereo (90ms/1kHz and untrained duration (220ms/1kHz). Circles represent pre-test scores with triangles showing the post-test scores. Experimental groups are differentiated by colours (2d training - red, 4d - green, 10d - blue). Error bars indicate ± 1 SEM. All groups improved on each condition form pre to post-test but there was only a significant group difference for generalization for the untrained interval condition where learning transfer was only found for the 10 day condition.

2.4.0 Experiment 2 Long term maintenance of perceptual learning

Experiment 2 was conducted to ascertain whether the perceptual learning achieved in Experiment 1 was maintained over time. Listeners from the 10 day training group were invited back to take part in another test phase session. This session was identical to the pre and post-test sessions of the original experiment (i.e. 4 conditions – 5 blocks per condition). Of the original group of listeners, seven of eight returned for testing. This group was further partitioned based on the time that had elapsed since finishing the Experiment 1 post-test. For

five listeners this time was equal to 6 months whilst for the remaining two, 3 months. Both the experimental setup and location were exactly the same as Experiment 1.

2.4.1 Results - Long term specific learning on the trained stimulus

Figure 2.5. summarises the results for the specific learning on the trained duration (90ms/1kHz,-85dB) after either 3 or 6 months from post-test. Collectively there was a significant improvement from pre-test to follow up session shown by a two time (pre-test, follow up) x two group (6months, 3months) ANOVA, with time as a repeated measure ($F(1,5)=19.482$, $p=0.007$, $\eta p^2=0.800$). However, as there was no significant time x group interaction ($F(1,5)=1.069$,$p=0.349$, $\eta p^2=0.176$) the duration from completion of experiment 1 had no significant influence on the maintenance of the perceptual learning. While group differences were not significant the average improvement for the 3 month group ($M=0.122$) was larger than the 6 month group ($M=0.076$).

When considering difference from post-test to follow up for the trained duration there was neither a main effect of time ($F(1,5)=0.003$, $p=0.986$, $\eta p^2=0.0006$) or group x time interaction ($F(1,5)=3.005$,$p=0.144$, $\eta p^2=0.375$). The mean scores showed that while there was a small decline in scores from post-test to follow up for the 6 month group ($M= -0.009$) the 3 month group actually improved between these two points on the time course ($M=0.002$).

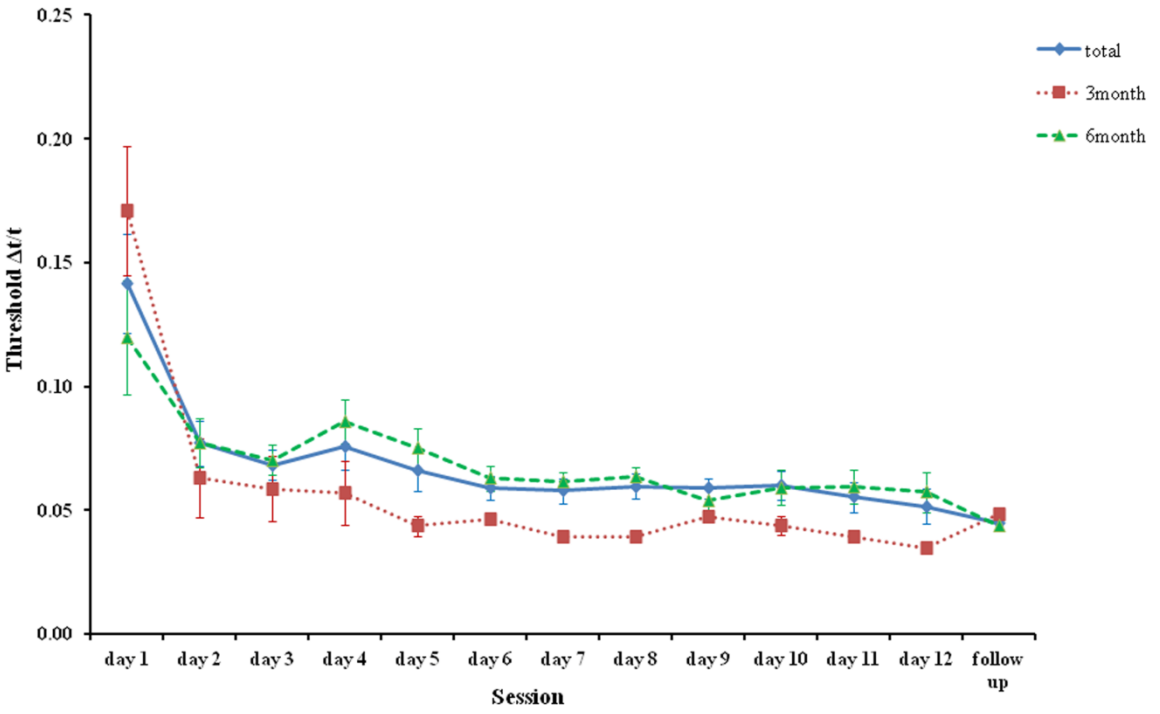

Figure 2.5. Learning curves showing mean temporal-duration discrimination thresholds (Δt/t for 79% correct performance) on the trained standard interval (90ms,1kHz,-85dB). The filled line (blue) represents the full data set for this experiment with the dotted line (red) showing the curve for the '3months since experiment 1 post-test' group and the dashed line (green) that for the '6 months from experiment 1 post-test' group. Session 1 is the pre-test, 2-11 are training days, 12 is the post-test and session 13, whilst not shown to scale, is the follow up at 3 or 6 months after the post-test session.. Error bars indicate ±1SEM

2.4.2 Long term Generalization

Figure 2.6. summarises the results for generalized learning to the untrained frequency (90ms, 4kHz, -85dB), stereo (90ms,1kHz,-85dB), and duration (220ms,1kHz,-85dB) conditions from both pre and post-tests. Considering frequency first, a two time (pre-test, follow up) x two group (6 month, 3 month) ANOVA with repeated measures showed that there was a borderline significant main effect of time for the full data set ($F(1,5)=6.486, p=0.051, \eta p^2=0.565$) and that this improvement was not dependent on group ($F(1,5)=0.259, p=0.632, \eta p^2=0.049$). All listeners showed lower discrimination thresholds at the long-term follow up than at pre-test implying that improvements due to the training on the

standard duration between pre and post-tests were at least maintained if not improved over durations of 3 and 6 months even without any additional training. During experiment 1 there was a considerable improvement between pre and post-tests for the alternate frequency condition so the carry over in performance improvement to follow up is not surprising. The results from experiment 1 show that an apparent ceiling level threshold is reached displaying a maximum benefit of training that would not be exceeded with additional sessions. Therefore when comparing post-test (where most listeners had attained this definitive threshold) and follow up I would not expect any further improvement. Indeed when contrasting these two points on the time course for the frequency condition there was no significant main effect of time ($F(1,5)=0.369, p=0.570, \eta p^2=0.069$) or time x group interaction ($F(1,5)=4.691, p=0.083, \eta p^2=0.484$). However, while the 3 month group showed a diminishment in the amount of learning transfer (M= -0.019) the 6 month group actually showed a small , non-significant level of improvement at follow up compared to post-test (M= 0.002); at the very least this suggests that the subjects maintained the level of performance achieved at the end of the training 6 months earlier.

A similar counter- intuitive result was also found when looking at the stereo condition. From pre-test to follow up, while there was no significant main effect of time ($F(1,5)=4.106, p=0.099, \eta p^2=0.451$) or time x group interaction ($F(1,5)=0.018, p=0.898, \eta p^2=0.003$), there was a small mean improvement for both the 6 month (M=0.024) and the 3 month conditions (M=0.026). When contrasting the post-test to follow up again there was no significant effect of time ($F(1,5)=0.114, p=0.749, \eta p^2=0.022$) or time x group interaction ($F(1,5)=5.764, p=0.62, \eta p^2=0.535$). However, on looking at the means the 6 month group showed a lower discrimination threshold at follow up than at post-test with an improvement of 0.029. This was not evident for the 3 month condition where there was a diminishment in threshold at follow up of M=-0.022.

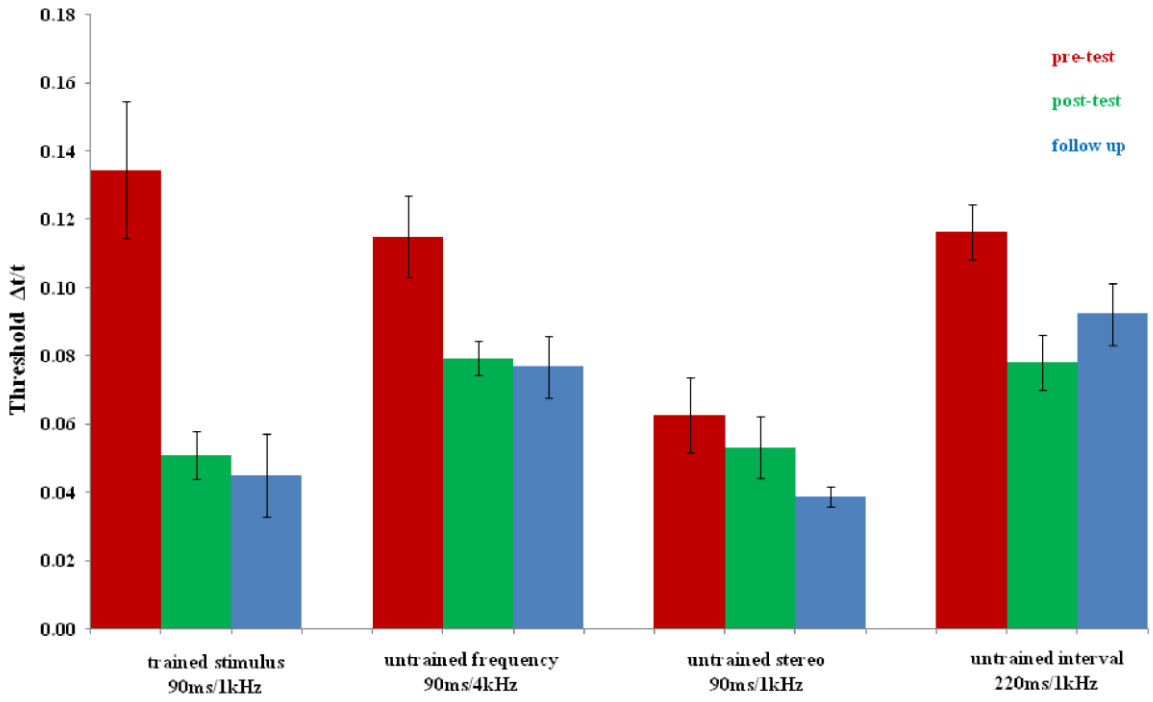

Figure 2.6. Discrimination thresholds (Δt/t for 79% correct performance) on the trained standard interval (90ms,1kHz,-85dB) and untrained frequency (90ms, 4kHz, -85dB), stereo (90ms, 1kHz, -85dB), and interval (220ms, 1kHz, -85dB) conditions. Red bars indicate pre-test scores and green bars post-test scores from Experiment 1. Blue bars indicate scores from the follow up study, Experiment 2.. Error bars indicate ±1SEM

When looking at the follow up data for the alternate untrained duration (220ms) there was an overall significant improvement from pre-test to follow up ($F(1,5)=10.405, p=0.023, \eta p^2=0.675$) and although this was not dependent on group ($F(1,5)=2.406, p=0.182, \eta p^2=0.325$) the mean scores showed that the performance improvement was larger for the 3 month ($M=0.044$) than the 6 month condition ($M=0.015$). However, as the initial baseline thresholds were considerably higher (worse) for the 3 month group, the implications are that the main improvement was within the training days rather than in the 'break' post-test. From post-test to follow up there was no main effect of time ($F(1,5)=2.280, p=0.191, \eta p^2=0.313$) or time x group interaction ($F(1,5)=0.961, p=0.372, \eta p^2=0.161$) and whilst the means showed that both groups performed worse on the alternate duration after the respective break, the

diminishment of performance was minimal and therefore the perceptual learning had been maintained over 3 and 6 months.

2.5.0 General Discussion.

The purpose of this study was to evaluate whether the perceptual learning of complex auditory stimuli might result in greater and longer lasting generalization than previously reported in the literature. Firstly, similar specific learning results were found despite the increased complexity of the stimuli. Secondly, I discovered the first instance, to my knowledge, of generalization to a novel temporal duration within stimulus type in contrast to the prior temporal generalization found by Lapid et al (2009) where the generalization was across stimulus types (empty to filled). I deem this within stimulus temporal generalization important as this is ecologically valid to everyday processes such as speech, which predominantly consist of filled soundscapes. Thirdly, I also assessed for the first time whether the improvements brought about through auditory perceptual learning could be maintained over a long delay period of three to six months, and indeed found that the benefits of specific and generalized learning were retained.

As with the results in the foundational work in interval discrimination by Wright and colleagues (2010), specific learning of the trained duration (90ms, 1kHz) occurred early in the time course with a statistically significant improvement shown by the first test day (after two 540 trial training days). Indeed there was a significant improvement after only 1 training day but due to a disparate number of blocks between test and training phases this should be approached with caution. Generalization to untrained conditions occurred later in the time course implying different neural processes for specific and generalized learning. A significant improvement for the untrained frequency (90ms, 4kHz) was found somewhere between 2 and 4 days of training, more rapidly than in the simple unisensory paradigm, with the novel

findings of generalization to the untrained duration (220ms, 1kHz) occurring later in the time course (between 4 and 10 days). I can therefore draw similar conclusions to Wright and colleagues in that generalization to novel stimuli requires a distinct amount of training. The use of complex stimuli extends the previous work in that it not only appears to decrease the amount of training required for generalization to occur (frequency) but also to increase the breadth of generalization (duration) facilitated by training on a standard duration. Utilising complex and multisensory stimuli also draws comparisons with speech perception where complex signals and auditory-visual multisensory processes are the norm (Skipper, van Wassenhove, Nusbaum, & Small, 2007; van Wassenhove, Grant, & Poeppel, 2005).

In postulating an explanation for the novel results found in the complex stimuli learning paradigm, here I consider theories of perceptual learning, the possible neural networks involved, whether the complex composition of the stimuli would facilitate the use of alternate networks, and finally if the stimuli are actually being processed solely as auditory signals.

While Wright and colleagues (B. A. Wright et al., 1997; B. A. Wright & Sabin, 2007; B. A. Wright et al., 2010) speculated that, due to different positions on the time course, specific and generalized learning may utilise different or modified neural circuitry, this may be further influenced by the neural networks that facilitate spectral and temporal processing and how they are integrated in the multisensory signal. Auditory processing is assumed to be analogous to the visual system in that two functional pathways are utilised to process ‘what’ and ‘where’ information (Zatorre, Bouffard, Ahad, & Belin, 2002). For the latter the posterodorsal pathway from the primary auditory cortex (A1) through the posterior temporal lobe and posterior parietal lobe to the dorsolateral frontal lobe, has been proposed for spatial processing with the anteroventral pathway from A1 through anterior temporal lobe to inferior frontal lobe coding the ‘what’ features of the signal.

Whether the signal is being processed as unisensory or multisensory, the Reverse Hierarchy Theory of perceptual learning provides a theoretical explanation for the results found utilising complex stimuli (Ahissar & Hochstein, 2004). Primary to this theory is that perceptual learning can happen at any level of processing and it is the complexity and difficulty of the task that guide the level at which the processing occurs. Difficult tasks, where more specific discrimination is required, focus attentional resources to primary sensory areas. However, if the task can be accomplished utilising more general object features the processing drives attention to higher levels. The reverse hierarchy theory can be applied to both unisensory and multisensory learning using similar mechanisms. Modality specific unisensory learning is supported by either low-level auditory areas for specific learning or high-level auditory areas for general learning (Ahissar, 2001; Ahissar & Hochstein, 2004). Learning utilising multisensory stimuli can lead to correlated activity in higher-level multisensory areas (Shams & Seitz, 2008) or learning can progress from primary sensory areas to higher-level multisensory areas under complex unisensory stimulation. Activity may then cascade back down the hierarchy such that generalization across modalities occurs when these higher-level multisensory areas are implicated in learning either unisensory or multisensory tasks (Proulx et al., 2014).

This naturally raises the question as to whether the novel generalization found to the alternate temporal duration in the multisensory paradigm is due to the spatiotemporal composition of the sonified image or can be attributed purely to the complexity of the signal. Future research could evaluate this by creating auditory stimulus sets which incorporate a number of frequency bands superimposed over each other to create a complex, but still unisensory, signal. If generalization to the alternate temporal duration is not found in this complex signal then the implications are that it is the multisensory nature of the signal that is driving temporal generalization rather than complexity per se.

A final consideration concerns the results from Experiment 2. To my knowledge this is the first evaluation of such long term benefits of perceptual learning in the auditory or multisensory domains and provides invaluable information for developing long-term training protocols. Performance in all conditions was not only superior to results for the pre-test phase but, alternate duration aside, also superior to the post-test phase. This implies that the specific and generalized learning attained through training is maintained over considerable lengths of time even without additional training. The results from post-test to follow up - that is, participants continue to improve over periods of no training - is somewhat counter-intuitive. I have to be wary of stating this is an experimental effect due to the low number of listeners in the 3 month group, however, I can theorise why these incongruous results occur. In structured interviews with long term users of The vOICe it was reported that one user experienced vOICe like visual percepts evoked by auditory stimulation even when not using the device. These were elicited by environmental sounds that were vOICe-like in composition but not multisensory in nature (Ward & Meijer, 2010b). It may therefore be possible that exposure to such sounds is strengthening neural networks instigated or unmasked through device use. If this is so, then the 6 month follow up group may have been exposed to more of these sounds than the 3 month group and therefore the learning network is further strengthened, hence the greater improvement for those with the longer absence. Future experiments could test for this by providing post-training listeners with complex unisensory sound stimuli in a non-structured setting between post-test and follow up.

While the main rationale of the experiment was theoretical, in evaluation of how complex signals are processed in comparison to simple stimuli the results also have application. Indeed the theoretical implications are salient to SSD use as they inform on how signals from the natural environment may be processed in the paradigm. It is unusual, after all, for us to encounter simple unimodal stimuli, as tested in the lab, in everyday situations. Within

training protocols the results offer a number of possibilities. Firstly, generalization to novel intervals could impact on training on recognition of object size invariance, as x-axis (time) is coding for the length of the object. For example, would training on a square with a length equivalent to a 200ms pan/stereo scan generalize to larger or smaller squares with different durations? Naturally for a square, y-axis dimensions are locked to x-axis, so a frequency judgement is to be made as well, but as frequency is generalizable, this should be additional and advantageous. Indeed in an app developed in the lab to train on sensory substitution a 4AFC involving various sized squares was problematic to naïve users. Secondly, if complex signals convey an advantage in learning there is justification for a rapid advancement in the complexity of stimuli across the time course. Generally, naïve users are presented with simple stimuli to aid an understanding of the algorithm and success is good. However, perhaps this is a too simplistic approach and if replaced with more complex stimuli the learning could still be effective. In Chapter 3 I expand on the idea of complexity and offer results that build on this idea of using more complex stimuli.

Another consideration is the maintenance of learning over periods of time with no training. That a person can break from the training for up to 6 months and not suffer a degradation in performance is promising. While this was a simple task it demonstrates rapid learning and this may extrapolate to other more complex learning tasks. This implies that while immersive device use is preferential in learning, the user can remiss from training with little negative impact.

As far as device use goes – to elicit best performance from the device use both headphones for the dual factor coding of horizontal features. While this conveys ecological disadvantages in reducing input from the environment this may be negated using bone conduction headphones. Finally, advantages in learning sensory substitution may also be conveyed by

incorporating complex unisensory auditory tasks into the training protocols. Indeed this could also be bi-directional in utilising sensory substitution in auditory training such as speech therapy.

Chapter 3.

In the Chapter 2 I demonstrated that the speed and breadth of perceptual learning on a low-level discrimination task is increased using complex stimuli from a sonified image. While it is tempting to therefore regard the signal as different from a purely auditory signal it is doubtful whether the information is being processed crossmodally due to the use of naïve listeners (i.e. insufficient time for plasticity). More likely is that the complexity of the auditory signal facilitates duration discrimination from spatial components (i.e. frequency).

In Chapter 3 I look at the formation of auditory objects and assess how much information is required to elicit successful recognition of sonified two dimensional objects, hypothesising that phase locking, dependent on resolution of the sonified image, dictates where in the cortical hierarchy the object is formed.

This Chapter is an adaptation of

Brown, D.J., Simpson, A.S., & Proulx, M.J. (2014). Visual Objects in the Auditory System in Sensory Substitution: How much information do we need? *Multisensory Research*, 27, 337-357

Visual Objects in the Auditory System in Sensory Substitution: How much information do we need?

David J. Brown[1,2],

[1] Crossmodal Cognition Lab, Department of Psychology, University of Bath, UK

[2] School of Biological & Chemical Sciences, Queen Mary University of London, UK

Abstract.

Sensory substitution devices such as The vOICe convert visual imagery into auditory soundscapes and can provide a basic 'visual' percept to those with visual impairment. However, it is not known whether technical or perceptual limits dominate the practical efficacy of such systems. By manipulating the resolution of sonified images and asking naïve sighted participants (n=19) to identify visual objects through a six-alternative forced-choice procedure (6AFC) I demonstrate a 'ceiling effect' at 8x8 pixels, in both visual and tactile conditions, that is well below the theoretical limits of the technology. I discuss the results in the context of auditory neural limits on the representation of 'auditory' objects in a cortical hierarchy and how perceptual training may be used to circumvent these limitations.

3.1.0 Introduction.

Visual impairment affects 285million people worldwide with 39 million of these legally blind, defined by a visual acuity of less than 20/200 or visual field loss to less than 10^0 (Pascolini & Mariotti, 2012). While a proportion of cases can be treated through surgical procedures such as the removal of cataracts, the development of compensatory techniques is essential for providing a basic visual percept for non-treatable patients. These techniques can be divided into invasive and non-invasive. Invasive techniques involve electrodes implanted in the eye ,epi or sub-retinal-retinal, (Benav et al., 2010; Eickenscheidt, Jenkner, Thewes, Fromherz, & Zeck, 2012; Fujikado et al., 2011; Keseru et al., 2012; Weiland et al., 2005; Zrenner et al., 2011)optic nerve (Chai, Li, et al., 2008; Chai, Zhang, et al., 2008; Veraart et al., 2003)or cortex (Brindley & Lewin, 1968a, 1968b; Dobelle & Mladejovsky, 1974; Dobelle et al., 1974; Normann, Maynard, Rousche, & Warren, 1999; Schmidt et al., 1996).

In the case of retinal implantation, assuming that all implanted electrodes contact the targeted retinal cells, state of the art technology incorporating 100 channels provides a theoretical working resolution equivalent to 10 x 10 pixels. However, the simulations of Weiland and colleagues (2005) have suggested that up to 1000 electrodes (e.g., around 30 x 30 pixels) would be necessary for visual processes such as face recognition or text reading. This is supported by Li et al's evaluation of object recognition with retinal implants, which implied an upwards ceiling effect at 24 x 24 pixels (Li, Hu, Chai, & Peng, 2012).

Non-invasive compensatory techniques rely on technology and neural plasticity to transmit information usually attributed to an impaired sense via a neural network of an unimpaired modality. This 'sensory substitution' generally substitutes for impaired vision with the substituting modality being touch (Bach-y-Rita, 2004; Bach-y-Rita, Collins, White, et al., 1969; Bach-y-Rita & S, 2003; Y. Danilov & M. Tyler, 2005; Danilov, Tyler, Skinner, Hogle,

& Bach-y-Rita, 2007) or audition (Abboud et al., 2014; Arno, Vanlierde, et al., 2001; Capelle et al., 1998; P. Meijer, 1992).

The sensory substitution device (SSD) is a 3 component system: a sensor (camera) to record information, an algorithm (on PC or smartphone) to convert it, and a transmitter (headphones or tactile array) to relay converted information back to the user. Perceptual resolution, or acuity, of visual-to-tactile (VT) devices are constrained by the distribution of touch receptors at the point of contact (back, fingers, tongue) resulting in low resolutions ranging from simple 10x10 systems to the 20x20 electrode Brainport (Bach-y-Rita, 2004; Bach-y-Rita, Collins, White, et al., 1969; Chebat et al., 2007; Y. P. Danilov & M. Tyler, 2005; Sampaio et al., 2001).

Unlike VT devices, visual-to-auditory sensory substitution devices (VA) are not constrained by the density of surface area receptors but instead exploit the wide frequency resolution of the cochlea and the large dynamic range of the auditory nerve. This allows for a much higher theoretical and functional resolution (Haigh et al., 2013; Striem-Amit, Guendelman, et al., 2012). As with VT SSDs, resolution varies amongst VA devices. For example, the Prosthesis for Substitution of Vision by Audition (PSVA) has dual resolution function with an 8x8 pixel grid of which the four central pixels are each replaced by four smaller ones. The 60 large pixels in the periphery and 64 smaller central pixels (fovea) give the PSVA a functional resolution of 124 pixels (Capelle et al., 1998). The VA device used in the experiments reported here, The vOICe (P. Meijer, 1992), which has been used to demonstrate auditory object recognition and localisation (Auvray et al., 2007; D. J. Brown et al., 2011; Proulx et al., 2008), utilises a 176x64 pixel array for a functional resolution of up to 11,264 pixels.

This leads to the question: do such systems exhibit ceiling effects in object recognition performance similar to those reported using invasive systems? (Li et al., 2012)(Li et al.,

2012). The source of such limits on performance can arise at multiple points along the neural pathways processing such information. Many studies of trained users of The vOICe and other SSDs have shown neural activity in brain areas commonly thought of as visual. The sensory modality being stimulated (such as the auditory system) is also activated and likely relays the information to the visual system. Due to the necessary transduction of sensory information in the stimulated modality (such as auditory cortex) before being later processed by the target modality (such as visual cortex), it is fundamental to understand how the capacity of the auditory system impacts the information available for further computations.

In auditory-visual substitution, the features of a two-dimensional image which represent an object are encoded as independent spectro-temporal modulations within a complex acoustic waveform (P. Meijer, 1992). Such acoustic features are encoded independently in the peripheral auditory system and object-based representations emerge in primary auditory cortex (Ding & Simon, 2012; Mesgarani & Chang, 2012; Shamma, Elhilali, & Micheyl, 2011; Teki, Chait, Kumar, Shamma, & Griffiths, 2013)Auditory cortex maintains a two-dimensional topographic map of frequency (Humphries, Liebenthal, & Binder, 2010) and modulation-rate (Barton, Venezia, Saberi, Hickok, & Brewer, 2012) that are the so-called tonotopic and periodotopic axes, where individual regions on the map independently represent sound features occurring at a specific frequency and modulation rate (Barton et al., 2012; Simon & Ding, 2010; Xiang, Poeppel, & Simon, 2013). It is thought that auditory objects are formed, in cortex, according to temporal coherence between these independently-coded acoustic features (Shamma et al., 2011; Teki et al., 2013).

The representation of spectro-temporal modulation is increasingly rate-limited in the ascending auditory pathway. Phase-locking on the auditory nerve is limited to around 4,000 Hz (Joris, Schreiner, & Rees, 2004). By midbrain (inferior colliculus) this limit is reduced to

around 300 Hz (Baumann et al., 2011; Joris et al., 2004)and by primary auditory cortex it is further reduced to around 30 Hz (Barton et al., 2012). In superior temporal gyrus (part of Wernicke's speech area), this limit is further reduced to <16Hz in the object-based representation of speech (Pasley et al., 2012), which coincide with those established in human psychoacoustic studies (Simpson & Reiss, 2013; Simpson, Reiss, & McAlpine, 2013).

Therefore, different stages of the auditory pathway provide different limits on the visual-sensory substitution problem, where the information encoded in the rendering of the visual image is encoded with increasingly coarse temporal features as it ascends. This is consistent with the Reverse-Hierarchy Theory of multisensory perception and perceptual learning (Proulx et al., 2014), where primary sensory areas provide greater specificity, and higher order areas provide perception at a glance (Ahissar & Hochstein, 2004; Ahissar et al., 2009) If auditory objects are pre-requisite in VA substitution, this limit is placed earliest at primary auditory cortex. If auditory objects are further refined in higher cortical areas implicated in speech processing, this limit is further strengthened.

These postulations provide testable hypotheses. The image-to-sound rendering system (P. Meijer, 1992) breaks the visual image into arbitrary pixel sizes which correspond to a resampling of the acoustic modulations by which the image is represented. Shannon-Nyquist sampling theory dictates that the fastest modulations captured are at half the sample (in this case pixel) rate. By varying the pixel resolution of the rendered image it is possible to alter the upper limit (of modulations captured) in a way that is equivalent to the various limits seen on the auditory pathway. If object recognition performance is limited by modulation processing in primary auditory cortex, there should be ceiling effects seen at pixel sampling rates of around 50-60 Hz (giving a cut-off frequency of 25-30 Hz) equivalent to 16x16 pixel visual object

(Figure 3.1a.). If performance is limited by higher cortical processing (in speech related areas) then ceiling effects may be seen at even lower pixel (8x8) rates of around 20-30 Hz (giving a cut-off frequency of 10-15 Hz).

The frequency range and temporal length of the sonified stimulus may also be a factor in object recognition. Wright et al (2010) demonstrated generalization to untrained frequencies but not temporal intervals (B. A. Wright et al., 1997; B. A. Wright et al., 2010) and while evidence shows that increased complexity in sonified images increases the breadth of generalization to untrained temporal features (Brown & Proulx, 2013) the extended time course for the latter implies a dominance of frequency components. However, with the meta-modal theory of brain organization postulating that the auditory system is preferential for temporal processing, and visual for spatial, (Pascual-Leone & Hamilton, 2001; Proulx et al., 2014), and the hypothesized reliance of auditory characteristics in naïve users of SSDs (D. J. Brown et al., 2011) there is an argument for temporal dominance. To explore this I categorized the test stimuli into ‘short’ with a wide frequency range (M=3951Hz) and short temporal length (M=758ms) and ‘long’ with a narrow frequency range (M=2280Hz) and long temporal length (M=951ms; see Figure 3.1b.). This allows us to evaluate whether there is dominance of the spectral (frequency) or temporal (signal length) features of the algorithm in object recognition.

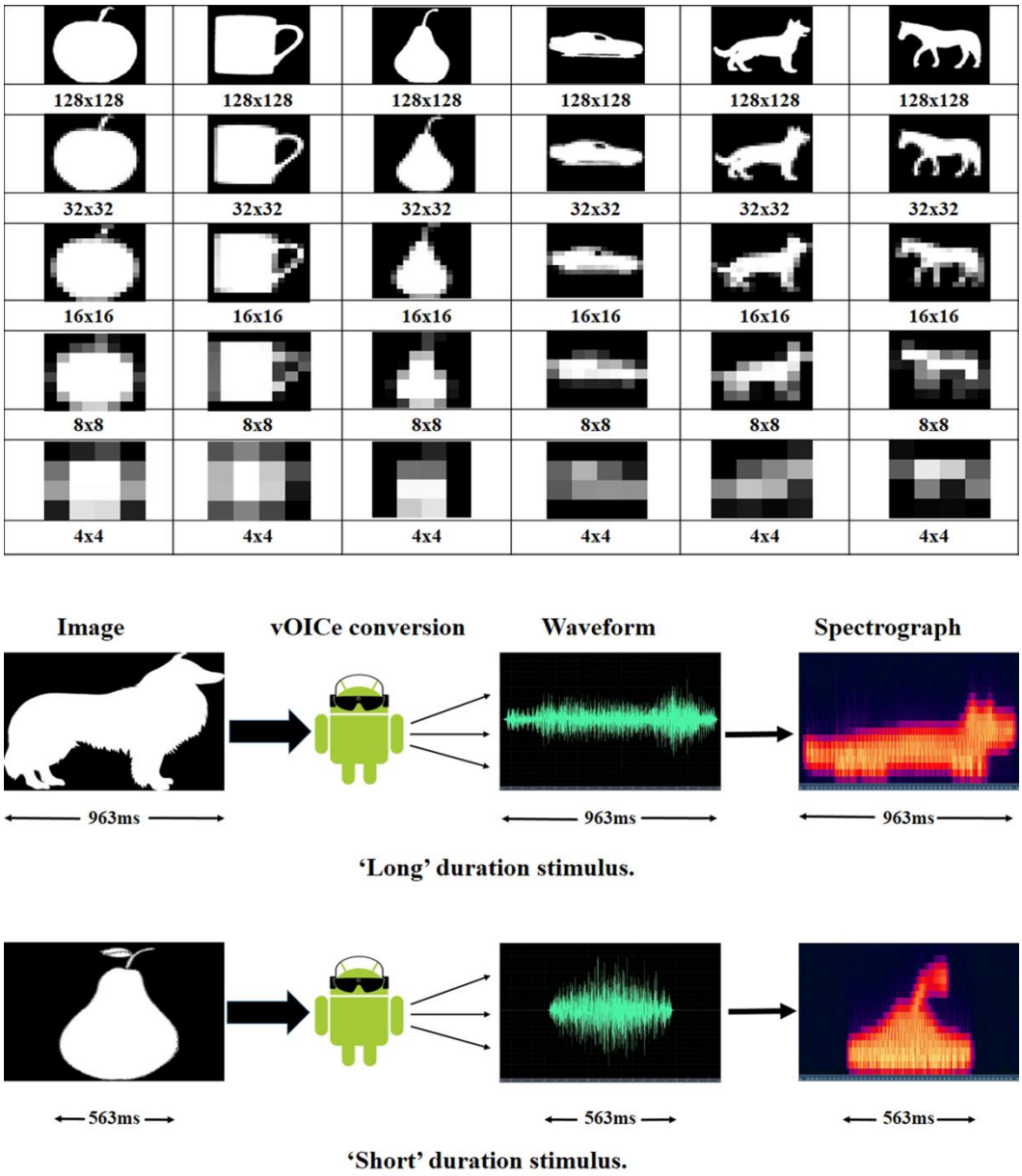

Figure 3.1a. (top): Visual representation of the sonified objects used in the test phases of the experiment. Objects presented to the participant (visually or haptically) were always at the 128x128 resolution. The objects at 32x32, 16x16, 8x8, 4x4 resolution were sonified using The vOICe and presented as auditory soundscapes only. The participants were never exposed to the visual or tactile objects at the reduced resolutions.

Figure 3.1b. (bottom): The sonification of one 'long' category object and one 'short' category object. The original visual image is shown along with the waveform and spectrograph of the sonified object.

A second stimulus consideration was the use of both visual and tactile objects. The target population for SSDs are those with visual impairment, rendering the association between soundscape and visual object meaningless. The reasoning behind the visual component of the task was due to demographics of the participants who were all sighted and naïve to the device. In attempting to demonstrate a proof of concept it seemed logical to train in a familiar modality (vision) for relative simplicity, and a modality relevant to application (tactile).

The rationale is threefold. Firstly, to evaluate the minimal level of information required for successful object recognition in VA SSD. Based on comparable studies with retinal implants and the visual information displayed in Figure 3.1a. I predict a ceiling effect at either 8x8 or 16x16 pixels after which an increase in resolution will not elicit superior performance. Secondly, to utilise a behavioural paradigm to assess where in the auditory hierarchy resolution based objects are processed. For the larger of the predicted ceiling effects I hypothesise auditory object recognition in primary auditory cortex, with lower ceiling effects further up the auditory pathway. Finally I am interested in whether recognition would be better for stimuli with a ‘short’ duration and wide frequency range than for those with a ‘long’ duration and narrow frequency range. As this is exploratory I make no directional hypothesis.

3.2.0 Method.

Listeners.

I recruited 19 undergraduate students (12 female) from 18 to 28 years of age (M=20.42, SD=3.22) from Queen Mary University of London. Two listeners withdrew from the study after the training session so 17 listeners (10 female) age range 18 to 28 (M=20.71, SD=3.29) took part in the test phase. All listeners reported normal or corrected vision and normal hearing. 16 (training) and 14 (test) were right handed. The study was approved by Queen Mary University of London ethics Committee REC/2009 and all listeners provided written consent prior to the study onset. Remuneration was via the undergraduate course credit scheme with an additional £0.05 per correct response in the test phases.

Materials

'Auditory' stimuli were created using The vOICe (Meijer, 1992), Adobe Audition 3 and Adobe Photoshop CS3. (see stimulus design below). The script was run in E-Prime2.0 (Psychology Software Tools, Pittsburgh, PA) on a Windows 7 desktop PC. All auditory signals were transmitted via Sennheiser HD555 full ear headphones. Images to be sonified were obtained from EST 80 image set (Max Planck Institute, Germany) and Clipart. The blindfold was the Mindfold (Mindfold Inc. Tucsan, AZ).

Stimulus design.

Images were transformed to soundscapes using The vOICe's image sonification feature at default settings (1 second scan rate, normal contrast, foveal view - off). Visual images were white on a black background with a 1 second duration on the x- axis and a 500-5000hz frequency range on the y axis. Tactile stimuli were created by cutting the object shape (white area) from 5mm foam board and attaching this to 90mm x 55mm card backgrounds. For the training days there were 40 different objects in total (34 on day one)

Test day stimuli – object resolution and categorization. During the test phases only six visual and six tactile stimuli were presented to the listeners. These were all at 128x128 pixels. These visual images were manipulated in Adobe Photoshop to produce variants at four pixel resolutions (32x32, 16x16, 8x8, 4x4) and then sonified (Figure 3.1a.). Hence the tactile or visual objects were always at 128 x 128 pixel resolution while the soundscapes were at various lower resolutions subdivided into two categories based on the temporal and spectral features of the rendered soundscape. 3 objects were 'long' on the x axis but narrow on the y axis (car, dog, horse) ,with the other 3 relatively 'short' on the x-axis but with a broad range of frequencies on the y-axis (apple, pear, cup . When sonified this resulted in either long, spectrally sparse or short spectrally dense signals as shown in Figure 3.1b.

Procedure.

Training day one. Listeners were shown a PowerPoint presentation about The vOICe algorithm, including worked audio-visual examples, and an explanation of the experimental task

For each task trial listeners listened to a soundscape (repeated 4 times) while looking at a blank screen. The soundscapes were then repeated accompanied by four numbered images on the screen. The listeners indicated, using 1-4 on a numeric keypad, which image had been sonified to create the soundscape. The soundscape could be repeated by pressing 'R' and visual feedback was given post-response in a correct/incorrect format prior to onset of the next trial.

There were 32 trials in each of 2 blocks. Each block had 4 categories of trial, varying in difficulty based on object features. For example, in the first 8 trials the correct object varied greatly from the 3 alternates. For the second set there were 2 obviously different alternates and so on. The trials alternated between filled and empty objects (object outline only) to

evaluate The vOICe's edge enhancement feature in early stage training. The second day one training phase replicated the first aside from that images were sonified at a 2-second scan rate. For the final 2 blocks on training day one the listeners were blindfolded and undertook a similar 4AFC procedure involving associations to be made between the soundscapes and the haptically explored tactile objects. Responses and requests for repeat presentations were instigated by the experimenter. Tactile blocks were completed after the visual ones for all listeners. Otherwise all presentation orders were counterbalanced.

The second training day was a replication of day one (minus the PowerPoint presentation), utilising different 4AFC's, and reversing the procedure so the listeners was presented with one object (visual or tactile) and 4 soundscapes (each repeated 4 times). The six test day objects (at 128 x 128 pixels) were introduced during this session, although the listeners were unaware these were the test day objects. 1- or 2-second scan rate order was counterbalanced across days. After the second training day, listeners who had a $\geq$ 50% correct response rate (based on a pilot study with different listeners) were invited to return for the test phases.

Test Day One. Methodologically this was similar to the training phases but with a number of alterations. Firstly, there were 6 presented objects in each trial (6AFC) with the same 6 objects being presented for each trial. Secondly, there was no post-trial feedback. Thirdly, there were 72 trials in each block of the visual test phase and 36 in each tactile block. Listeners were given 6 visual or haptic objects and required to match the soundscape to one of them, either by responding 1-6 on the keyboard (visual) or verbalising a response (tactile). Again a repeat feature was available to listen to the soundscape again prior to responding.

Test Day Two. As with the training days this was a reversal in procedure. For each trial listeners were presented with six soundscapes (each repeated 4 times) and 1 visual or tactile

object. The task was to indicate which of the 6 objects had been sonified.. As in test day 1, there was no post-trial feedback. The order of test days was counterbalanced across listeners but the visual-soundscape association was always performed first.

3.3.0. Results.

The primary objective of the experiment was to evaluate auditory object recognition, at increasingly coarse resolutions, using a VA SSD. I was also interested in whether the temporal and spectral composition of the stimuli were influential in successful object recognition, and finally, in the initial training sessions, if empty or filled objects and different duration scan rates would elicit superior performance.

3.3.1. Object Resolution – Visual/Soundscape.

Figure 3.2. and Table 3.1. shows performance accuracy (%) as a function of resolution for the visual/soundscape matching condition. The means and standard deviations for each resolution category are displayed in Table 3.1. While successful recognition was better than the 6AFC chance level of 16.67% for all resolutions ($p<0.05$), implying successful use of the device irrespective of object resolution, there was a significant difference between the performance in the four categories, ($F(3,48)=28.686$, $p<0.001$, $\eta p^2=0.642$) . Bonferroni corrected planned contrasts showed that the highest resolution, 32x32 was better recognised compared to 4x4 (M=26.471%, 95% CI [17.69,35.25], $p<0.001$), but not compared to 16x16 ($p=0.988$) or 8x8 ($p=0.556$). Performance on the 16x16 resolution was superior to 4x4 (M=25.000%, 95% CI [12.99,37.01], $p<0.001$) but not 8x8 ($p=0.974$). The final contrast demonstrated that recognition of stimuli at 8x8 was significantly better than 4x4 (M=21.732%, 95% CI [11.59,31.87],$p<0.001$).

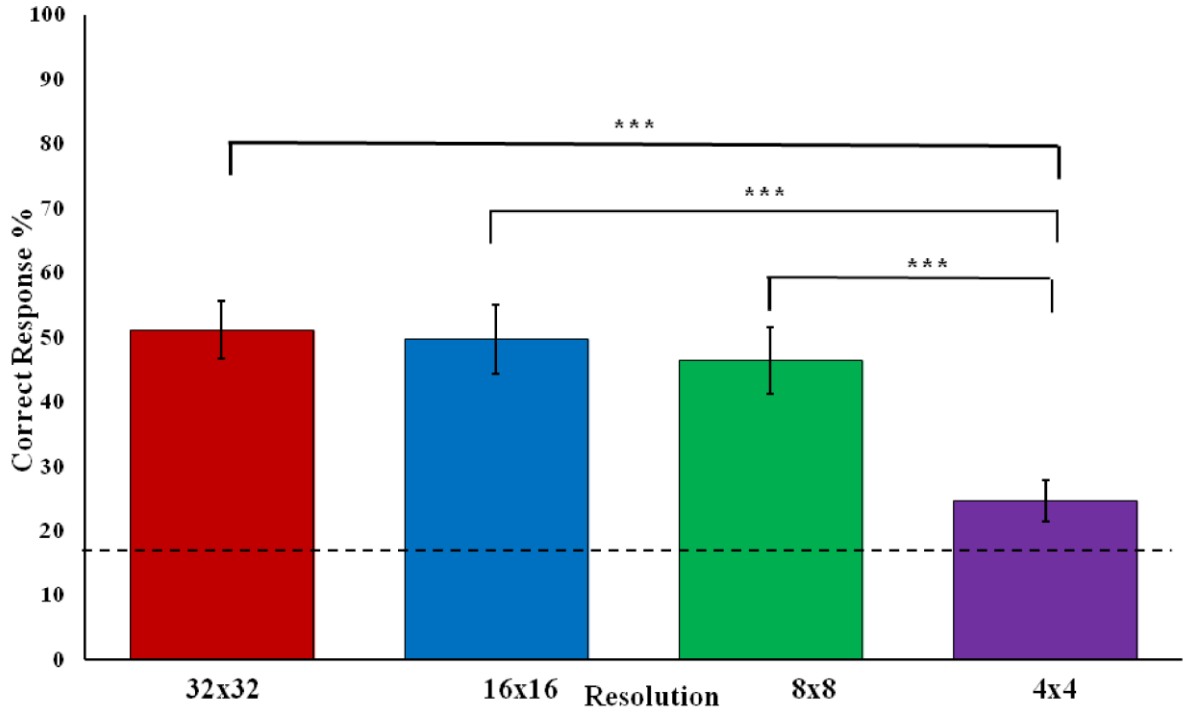

Figure 3.2.: Successful object recognition in the visual-to-auditory→ visual matching condition based on object resolution. The dashed line represents what would be expected by chance. Contrast bars indicate significant differences between conditions with error bars displaying ±1 SEM.

*** significant at <.001

Table 3.1.: Mean correct scores (%) and standard deviations in the visual-to-auditory→visual matching and the visual-to-auditory→tactile matching tasks. Results given for individual resolutions and total by modality.

Resolution	Visual (mean%)	Visual (SD)		Tactile (mean%)	Tactile (SD)
32 x 32	51.14	18.08		56.62	23.54
16 x 16	49.67	21.89		48.16	20.46
8 x 8	46.41	21.40		35.64	16.31
4 x 4	24.67	13.17		20.39	12.13
Total	42.61	17.15		39.71	14.67

3.3.2. Object Resolution – Tactile/Soundscape.

Figure 3.3. and Table 3.1 show the results for the tactile/soundscape matching condition. Performance was above chance for the three higher resolution stimuli but, unlike the visual matching condition, not for the 4x4 ($t(16)=1.269$, $p=0.223$,$d=0.635$). There was a significant main effect of resolution on tactile – soundscape matching ($F(3,48)=23.019$,$p<0.001$, $\eta p^2=0.590$) with the 4x4 soundscapes poorly matched compared to 32x32 (M=36.225%, 95% CI [21.09, 51.37],$p<0.001$), 16x16 (M=27.770%, 95% CI [12.94, 42.60],$p<0.001$), and 8x8 (M=15.248%, 95% CI [4.87, 25.63],$p=0.003$), demonstrating that recognition of the lowest resolution soundscapes was difficult irrespective of object modality. Unlike the visual matching condition where performance varied little above the ceiling effect of the 8x8 trials, there was a distinct advantage for the higher resolution objects in the haptic condition: recognition in 32x32 was better than 8x8 (M=20.978%, 95% CI [4.61, 37.34],$p=0.008$),and 16x16 , although not quite at significance for the latter ($p=0.059$)

T-tests were performed to compare 'visual' and tactile conditions for each resolution. Tactile performance at the highest resolution was better than its visual counterpart, although non-significant ($p=0.113$). Visual recognition was superior for the other 3 resolutions, with this difference significant at 8x8 ($t(16)=3.272$,$p=0.005$, $d=0.794$) but not for 16x16 ($p=0.740$) or 4x4 ($p=0.118$).

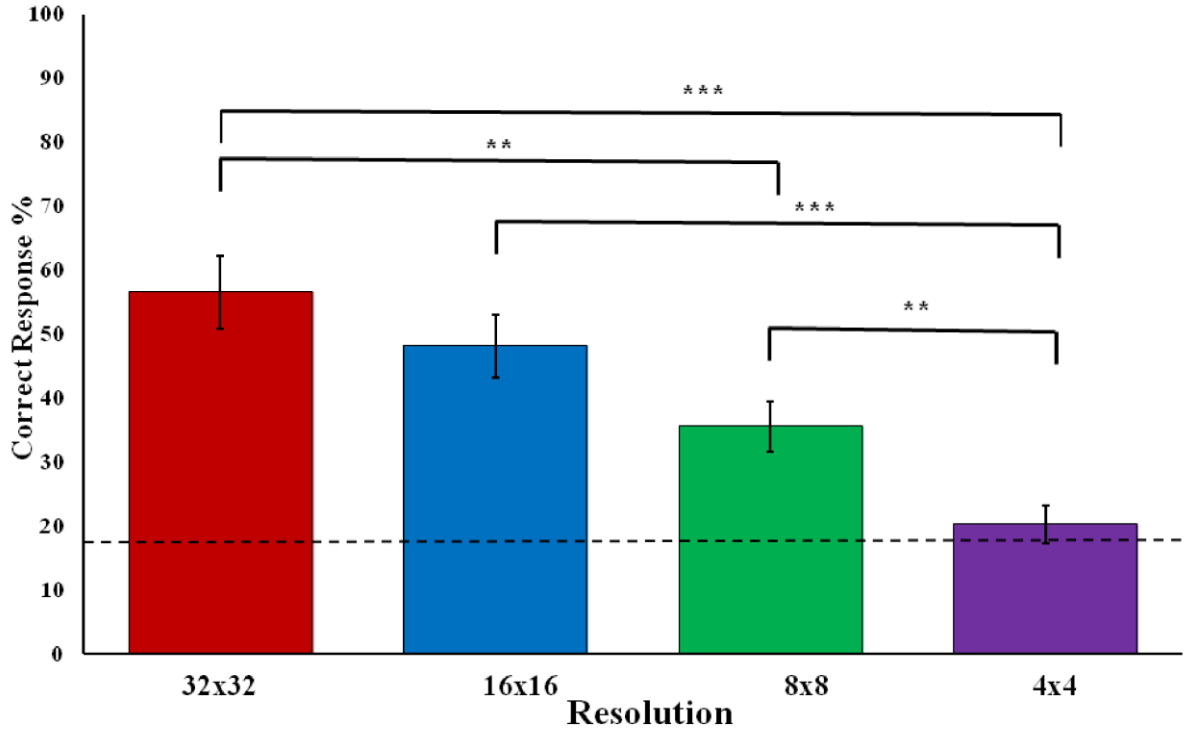

Figure 3.3: Successful object recognition in the visual-to-auditory→ tactile matching condition based on object resolution. The dashed line represents what would be expected by chance. Contrast bars indicate significant differences between conditions with error bars displaying ±1 SEM

** significant at <.01, *** significant at <.001

3.3.4. Object Type.

The secondary analysis considered object recognition as a function of stimulus type. Three of objects were classified as 'long' and the other three as 'short' based on the temporal duration of the signal. The latter group also were composed of a wider range of frequencies compared to the former. Figure 3.4. and Table 3.2. show the results for the individual objects. Collapsed across the two categories (long + short) there was no significant difference between 'long' (M=44.20%, SD=17.34 and 'short' (M=41.42%, SD=19.13) in the visual matching task (t(16)=0.969,p=0.347, d=0.235). In the haptic condition, recognition for objects in the 'short' category (M=44.51, SD=17.91) was superior to those in the 'long' category (M=35.29, SD=13.15; (t(16)=3.417,p=0.004, d=0.860).

3.3.5. Object Type – Individual Objects.

To find the source of these differences, the individual objects were analysed looking at both intra and intergroup comparisons. In the visual condition there was an overall main effect of object type ($F(5,80)=3.543, p=.006, \eta p^2=0.181$) with intragroup differences between cup versus pear (short;=0.014) and dog versus car (long;=0.009). Intergroup contrasts demonstrated performance differences for dog versus pear (p=0.006), horse versus pear (p=0.034) and a borderline effect for cup versus car (p=0.057).

There was also a main effect of stimulus type in the tactile/soundscape matching condition ($F(5,80)=4.053, p=0.002, \eta p^2=0.202$) with contrasts showing intragroup differences for dog versus horse (p=0.026), dog versus car (p=0.02) and a borderline result in the apple versus pear (p=0.067). Intergroup contrasts in this condition were significant for cup versus horse (p=0.007), cup versus car (p=0.034), apple versus car (p=0.003), apple versus horse (p=0.003) and borderline for pear versus horse (p=0.055).

Table 3.2.: Mean correct scores (%) and standard deviations for individual object recognition. Percentages are given for each object and a total for both the 'long' and 'short' conditions.

Object	Visual Matching			Tactile Matching	
	Mean %	S.D.		Mean %	S.D.
Pear	32.84	27.20		38.24	25.55
Apple	42.16	18.39		48.82	25.95
Cup	49.27	21.71		46.47	18.69
'Short' category	41.42	19.13		44.51	17.91
Dog	48.78	22.62		44.61	17.66
Horse	46.81	15.20		28.88	17.37
Car	37.01	24.20		31.55	20.86
'Long' category	44.20	17.34		35.29	13.15

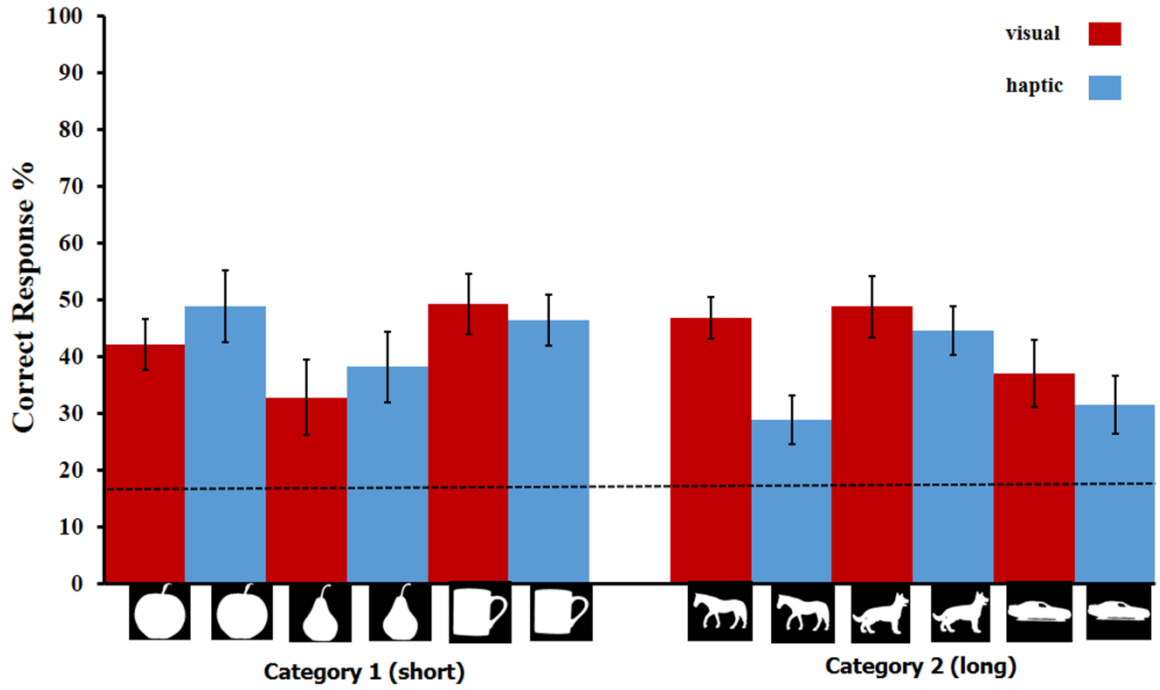

Figure 3.4.: Successful object recognition for each individual object in both visual-to-auditory→ visual matching and visual-to-auditory→ tactile matching. Objects are categorised into 'long' and 'short' conditions based on the temporal length of the active part of the soundscape. The dashed line indicates what would be expected by chance, with error bars displaying ±1 SEM.

3.3.6. Procedure Comparison.

The final analysis in the test phase contrasted performance over the two test sessions. Training effects would suggest superior performance for day two. Conversely I found overall performance on the second day (M=39.59%, SD=18.15) to be worse than day one (M=42.72%, SD=14.39) although not reaching significance (t(16)=1.447,p=0.167, d=0.351). If this comparison is made with the data divided by stimulus type, visual performance on day one (M=45.18%, SD=16.66) is significantly better than for day two (M=40.03%, SD=18.63) (t(16)=2.492, p=0.024, d=0.604) but this is not found for the tactile condition (t(16)=0.333, p=0.744, d=0.081). The two test days differed in the presentation of the 6AFC. On day one the listeners was presented with 6 visual/haptic objects and 1 soundscape. This method of

presentation is clearly less problematic to the listener than if given 1 object and 6 soundscapes, as on day 2.

3.3.7. Training

The structure and stimuli in the training regime allowed us to evaluate device settings in naïve users. Objects were either filled, where the whole object was white, or empty, where only the object outline was in white. Device scan rates were either 1 second or 2 seconds to give four stimulus conditions. Table 3.3. displays the mean performance for these conditions. For visual/soundscape matching analysis of variance showed a main effect of performance as a function of condition ($F(3,54)=4.366, p=0.008, \eta p^2=0.195$). Bonferroni corrected contrasts found no significant pairwise comparisons. However trends suggested that the 1 second filled stimuli were poorly recognised compared to 2 second filled ($p=0.059$), and 2 second empty ($p=0.061$) implying that the time scan may have had some effect. Analysis on this data collapsed into 'time scan' and 'filled/empty' groups showed that performance on the 2 seconds scan rate ($M=64.31\%$, $SD=14.22$) was superior to its 1 second counterpart ($M=57.81\%$, $SD=10.90$), ($t(18)=2.914, p=0.009, d=0.668$) but not reaching significance for filled ($M=62.66\%$, $SD=12.53$) vs empty ($M=59.46\%$, $SD=12.81$) shapes ($t(18)=1.438$, $p=0.168$, $d=0.330$)

These results contrast with those of Brown et al (2011) who evaluated different vOICe device settings in object recognition and found no significant advantage for the 2 second scan speed over the 1 second. This can be attributed to paradigm differences with the former using the device in real time with real objects at multiple perspectives and the later utilising sonified 2 dimensional images. There is clearly an advantage to a slower scan speed if the objects are simple and the soundscape consistent over time.

Table 3.3.: Mean correct scores (%) and standard deviations for the different conditions in the training phases of the experiment.

	Visual Matching			Tactile Matching	
	Mean %	S.D.		Mean %	S.D.
1 second filled	60.86	11.48		67.43	11.66
1 second empty	54.77	14.60			
2 second filled	64.47	15.84			
2 second empty	64.14	14.71			
Filled total	62.67	12.53			
Empty total	59.46	12.81			
1 second total	57.81	10.90			
2 second total	64.31	14.22			

3.4.0 Discussion.

In this study I evaluated object recognition performance in naïve users of a VA SSD, The vOICe. Images, and their soundscapes, were manipulated by pixel resolution to ascertain the minimal amount of visual/tactile/soundscape information that is needed for successful recognition. As secondary considerations I looked at the spectral/temporal composition of the stimuli and presentation order within the 4AFC as factors in recognition, and replicated various device settings in training to assess for any preference. The results demonstrate a lower ceiling effect of 8x8 (64) pixels in both the visual-VA and tactile-VA conditions for object resolution. While this is informative for structuring effective training regimes it also allows postulations on cortical representation of sonified objects.

In both invasive and non-invasive SSD systems the central ‘visual’ system (i.e., cortex) is implicated in the processing of visual objects. Imaging studies have demonstrated the recruitment of ‘visual’ areas in VA SSD use, even in naïve users (Arno, De Volder, et al., 2001; Poirier, De Volder, et al., 2006) with transcranial magnetic stimulation (TMS) to visual

cortex impeding pattern recognition tasks using SSDs (Collignon et al., 2007). Output from The vOICe also shows activation in areas of lateral occipital cortex, an area not associated with auditory input, implying that the 'auditory' signal from the device is not only processed in the auditory pathway (Amedi et al., 2007; Haigh et al., 2013; Plaza, Cuevas, Grandin, De Volder, & Renier, 2012). This is further corroborated by evidence of a correlation between musical ability and performance using a VA SSD (Haigh et al., 2013). This leads to the further question: are the limits of such systems to be found in auditory or visual neural circuits?

If auditory object recognition is a limiting factor, then information processing in primary auditory cortex is crucial; phase locking in auditory cortex is limited to around 30Hz, thus I would expect a ceiling effect at the 16x16 image resolution (Barton et al., 2012). However, the ceiling effect at 8x8 pixels instead suggests that object recognition is processed further up the auditory pathway, such as in the superior temporal gyrus (STG) where phase locking is reduced to <16HZ. This is consistent with performance by higher cortical representations optimized for speech processing (Pasley et al., 2012). The implications of this are that the pre-lexical, higher-cortical object-based representation constitutes the ultimate token that allows the listeners to recognize a rendered object and places strict limits on the potential success of the substitution system, and subsequent processing in visual or supramodal cortical areas. This does not mean that these limits, as implicit in the use of a higher cortical speech processor, negate the viability of SSDs and indeed may be circumvented by building crossmodal networks at the earlier level of primary cortex (or even midbrain). Extensive training and learning on the devices might, via synaptic plasticity, produce crossmodal networks capable of exploiting earlier, wider-bandwidth representations thus bypassing the limitations of the speech processor. Indeed recruitment of higher multisensory processing cortical areas, such as the STG, may be key in allowing information transfer between primary

sensory areas thus giving rise to higher fidelity information processing and even visual imagery in some long term device users (Proulx et al., 2014; Ward & Meijer, 2010b).

The ceiling effect at 8x8 draws interesting comparisons with Weiland and colleagues (2005) simulations for retinal implants. Their estimation of a 30x30 electrode/pixel array being a requisite for face recognition and text reading may be overstated. While noting I was comparing invasive and non-invasive techniques and different paradigms, the 8x8 ceiling with minimal improvement at higher resolutions, implies the brain can extract enough salient information from coarse SSD input for effective object/pattern recognition.

The 8x8 ceiling effect may also have been influenced by how the image resolution was reduced, and the subsequent soundscapes. With reference to Figure 3.1a, for the 32x32 and 16x16 images there is a distinct contrast used – white images (maximum volume) on a black background (silence) – and therefore recognition is based on frequency and temporal features only. However the image reduction for 8x8 and 4x4 introduces grey pixels of various shades bringing amplitude into the processing. At 4x4 this is considerable with only 1 or 2 pixels at maximum volume and therefore unsurprising that recognition is poor. At 8x8 grey/quiet pixels are very much to the periphery with the loud pixels retaining the basic shape.

As well as being affected by resolution, object recognition was also influenced by stimulus type (visual/tactile), stimulus features (long/short temporal length), and task procedure. The soundscapes in both the visual and haptic matching tasks were identical and therefore any performance differences can be attributed to modality-specific difficulties in object identification rather than processing of the SSD signal. Unsurprisingly, visual/soundscape matching was more successful than the haptic counterpart. All listeners were sighted and therefore their primary modality for ‘everyday’ object recognition is vision.

Visual object recognition utilises a number of cues such as shape, luminance, depth, motion, shading and colour which are processed in parallel to allow a rapid identification of the object, usually in about 1 second (Martinovic, Gruber, Hantsch, & Muller, 2008) Object recognition via haptics is less rapid and usually serial (Overvliet, Smeets, & Brenner, 2007b) as individual object features have to be explored sequentially, committed to memory, and mentally reassembled to give a percept of the object (Craddock & Lawson, 2008). If time based haptic exploration is slower (and logic dictates that larger objects require more exploration time), then the advantage for 'short' objects in the haptic condition, compared to 'long', is understandable. This would be salient if a time limit was placed on the trial forcing object identification to be rapid. In the present experiment there was no 'official' time limit placed on the task, but having completed the more rapid 'visual' task first listeners may have responded in the haptic task at a speed familiar to the procedure.

The procedure was certainly a main effector on the results. On Test Day One all stimuli in the trial (all visual/haptic objects + 1 repeated soundscape) were presented to the listeners 'online' simultaneously for the duration of the trial. Visual-auditory feature matching and, saliently, comparison between features of different objects can be done quickly with little memory load. On Test Day Two the visual/haptic object is available for the trial duration but the 6 soundscapes are sequentially presented. Feature matching, particularly comparisons, requires memory load in the retention and recall of previous soundscapes. While all 6 tactile objects on day one are 'available' to the listeners for the duration of the trial, haptic exploration is still serial as all objects cannot be haptically explored concurrently.

The level and duration of visual impairment in the target group may also be influential on the ability to use different levels of resolution in sensory substitution. While the data collected on sighted listeners may be extrapolated to inform sensory augmentation (e.g. expansion of the

field of view), where the device is not substituting for an impaired sense but providing additional information to a fully functioning perceptual system, processing differences in late, and particularly, congenitally blind users, may elicit different results. Behavioural and neural differences between sighted, late and congenitally blind have been demonstrated for, amongst other things, false memories, the mental number line, and spatial representations (Pasqualotto, Lam, & Proulx, 2013; Pasqualotto, Taya, & Proulx, 2014). Pasquallotto and colleagues found in a spatial task that while sighted and late blind showed a preferential use of an object-based or 'allocentric' reference frame, the congenitally blind preferred a self-based 'egocentric' reference frame (Pasqualotto, Lam, et al., 2013). This corresponds with ideas that at least some visual experience is a requisite of developing multisensory neurons, spatial updating tasks, multisensory integration and higher cognition (Pasqualotto & Proulx, 2012; Reuschel, Rosler, Henriques, & Fiehler, 2012; Wallace, Perrault, Hairston, & Stein, 2004). With two algorithm principles coding spatial factors and multisensory integration integral in SSD use, task based comparisons between the three should feature heavily in future research.

The results of the present study feed directly into theories regarding standardization of working resolutions across devices. SSDs are limited in the information they can convey by their conversion algorithms; that is, three principles can only transmit three aspects of visual perception. One way to overcome this is to utilise numerous SSDs (VT + VA) or a combination of invasive and non-invasive devices. Should we establish a consistent working resolution across devices to develop effective training protocols that maximise the effectiveness of multiple device use? A functional limit (24x24) for basic object recognition has been ascertained for retinal implants (Li et al., 2012). If it holds that successful object recognition can be achieved at lower resolutions in SSDs then this informs on the use of each device in an invasive/non-invasive combination, i.e the SSD for fine grained recognition and

the implant for more coarse spatial/navigation information. A final consideration in applying these results to developing training protocols is ‘how high a resolution is sufficient/desirable for successful object recognition in sensory substitution?’ As stated by Paul Bach-y-Rita

“A poor resolution sensory substitution system can provide the information necessary for the perception of complex images. The inadequacies of the skin (e.g. poor two-point resolution) do not appear as serious barriers to eventual high performance, because the brain extracts information from the patterns of stimulation. It is possible to recognise a face or to accomplish hand-eye coordinated tasks with only a few hundred points of stimulation.”

Pg 543 (Bach-y-Rita & Kercel, 2003).

If the brain is able to extract enough salient information from low resolution input to disriminate objects, the provision of more complex objects at early stages of training, as alluded to in Chapter 2, may be valid. The extra information in the high resolution images has no negative effect on recognition, compared to the lower level counterparts, giving little reason to remove it. However, provision of high levels of information may be advantageous in that it gives more data to extract salient feature information from. Of course there may be an upper limit in which the amount of information in the signal hinders recognition and this is evaluated in Chapter 5.

In conclusion, I have demonstrated an apparent resolution ceiling effect (8x8 pixels) in which successful object recognition is possible in naïve users of a VA SSD and postulated that in such users the ascending auditory hierarchy may place limitations on such a task. Further research should be undertaken to evaluate how this can be extrapolated to extensively trained users, late and congenitally blind users and situations in ‘real time’. A more comprehensive understanding of this would allow the development of more effective training

protocols for sensory substitution and give a better understanding of the associated brain processes

Chapter 4

In Chapter 3 I showed that while simple object recognition is possible with degraded input, phase locking at different frequencies limits object formation to different levels of the auditory hierarchy. In Chapter 4 I further analyse the potential theoretical limitations of the algorithm based on principles of auditory scene analysis, primarily proximity and harmonicity. The algorithm requires concurrent processing of the auditory representations of horizontally spatial information to segregate features into independent auditory objects. A failure to segregate these features will result in potential misidentification of the object.

Considering integrated audio-visual information has a positive impact on perception, in the second part of the experiment congruent and incongruent audio-visual information is used to assess whether this reduces and potential conflicts in the auditory stream processing.

Visual objects in the auditory stream: Auditory scene analysis in sensory substitution.

David J Brown [1,2]

[1] Biological and Experimental Psychology Group, School of Biological and Chemical Sciences, Queen Mary University of London.

[2] Department of Psychology, University of Bath.

Abstract.

A critical task for the brain is the sensory representation and identification of perceptual objects in the world. When the visual sense is impaired, hearing and touch must take primary roles and in recent times sensory substitution devices (SSD) have been developed that employ the tactile or auditory system as a substitute for the visual system. Visual-to-auditory devices provide a complex, feature-based auditory representation that must be decoded and integrated into an object-based representation by the listener. However, we don't yet know what role the auditory system plays in the object integration stage and whether the principles of auditory scene analysis apply. Here I used a well-established visual-to-auditory SSD to test whether auditory feature-based representations of visual objects would be confounded when their features conflicted with the principles of musical harmonic grouping. I found that listeners (N = 36) performed worse in an object recognition task when the auditory feature-based representation was harmonically consonant. I also found that this conflict could be partially suppressed by using congruent visual cues. The findings suggest that early auditory processes of harmonic grouping dominate the object formation process and that the auditory-to-visual SSD may require modification.

4.1.0 Introduction.

Our sensory systems provide a rich coherent representation of the world through the integration and discrimination of input from multiple modalities (Spence, 2011). These low-level processes are modulated by high-order processing to selectively attend to task relevant stimuli. For example to attend to a speaker at a cocktail party we must select the low-level acoustic features that are relevant to the target, that is the person you are speaking with, from the environmental noise (Cherry, 1953). To accomplish this, feature-based sensory representations must be recombined into object-based representations in a rule based manner. In visual perception this is through scene analysis. Visual input is grouped into distinct objects based on Gestalt grouping rules such as feature proximity, similarity, continuity, closure, figure ground, and common fate (Ben-Av, Sagi, & Braun, 1992; Driver & Baylis, 1989). Similarly, there are rules that govern the arrangement of low-level audio input into 'auditory' objects. This process is called auditory scene analysis (ASA). Grouping in ASA is either at a temporal or melodic level and governed by proximity or similarity in time, pitch/loudness continuation, or at spectral levels including a common fate, coherent changes in loudness, frequency or spectral envelope, or harmony (A.S. Bregman, 1994)

Shape and contour are crucial for the organisation and recognition of visual objects. In parallel to this the temporal contour of a sound, known as its envelope, is critical at recognizing and organising auditory objects (Sharpee, Atencio, & Schreiner, 2011). Visual-to-auditory SSD code visual characteristics (brightness, spatial position) into auditory ones (pitch, loudness, temporal and stereo scan) to convert visual features to 'auditory' objects (P. Meijer, 1992). It is therefore intuitive to assume that when mapping visual shapes to auditory envelopes there will be an equivalent object formation and recognition process. This is exemplified in Figure 4.1. which shows the spectrograph of a 2D visual object sonified using The vOICe, and how basic shape is retained in the auditory output.

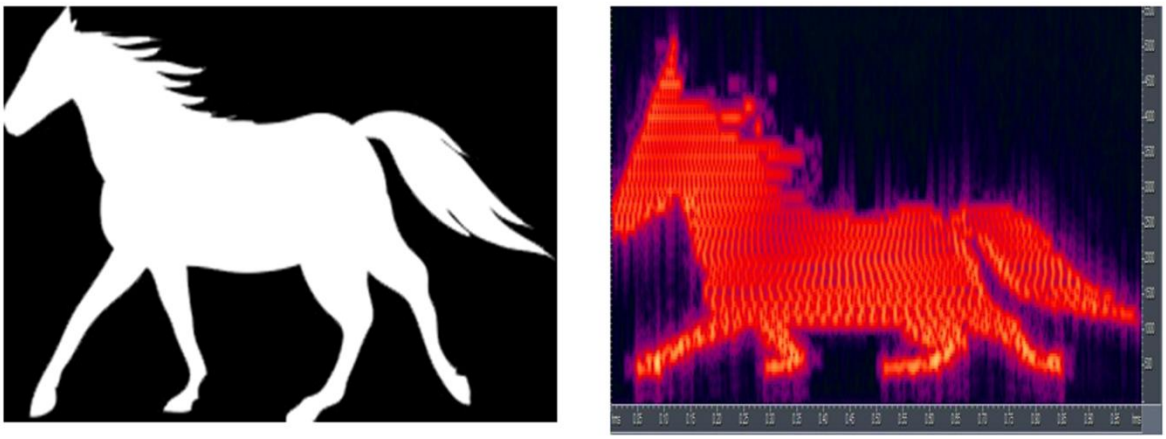

Figure 4.1. Spectragraph (right) of 2D visual object after sonification using The vOICe SSD.

In the early stages of learning to use visual-to-auditory SSD it is posited that discrimination of the signals' auditory characteristics are salient (D. Brown, T. Macpherson, & J. Ward, 2011) as insufficient time has elapsed to elicit the crossmodal plasticity attributed to long term use (Proulx et al., 2014). It is therefore a logical jump to infer that successful visual object formation in The vOICe would be modulated by the rules that govern ASA, with a failure to segregate auditory objects translating to a reduction in task based performance.

This could be exemplified in tasks such as the recognition of alphanumeric characters. These stimuli are ideal for training as they are simple, have defined features, are familiar in a 'known' category, and are commonly used in tests of visual acuity, in the sighted and SSD use (Haigh et al., 2013; Levy-Tzedek, Riemer, & Amedi, 2014; Striem-Amit, Guendelman, et al., 2012). However the structure of some characters could be susceptible to confounds due to ASA rules. For example, a sonified letter 'E' consists of four lines: one vertical and three horizontal. If there are harmonic relations, and subsequent segregation failure, between the tones representing the three horizontal lines then this may lead to a misidentification of the character. Consonance between the middle and bottom line could result in object

identification as an ‘F’ as these two lines would be perceived as one. Visually, segregation of these lines is non-problematic therefore a feature segregation failure must be in the auditory processing of The vOICe output signal. If we consider this signal as auditory, then the principles of ASA may be salient to this task.

Is The vOICe output auditory, or visual? Long term users describe a visual experience (Ward & Meijer, 2010a) and activation is found in typical visual areas (Amedi et al., 2007; Striem-Amit & Amedi, 2014). However, prior to training activation is only in cortical areas associated with the substituting modality (audition).

I tested potential limitations in feature segregation due to ASA using The vOICe, which renders horizontal visual lines as tones that are frequency consistent over time. Thus for two parallel lines, the signal output is concurrent tones at different frequencies. In the experimental design the duration of the tones was consistent and therefore perception of the lines as segregated objects would be dependent on spectral (frequency) principles of ASA. I was interested in two principles: harmonicity (consonant and dissonant stimuli), and proximity, that is distal features more likely to be segregated. Both of these principles were manipulated in the design in which the listeners were required to indicate whether they heard the sonification of one or two visual lines. The stimuli were designed to test both proximity and harmonicity in the same task with a second type of one-line stimuli to evaluate the weighting of each principle, that is the filled single line stimuli had the same top and bottom frequencies as their parallel line counterpart (proximity) but being visually and aurally ‘filled’ lacked the interval (harmonicity).

I hypothesised that there would be a general linear improvement in discrimination as the gap between parallel lines increased based on proximity, and that any potential conflicts would be for tones showing harmonic consonance.

Experiment 1

4.2.0 Method

Listeners

I recruited 36 listeners (28 female) via an Undergraduate Research Assistant module. Listener age ranged from 18 to 25 years old (M=20.17, SD=1.30). All listeners provided informed written consent, and had normal or corrected eyesight, normal hearing and educated to undergraduate level. Four listeners self-reported as left handed and all were naïve to The vOICe and the concept of sensory substitution. The study was approved by the University of Bath Psychology Ethics Committee (#13-204).

Materials and stimulus design.

Visual stimuli were created using Adobe Photoshop CS 3.0. and sonified using The vOICe sonification feature at default settings (1 second scan rate and normal contrast, but with foveal view and high contrast off). Cool edit Pro 2.0 was used for frequency analysis of sonified stimuli. Auditory and visual stimuli were presented in E-Prime 2.0 running on a Windows 7 PC with the output signals transmitted to the listeners via Sennheiser HD 585 headphones.

Stimulus design.

Visual stimuli were created in Adobe Photoshop CS 3.0. A black background, dimensionally consistent with the resolution of The vOICe 'visual' field, was created in Photoshop. A grid of 48 x 1.5 pixels rows was overlaid across this. Two horizontal white lines, each 1 pixel row thick and full background width, were drawn across the centre (y axis) of the image. For subsequent 'parallel line' stimuli each of the lines was moved 1 row up or down effectively doubling the width of the gap. Filled line stimuli were created in an identical manner except

the gap between the two lines was filled with white pixels (noise) rather than black (silence). The single line stimuli consisted of a double width horizontal white line, moved to 24 points on the y axis to give the stimulus set. In total there were 23 parallel line (the first parallel line stimulus with no gap was classified as a single line), 24 single line, and 24 filled line visual stimuli. The 71 sonifications were normalized and frequency analysed in Cool Edit Pro 2.0 to predict roughly which line pairs would show consonance. This was done by assigning musical notation theory to the frequencies and looking for octave, perfect 4^{th}'s and 5^{th}'s and major and minor 3^{rd}'s and 6^{th}'s.

Procedure.

Listeners watched a PowerPoint presentation that gave a brief overview of The vOICe including audio-visual (AV) examples of how the sonification algorithm converted information. Examples of parallel, filled, and single line stimuli were given along with example trials to explain the task procedure.

For each trial the listener was presented with a soundscape which was created from either 1 or 2 horizontal lines. The task was to indicate using the PC keyboard whether it was a 1 line (single or filled) or a 2 line stimulus. There was no post-trial feedback given. Each of 4 blocks consisted of 94 randomised trials (46 parallel, 24 filled, 24 single) for a total of 386 trials.

4.3.0 Results.

First I analysed the data for the correct performance in recognising parallel lines. Figure 4.2. shows correct performance for each of the frequency range gaps. The omnibus main effect showed superior discrimination as a function of inter-tonal gap ($F(8.52, 298.04)=21.937, p<0.0005, \eta p^2=0.385$) in that some parallel line pairs were more often recognised as such. It also broadly fitted with the pattern of consonance and dissonance in the

stimulus frequency analysis and showed a distinct pattern of successful discrimination (above 50%) followed by ranges of integration (below 50%). These were grouped into consonant and dissonant categories.

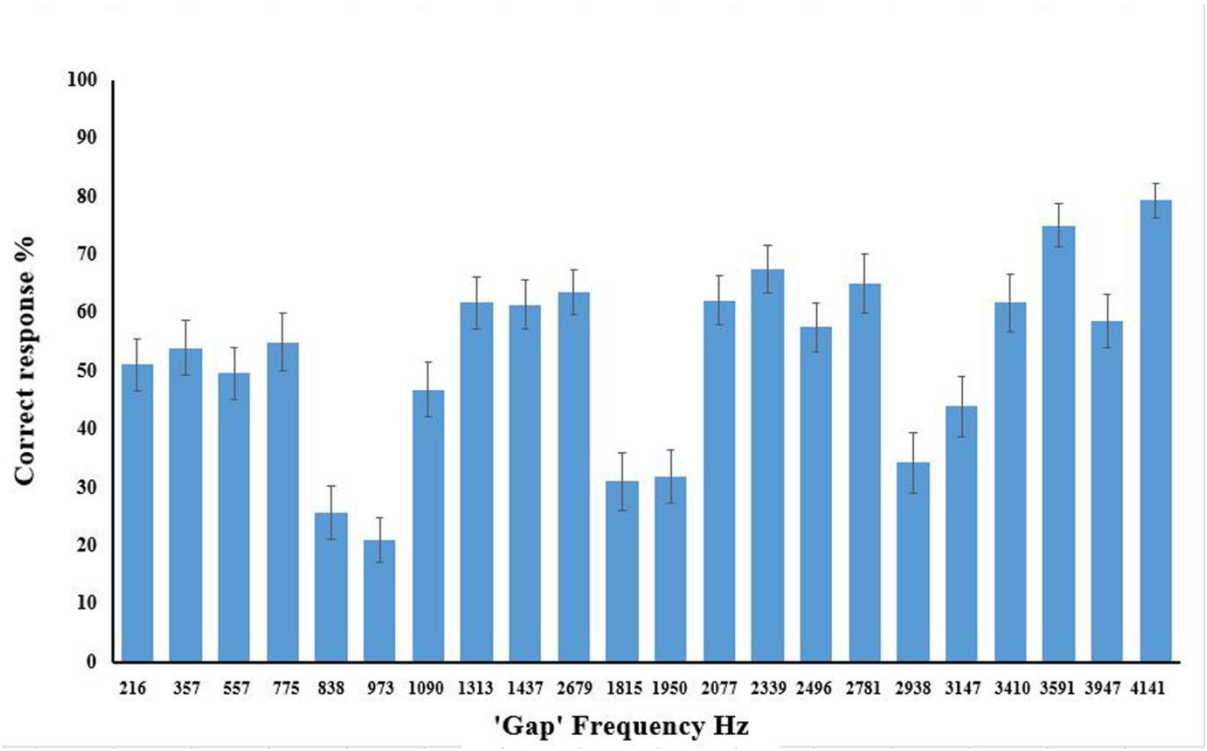

Figure 4.2.: Correct response to parallel line stimuli for each frequency gap prior to categorization into consonant and dissonant groups. Error bars show ±1 SEM

The overall effect of consonance/dissonance on parallel line feature segregation, irrespective of gap size per se, is shown in Figure 4.3. Segregation of the two tones was significantly easier when the signals showed dissonance (M=59.48, SD=19.54) compared to consonant (M=30.73, SD=22.86) features, ($t(35)=9.513, p<0.0005, d=1.58$). Analysis of variance was run the seven frequency categories and showed a significant main omnibus effect ($F(3.19,111.52)=42.182, p<0.0001, \eta p^2=0.547$) with parallel line performance for these groups shown in Figure 4.4. and Table 4.1.

Table 4.1.: Correct performance for all parallel and filled line presentations in audio, audio-visual congruent and audio-visual incongruent. Consonant stimuli are marked in blue

Category Interval	Bottom Line	Top Line	Audio Mean correct (%)	S.D.	AV Mean congruent (%)	S.D.	AV Mean incongruent (%)	S.D.
Parallel 498Hz	1248Hz	1746Hz	50.95	24.15	59.24	30.55	49.46	30.44
Parallel 917Hz	1119Hz	2036Hz	23.26	22.19	41.30	30.72	25.00	25.28
Parallel 1529hz	969Hz	2498Hz	57.03	22.68	71.74	23.07	54.08	26.56
Parallel 1913hz	839hz	2752Hz	30.73	26.41	36.41	27.16	30.98	28.67
Parallel 2666hz	689hz	3355Hz	61.98	21.76	75.82	19.88	59.24	25.76
Parallel 3111Hz	602Hz	3713Hz	38.19	28.96	50.54	33.39	42.39	33.87
Parallel 4047Hz	516Hz	4563Hz	67.97	19.25	75.54	23.07	55.43	28.54
Total			**47.16**	**15.52**	**58.66**	**15.18**	**45.23**	**12.05**
Filled 498hz	1248Hz	1746Hz	48.39	20.02	64.58	17.93	44.64	14.56
Filled 917Hz	1119Hz	2036Hz	41.43	25.32	65.48	18.07	43.75	24.83
Filled 1529hz	969Hz	2498Hz	50.71	26.98	60.71	16.90	44.94	23.69
Filled 1913hz	839hz	2752Hz	56.79	28.17	57.74	25.46	48.21	25.40
Filled 2666hz	689hz	3355Hz	55.18	33.30	64.58	17.60	52.38	22.05
Filled 3111hz	602Hz	3713Hz	58.21	33.48	70.24	20.34	50.59	22.87
Filled 4047hz	516Hz	4563Hz	60.89	34.04	68.45	21.51	51.79	21.02
Total			**53.09**	**6.20**	**64.54**	**3.96**	**48.90**	**4.81**

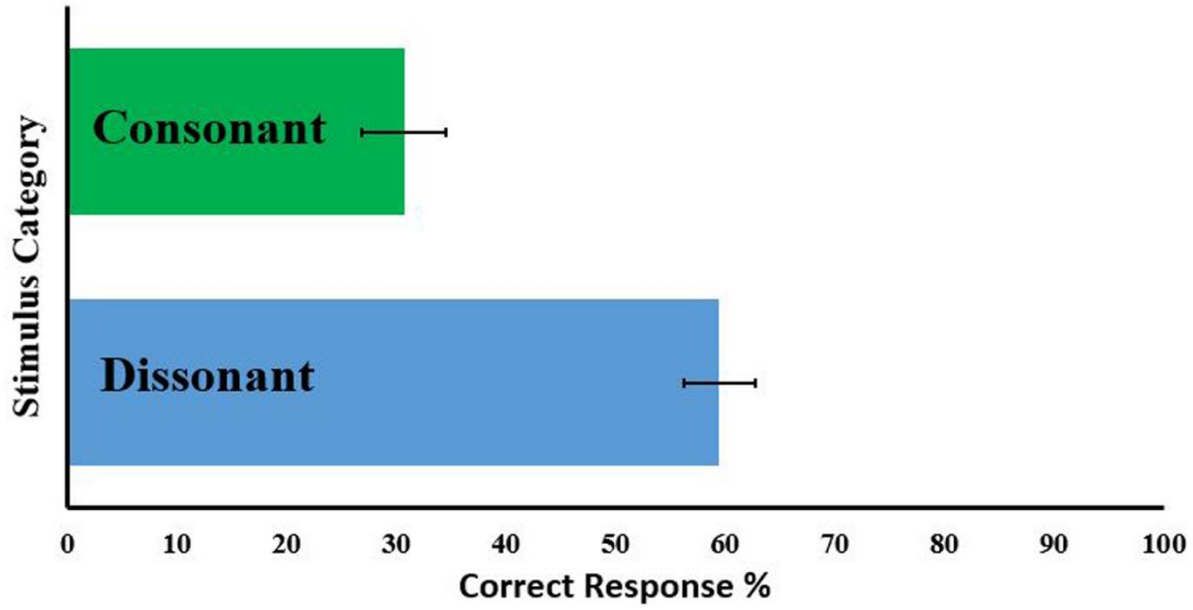

Figure 4.3.: Overall correct scores for parallel line recognition in the audio only condition as a function of consonance and dissonance. Error bars show ± 1 SEM

With the significant difference shown between the pooled consonance and dissonance data it was unremarkable that all contrasts between these two categories were significant, with dissonant sounds segregated better than consonant regardless of line ‘gap’.

This data shows confusion in feature segregation based on harmonicity but was there an independent effect of proximity? That is, are sonified lines further apart segregated more successfully when harmonicity is controlled for? Gestalt theories and ASA imply that the further apart they are the more likely to be segregated. The contrasts supported this for both consonant and dissonant conditions. When there was no harmonic interference (dissonant data) the larger gaps were more easily discriminated, although not significantly contrasted to the interval directly below it. For example discrimination for 4047Hz was better than 1529hz (M=10.94%, 95% CI [2.02,19.86],p=0.006) and 498Hz (M=17.01%, 95% CI [5.82,28.21],p<0.0005) but not for 2666Hz (p=0.174) and while 2666Hz was segregated more successfully than 498Hz (M=11.02%, 95% CI[1.10,20.95],p=0.018) it was not segregated successfully from 1529hz (p=0.944). Similarly in the consonant data 3111Hz performance was superior to 917Hz (M=14.93%, 95% CI[.77,29.09]p=0.031) but not 1913Hz (p=0.264) and there was no difference between the latter two (p=0.745).

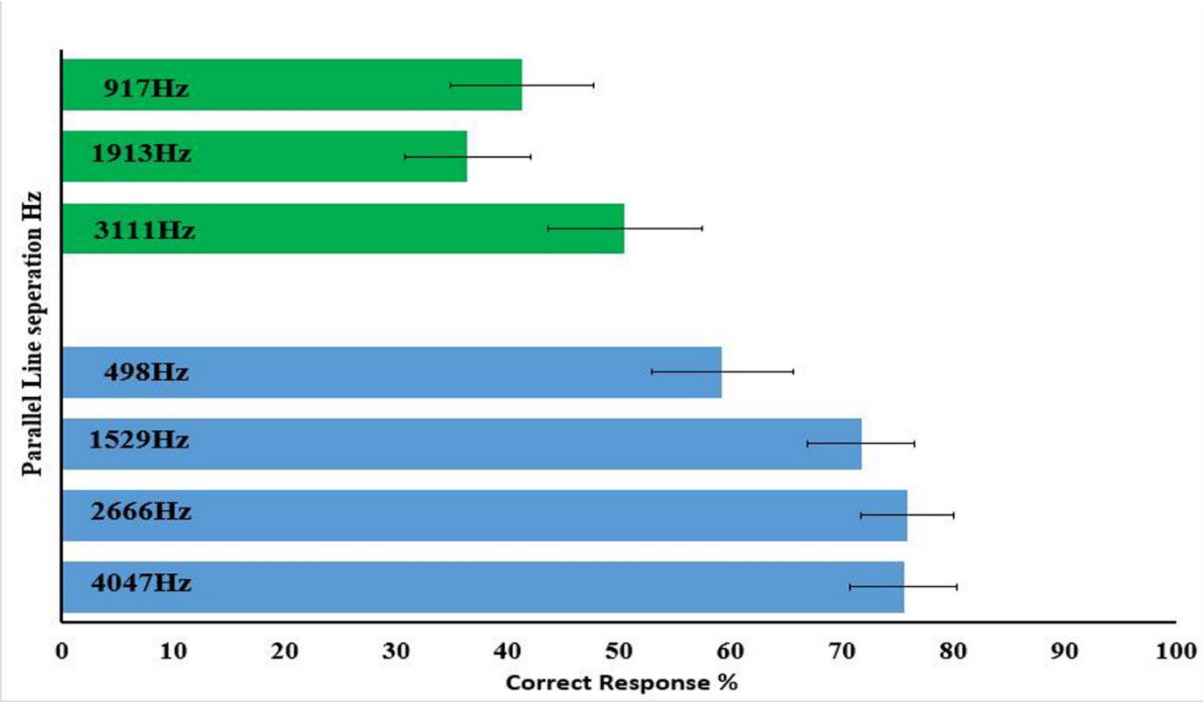

Figure 4.4.: Correct discrimination of parallel lines in the auditory only condition as a function of consonance and dissonance and size of interval 'gap' Error bars show ±1 SEM

As it seems both proximity and harmonicity influence parallel line segregation, use of the filled line stimulus modulated harmonic effects by 'filling' the interval with sonified pixels. The filled lines share spatial properties with their parallel line counterparts, that is, parallel line interval equals filled line bandwidth, so a similar categorization was used. As harmonicity is dependent on interval it was unsurprising that the filled line data was unremarkable. Figure 4.5. and Table 4.1. illustrate the lack of interference at frequency bandwidths equivalent to parallel line consonance, and thus discrimination was based on 'proximity'. This was gradual however as while there was main effect ($F(2.84, 96.54)=4.691, p=0.005, \eta p^2 =0.121$), only the largest 4097Hz bandwidth contrast with 917hz (M= 19.46%, 95% CI [0.58, 38.35], p=0.038) reached significance, implying the extra 'noise' (>1900Hz bandwidth) in the filled line stimuli allowed categorization.

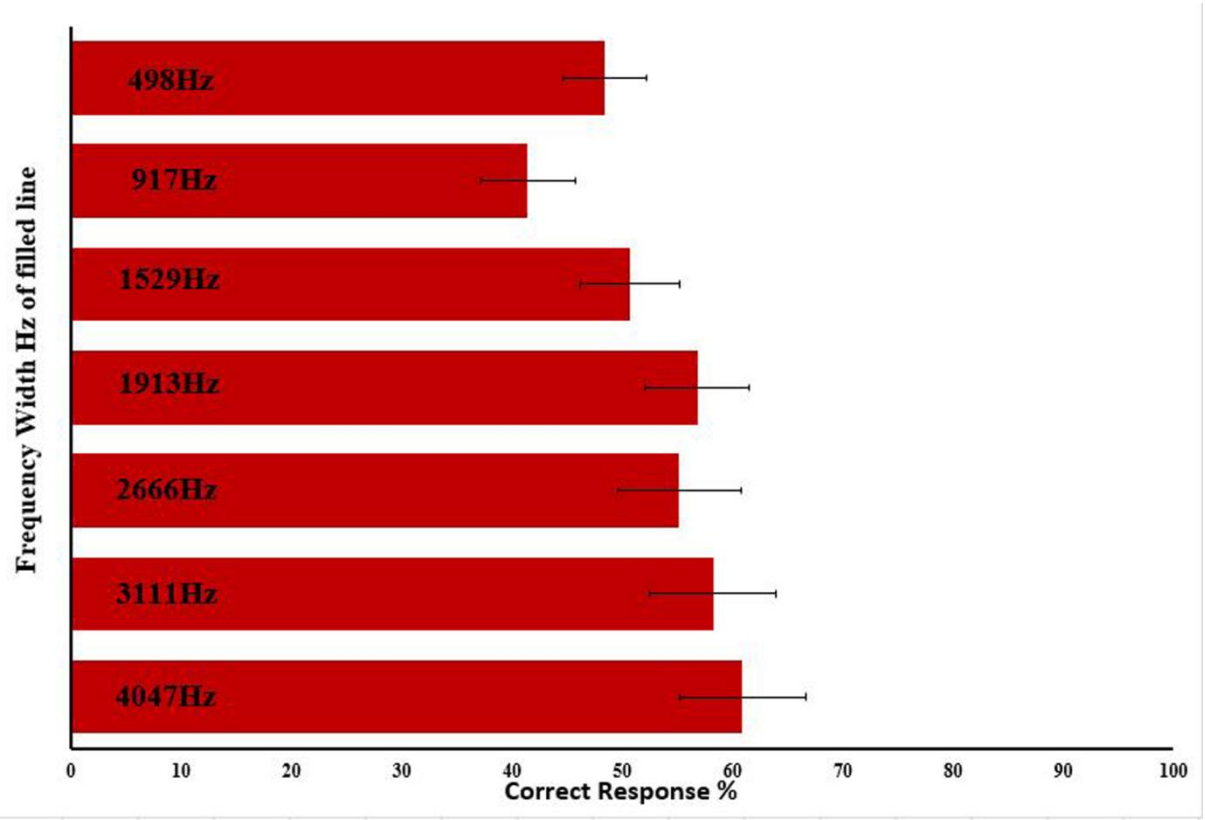

Figure 4.5.: Correct performance for the filled line stimuli in the auditory task. Error bars show ± 1 SEM.

Were filled line stimuli recognised more than the parallel line counterparts, taking into consideration the poor performance for consonance? Quite simply, no with p=.346 for all data, and p=.111 when consonance was modulated.

Experiment 2

The results from Experiment 1 demonstrate areas of confusion in the sonification of parallel lines due to consonance within some soundscapes. This is a potential limitation of the algorithm for certain situations, although these are specific. As SSD are devices for providing crossmodal information, by representing visual features in sound, could this limitation be negated by the provision of task-relevant sensory information in another modality? I tested for multisensory interactions in Experiment 2.

The provision of synchronous audio-visual information has been demonstrated to facilitate superior performance, compared to a unimodal counterpart in tasks such as visual search (Iordanescu et al., 2008) speeded classification (Ben-Artzi & Marks, 1995) with weighting of modality dependent on the nature of the task (Alais & Burr, 2004b; Driver & Spence, 1998). Considering this there should be an increase in performance if congruent visual line stimuli are presented synchronously with the soundscapes. Therefore for Experiment 2 I predicted an increase in performance for congruent audio-visual presentations (such as both hearing and seeing two lines) but no, or limited, increase in discrimination for incongruent presentations (such as hearing two lines but seeing one).

4.4.0 Method

Listeners.

I invited the same listeners back two weeks later to retake the task with the additional audio-visual component. Of the 36 participants 24 (19 female) returned, with an age range of 18-23 years (M=20.17, SD=1.01). Of this returning group only two self-reported as left handed.

Procedure.

There was no repeat of the PowerPoint presentation although listeners were asked if they remembered the task. For each trial the listener was presented with a single (single + filled) or parallel line soundscape and a simultaneous irrelevant visual line presentation. Visual lines

were either congruent (lines used in the soundscape creation) or categorically incongruent (i.e. 2 line soundscape – single or filled line visual). Categorically incongruent stimuli retained some spatial congruency in that a single line soundscape would be represented by one of the parallel lines in the visual presentation. Listeners were explicitly told that while it was a requisite to look at the screen for task timing they were NOT required to indicate how many visual lines were on the screen, but to discriminate between the audio tones as in Experiment 1. The task again was 386 trials split into 4 blocks of 96 trials.

4.5.0. Results.

First consider the effect of consonance and dissonance on successful segregation of parallel lines with concurrent visual presentation. Collapsing across congruency, there was a main omnibus effect in that feature segregation was better for dissonant (M=62.57,M=20.72) compared to consonant (M=37.77, SD=24.60) ‘gaps’ ($t(22)=5.845, p<0.0005, d=1.22$), similar to what was found in the audio-only condition. This is shown in Figure 4.6. illustrating clearly that interference in feature segregation due to auditory consonance is apparent in the audio-visual paradigm. However, when questioning whether audio-visual information can negate this harmonic interference, congruency is critical. Data was categorized by consonance as for as in Experiment 1 and subjected to the same analysis, accounting for congruency.

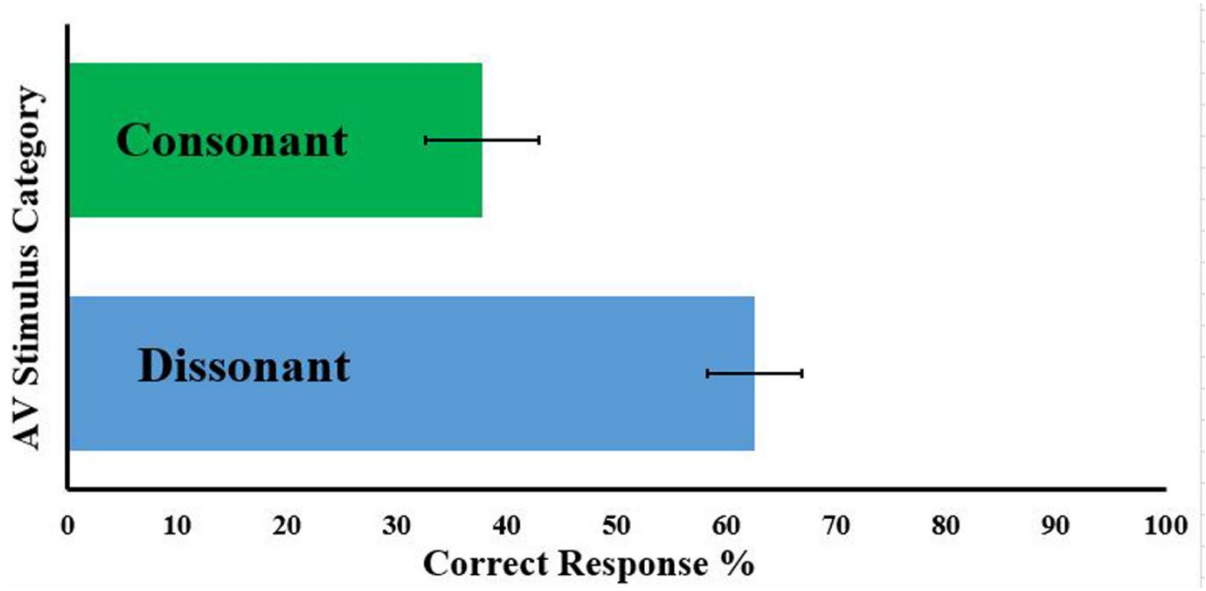

Figure 4.6.: Overall correct scores for parallel line recognition in the audio-visual condition as a function of consonance and dissonance. And disregarding congruency. Error bars show ± 1 SEM.

Figure 4.7. and Table 4.1 shows the results for congruent and incongruent AV presentation on parallel line discrimination. As can be seen a similar pattern is found as in the AO condition shown in Table 4.1., in that dissonant stimuli are feature segregated more successfully than consonant. Unremarkably when grouped there were significant differences in contrasts comparing dissonant and consonant frequency groups, illustrating the magnitude of the harmonicity effect, however the focus of the experiment was to assess whether AV stimulation would negate the harmonicity effect found in AO and so this is the focus of the analysis.

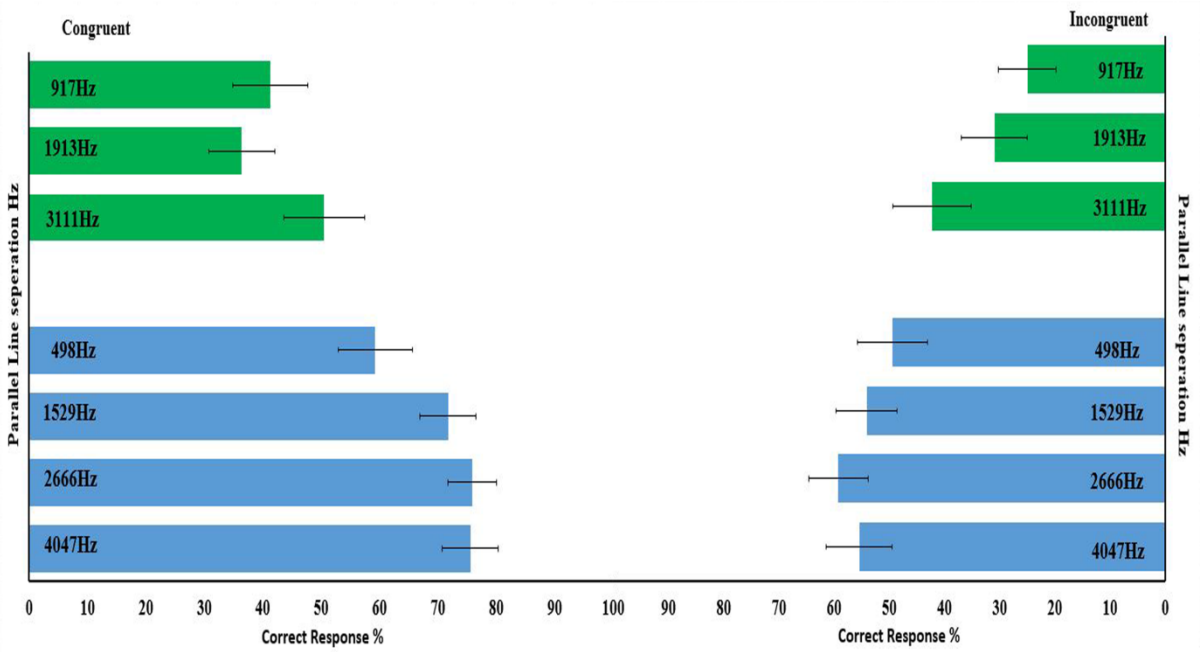

Figure 4.7.: Correct performance for congruent (left) and incongruent (right) in the audio-visual task. Consonant stimuli are shown in green and dissonant in blue. Dashed line represents chance level and error bars represent ± 1 SEM

Firstly I assessed the effect of AV presentation, contrasted to AO, for parallel line recognition, disregarding consonance and dissonance effects. An ANOVA with Bonferroni corrected contrasts showed a main omnibus effect for presentation type ($F(1.56, 34.42)=5.522$, $p=0.13$, $\eta p^2=0.201$), with contrasts illustrating the AV congruent ($M=58.66\%$, $SD=20.07$) was recognised significantly better than AV incongruent ($M=45.23\%$, $SD=22.57$) with a mean difference of ($M=13.432$, 95%CI [5.60, 21.26], $p=0.001$). Performance on the AV congruent was also better than the AO condition ($M=48.16\%$, $SD=18.47$) although this didn't reach significance ($M=10.501$, 95%CI [-.99, 21.99], $p=0.081$). Overall, provision of congruent AV information elicited superior performance compared to the AV incongruent, and almost AO, but is the magnitude of this effect different for the consonant and dissonant stimuli?

For the consonant stimuli, where harmonicity effects negatively impact on discrimination,

there was a no main omnibus effect for type of presentation ($F(2,44)=3.010$, $p=0.06$, $\eta p^2=0.120$), although there was a significant contrast in that AV congruent (M=42.75%, SD=26.33) presentation was better than AV incongruent (M=32.79%, SD=25.85) , (M=9.964, 95%CI [0.59, 19.34], p=0.035). AV incongruent was slightly worse than AO but nowhere near significance. Therefore, as with the total, AV congruent presentation improves recognition for consonant stimuli but this is still insufficient to counteract the effect of harmonicity as scores are still below what would be expected by chance. Incongruent AV information has no negative impact compared to AO presentation.

For differences in presentation type for the dissonant stimuli, where performance was above chance in the AO condition, an ANOVA with Bonferroni corrected contrasts showed a main omnibus effect for presentation type ($F(1.49, 32.73)=7.006$, $p=0.006$, $\eta p^2=2.42$). Contrasts showed a significant difference between AV congruent (M=70.58%, SD=19.33) and AV incongruent (M=54.55%, SD=24.15) with performance better for the former (M=16.03, 95%CI [8.46, 23.60], $p<0.0005$). Compared to the AO (M=59.95%, SD=19.65) the AV elicited a higher mean but the difference didn't reach significance (M=10.63, 95%CI [-1.21, 22.48], p=0.089). AO was better than AV incongruent but with a p=0.944 this was inconsequential.

A final analysis looked at the effect of congruent and incongruent audio-visual presentation on filled lines and showed higher mean scores for the congruent audio-visual (M=61.79, SD=17.38) compared to auditory-only (M=50.22, SD=28.21) and lower mean scores for incongruent audio-visual (M=46.01,SD=17.38) compared to auditory-only, but neither significantly different (p=0.094 and p=0.272 respectively).

In summary, the use of audio-visual stimuli improves feature segregation resulting in better recognition of auditory parallel lines compared to the audio-only condition. However, this is

only when the audio-visual presentation is categorically congruent. The effect is strongest for dissonant stimuli implying that the congruent visual lines, while aiding discrimination, do not override the effect of harmonicity. Incongruent visual stimuli had little effect on feature segregation compared to the auditory-only condition.

4.6.0. General Discussion.

In this study I used a sonified line discrimination task to evaluate the influence of auditory harmonicity and stimulus proximity in object discrimination using visual-to-auditory sensory substitution. Results of Experiment 1 demonstrated a general but weak influence of proximity, in that more distal sonified lines were more likely to be segregated into two auditory objects. However, more influential in segregation was the harmonic relations between the two sonified lines. If this was consonant, eliciting tonal fusion, it was more likely that the parallel lines would be integrated into one object. For dissonant frequencies with no harmonic interference, superior signal segregation was generally a factor of proximity. The influence of spectral principles of auditory stream analysis could therefore be seen as a limitation of the device conversion algorithm.

In Experiment 2 I evaluated whether the provision of categorically congruent visual information, presented synchronously with the soundscapes, would negate the effect of harmonicity at the consonant frequencies and increase the likelihood of false positive responses at dissonant intervals. In the opposite direction I tested whether incongruent audio-visual presentation would increase Type II errors compared to the audio-only condition. Results demonstrated that whilst congruent audio-visual information elicited superior performance relative to audio-only this was insufficient to negate the effects of auditory harmonicity

ASA regards auditory objects as the main unit of attention (Shinn-Cunningham, 2008) with these objects defined as the representation of a group of coherent sounds perceived to come from the same physical sound source (Alain, 2007). The rules that dictate how feature-based sensory input are grouped into object-based representations are spectro-temporal with continuity (pitch, loudness) and similarity (timbre) grouping features across time, and coherent changes in the spectral envelope and harmonicity for grouping concurrent stimuli. One way to ascertain whether a set of frequency components were emitted from a single source is evaluation of harmonic relations between components (A.S. Bregman, 1994; A.S. Bregman, Levitan, & Liao, 1990; A. S. Bregman, Liao, & Levitan, 1990).

As different physical objects in the environment vibrate they generate a harmonic spectrum that consists of partials (sine waves) that are all approximate multiples of the fundamental frequency (f_0) (A.S. Bregman, 1994) . The resolution of these partials by the auditory system elicits the perception of tonal-pitch. The harmonicity effect is driven by the coincidence of these partials for different tones, with consonance signified by a greater number of coincidences relative to dissonant sounds. If an auditory scenes' spectra contains partials that are not multiples of the f_0 then they are inharmonic or dissonant and unlikely to be grouped as coming from the same object, that is, they are segregated as separate auditory objects (A.S. Bregman, 1994).

The level of tonal fusion dependent on harmonics is exemplified and used in musical theory (DeWitt & Crowder, 1987). Musical notes, ordered by fundamental frequency, are arranged in a notational scale of 8 notes (A-G) with the frequency differential between notes termed the interval. For example, an octave is an interval between one pitch (e.g. 440hz) and another with half (220Hz) or double (880Hz) the f_0. This is perceived as highly consonant, as both notes share partials, with the likelihood that two 'visual' lines sonified at such intervals

would unlikely be segregated. Other highly consonant intervals are perfect 4^{ths} and 5^{ths} with imperfect, major and minor 3^{rds} and 6ths consonant to a lesser degree (Davies & Davies, 1978; Terhardt, 1974). The frequency analysis grouping for Experiments 1 and 2 by consonant and dissonant trials used musical notation theory for categorization.

Considering these rules that govern spectral grouping in ASA it is of no surprise that pairs of parallel lines at consonant frequencies were problematic in the segregation of the two tones, as hypothesised. However, Gestalt grouping and ASA also predict that more proximal objects will be less likely to be segregated. This was our secondary consideration. That there was a limited effect of proximity is unsurprising considering the literature on both auditory and visual grouping. While it is true that both Gestalt grouping and ASA posit the influence of proximity in successful segregation; that is, more spatially proximal objects are more likely to be grouped, the spatial configuration (and sonifications) of the stimuli used were beyond discrimination thresholds found in psychophysical evaluations in both the substituted modality vision, and the substituting modality audition. For the former vernier acuity paradigms have demonstrated thresholds, under optimal conditions, of about 2 seconds of visual arc (Berry, 1948), typically 5 to 10 times smaller than the closest spacing foveal cones (Westheimer, 1978). These threshold levels are dependent on high target visibility, high contrast and synchronous presentation (Berry, 1948; Klein, Casson, & Carney, 1990; Watt, 1984; Waugh & Levi, 1993a, 1993b; Westheimer & Hauske, 1975; Westheimer & McKee, 1977), all features of the stimuli used in the task. This fine grain discrimination of the visual system illustrates that from a psychophysical level segregation of the two lines should be a simple task.

From an auditory or substituted perspective the discrimination of the two parallel lines is a frequency discrimination task measured in auditory psychophysics using just noticeable

difference (JND) paradigms. In a typical task the listener is played two tones successively and required to indicate whether there is a difference in pitch, with the JND being the threshold at which change is perceived. Again the literature illustrates a perceptual system capable of fine grain discrimination and modulated by the baseline frequency of the stimulus. For example, the JND for simple low frequency tones (125Hz-2000Hz) is constant at about 3Hz. Baseline frequencies above this elicit larger JND's: 12Hz at 5000Hz, 30Hz at 10000Hz, 187Hz at 15000hz (Shower & Biddulph, 1931; Wever & Wedell, 1941). Kollmeier demonstrated JND's for sine waves and complex tones below 500Hz at about 3Hz and 1Hz respectively, implying an advantage to signal complexity (Kollmeier, Brand, & Meyer, 2008) relevant to the study as each sonified line was a complex of sine waves. Paradigms where the tones are presented concurrently elicit even smaller JND's compared to sequential presentations as the listen is then able to use beat frequencies for discrimination.

The JND's illustrated are far below the minimum frequency gap in the parallel line condition (2 x line width = 194Hz) implying that auditory segregation based on proximity should be a simple task. There is certainly an effect of proximity in that when consonance is accounted for superior segregation is found for larger gaps, but as this is also found in the filled line condition, with no harmonic confusion, we can safely posit that it is the harmonicity effect that facilitates object feature segregation, or lack of, and subsequent poor discrimination performance.

While the harmonic relations of the signal may be a potential limitation of the algorithm would this be negated by the provision of concurrent visual information? It is well established that we integrate incoming sensory information from multiple modalities to provide a coherent view of the environment and that bidirectional crossmodal influences have been shown to facilitate increased performance, compared to a unimodal counterpart, in tasks such

as speech perception (Frassinetti et al., 2002; Kollmeier et al., 2008; Seitz, Kim, & Shams, 2006). Crucial to crossmodal integration is stimulus congruence at either high or low levels or processing (Soto-Faraco et al., 2004b; Vatakis & Spence, 2006), and temporal synchrony. Secondly, the nature of the task influences the site of cortical processing.

Important in the formation of object-based representations from integration of crossmodal sensory features is synchronicity. If features are not temporally synchronous then they are likely to be segregated (Spence, 2011). In the present study onset, offset, and duration of all tones were equal, illustrating temporal consistency, and therefore y-axis discrimination is a spatial task. The weighting of auditory and visual stimuli in crossmodal integration is task-dependent supported by a meta-modal theory of brain organization (Pascual-Leone & Hamilton, 2001). This theory views the brain as a task based machine with computations based on function rather than being modality specific. Central to the theory is that brain areas for specific modalities are functionally optimal for particular computations; auditory areas for temporal features or tasks and visual for spatial (Proulx et al., 2014). This has been demonstrated behaviourally in that visual information has been shown to dominate over concurrent audio information in bimodal spatial perception (Alais & Burr, 2004b; Bertelson & Aschersleben, 1998; Driver & Spence, 1998) and motion (Kitagawa & Ichihara, 2002; Lewis et al., 2000; Soto-Faraco, Spence, & Kingstone, 2004a), while in temporal tasks the opposite is found with auditory dominance for interval duration (Burr et al., 2009; Grondin, 1993; Ortega et al., 2014; Romei et al., 2011), synchronization of auditory and visual flicker (Shipley, 1964) and rate perception (Recanzone, 2003).

Applying this to the audio-visual stimuli in Experiment 2, the spatial nature of the task adds weight to the visual features in the process of audiovisual integration. Overall, providing the

AV presentation is categorically congruent, this should increase false positives (i.e. a visual parallel line elicits an incorrect response to a single auditory line) to dissonant stimuli where performance is already above chance level (50%) and reduce errors in consonant conditions. As the results show, I found both of these effects. Conversely if the AV is categorically incongruent there could still be degradation in discrimination performance as the weighting of visual stimuli may elicit a type II error, although this was not shown in the present study.

The overall results clearly illustrate object-based representations in the SSD algorithm may be limited by the principles of auditory stream analysis and how this may manifest in confusion in audio-visual object recognition. While it posits an explanation for the misidentification of simple alphanumeric letters such as 'E' in training, the nature of the training task emphasises the issue. The sonifications in training were presented virtually, that is, the object is sonified using The vOICe but relayed to the listener as a static object soundscape. The soundscape is consistent as it is not modulated by sensor-object distance and angle. Therefore if the object feature lines are at consonant frequencies when recorded they will be for each presentation. This is negated if the device is used in real time as even slight movements of the sensor will realign the objects in the visual field, changing the soundscapes and negating, or moving the points of confusion. Sensor movement may also allow for higher-than-expected visual acuity due to the use of dynamic information to give a higher fidelity picture of the environment (Proulx et al., 2014).The results do however emphasise that if using static virtual objects in training to provide them at multiple object sizes to reduce interference from consonant frequencies.

This 'on screen' consonance issue become more salient in situations where sonifications are required outside the scope of using an SSD. For example, while the visually impaired have access to written text via screen readers the representation of graphical objects, such as

required in flow diagrams and bar graphs, is not facilitated by these devices. Attempts to represent these static images as sonifications on a screen should consider the frequency components of the object features to ensure that consonant lines are avoided. For example Figures 4.8a. and 4.8b. show hypothetical column graphs and 4.9a. and 4.9b. a flow chart with an additional frequency range on the y-axis. While the two bars in 4.8a. are more distal than their counterparts in 4.8b., consonance between the frequencies of the bar sonifications may result in a misrepresentation of the data in 4.8a, that is this may be perceived as one bar (top or bottom, dependent on high or low frequency preference). Similarly in the flow chart in 4.9a, the horizontal parallel lines of rectangles in 4.9a may be perceived as one due to the effect of harmonicity. As with compensation in SSD use (moving the sensor) negation of these sonification issues is a simple endeavour. Overlaying a grid on the workspace, (4.9b.) highlighting the frequency components of each grid line, would allow size-standardised objects that avoid consonance. Furthermore elements such as error bars, providing they were of different magnitudes, provides further information so that segregation could be achieved from temporal factors.

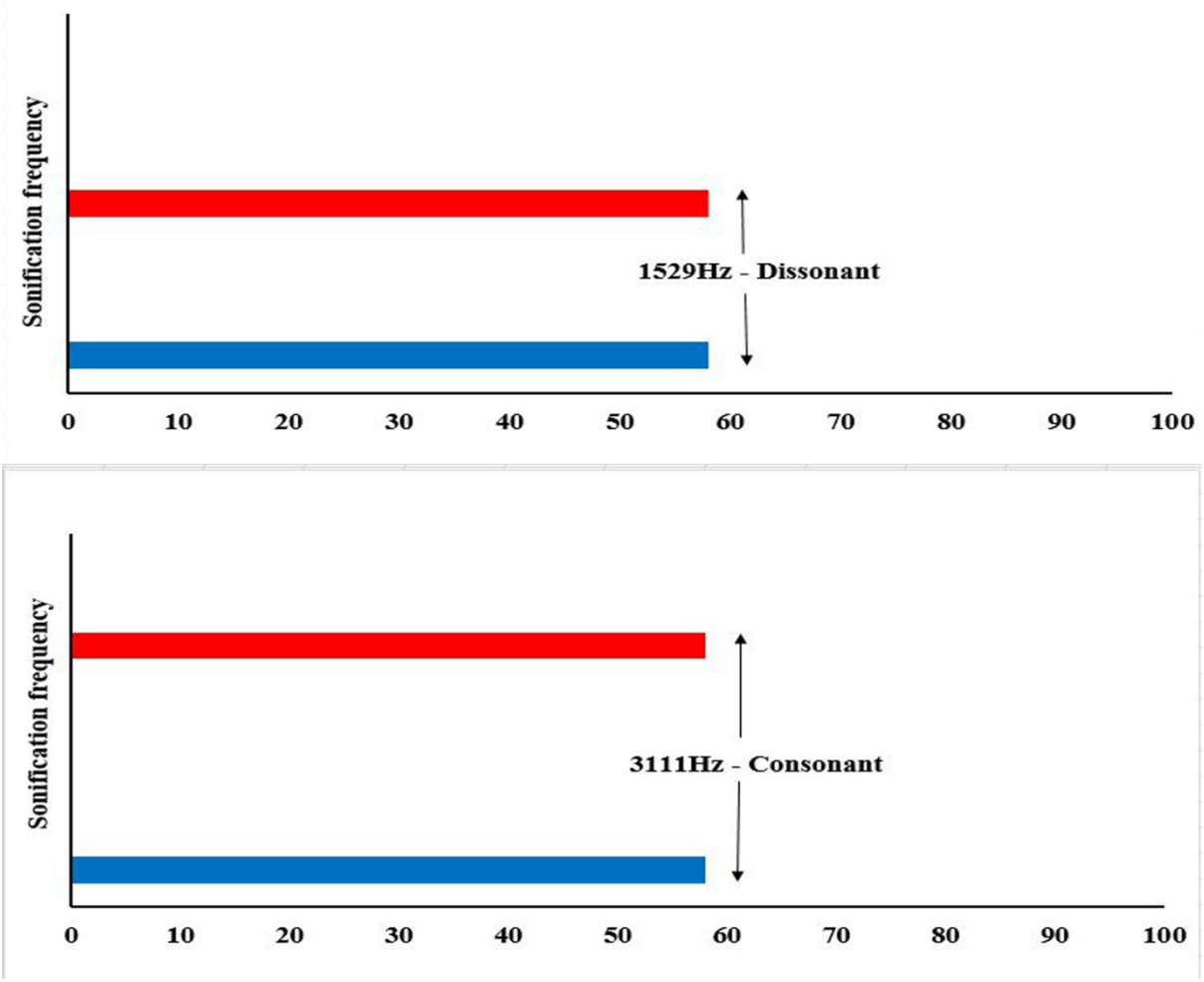

Figure 4.8a. and 4.8b.: Two hypothetical bar graphs to illustrate potential misidentification of bar graph elements due to harmonicity. While the two bars in 4.8b. (bottom) are more spatially distal than 4.8a. (top) the consonant frequency of the interval may result in a segregation failure leading to the perception of 1 bar.

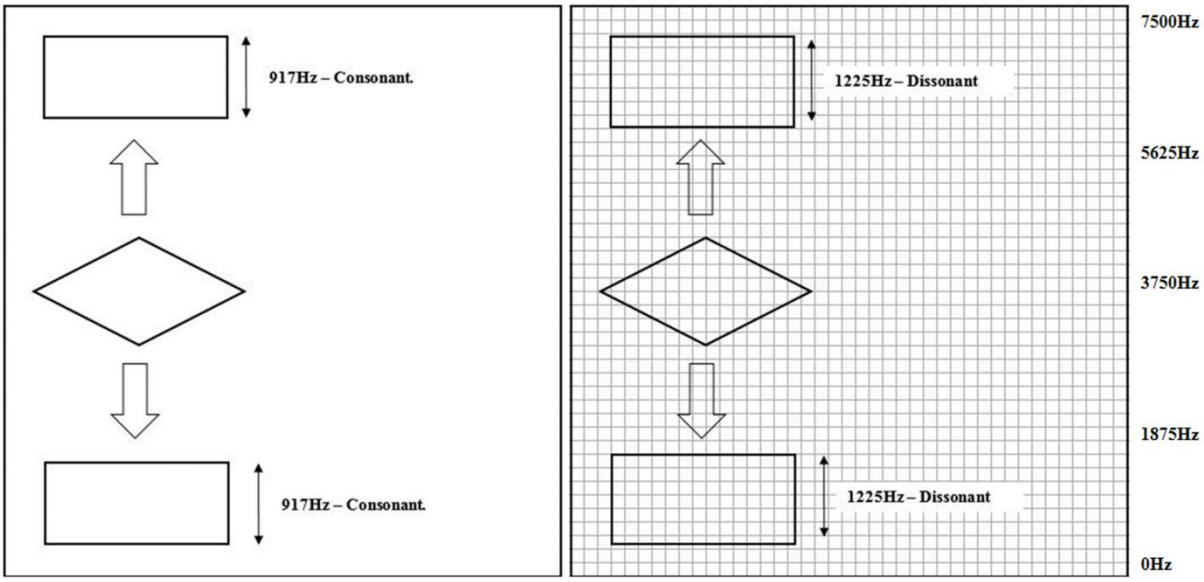

Figure 4.9a. and 4.9b.: Diagram to illustrate how harmonicity may influence the sonifications of flow chart elements. In 4.9a. the two horizontal lines in the rectangles have a consonant interval and thus likely to be misidentified. In 4.9b. a grid is overlaid to represent sonified frequencies allowing for a resizing of the rectangles to ensure dissonant frequencies for the rectangular elements.

While this experiment demonstrated potential problems in naïve listeners in future research it would be interesting to see if this applies in different populations. For example, in the blind, especially congenital, the audio-visual presentation would have to be replaced by a audio-tactile representation reducing the spatial dominance associated with visual processing. Would performance in the crossmodal condition be therefore degraded? Similarly in trained users of SSD acuity is high suggesting that the problem of proximity is accounted for in training and subsequent crossmodal associations. It would also be interesting to test whether training in audio-visual or audio-tactile would generalize to improved discrimination in the audio-only conditions.

In summary, I demonstrated that the formation of auditory based objects from cross modal sensory features using a visual-to-auditory SSD may be limited by the principles of auditory stream analysis, partially negated by the provision of synchronous crossmodal stimulation and how this can be applied using simple, active, strategies.

Chapter 5

In Chapter 4 the results demonstrated that the principles of auditory scene analysis, particularly spectral harmonic relations, may be theoretically limiting in auditory object formation and that this confound is strong enough to overpower additional visual input. In Chapter 5 I return to the idea of complexity discussed in Chapters 2 & 3. In Chapter 2 I demonstrated that increased complexity facilitated superior performance in a low-level task, while in Chapter 3 the results showed that a high-level task like object recognition can be achieved with a degraded level of information. In the final experimental chapter, again looking at complexity, I evaluate whether processing is limited by the amount of information provided. In this experiment a SIM/SUCC paradigm is employed to manipulate the density of information in a 2D object recognition task with the hypothesis that information capacity is limited and therefore reducing this lead to an increase in behavioural performance.

Splitting the load in visual-to-auditory sensory substitution. Capacity limits may hinder the recognition of complex objects.

David J. Brown [1]

[1] Biological and Experimental Psychology Group, School of Biological and Chemical Sciences, Queen Mary University of London.

[2] Department of Psychology, University of Bath

Abstract.

The formation of auditory objects in visual-to-auditory sensory substitution requires the matching of visual features corresponding auditory features. In naïve users of sensory substitution devices complex images are difficult to recognise compared to images with a low density of features. Cognitive load theories imply that perception has processing capacity limits that restrict the number of sensory features that can be attended to simultaneously. Using an object recognition task, sighted listeners (n=18) matched soundscapes from a visual-to-auditory sensory substitution device with visual and tactile objects in a 4AFC. Stimulus load was manipulated by simultaneous or successive presentation. Results show that recognition was significantly better for successive presentation, with low load, implying a capacity limit in object formation in sensory substitution. The design of the study and behavioural results show potential for direct translation sensory substitution training.

5.1.0 Introduction.

Numerous techniques have been developed to make the visual world accessible to those with blindness (legally defined as an acuity of 20/200 in the better eye) which affects almost 40 million people worldwide (Pascolini & Mariotti, 2012). Invasive techniques such as implants provide low resolution imagery by stimulating surviving retinal cells (Eickenscheidt et al., 2012; Keseru et al., 2012; Noorsal et al., 2012; Zrenner et al., 2011) or cortex (Brindley & Lewin, 1968b; Dobelle et al., 1974; Normann et al., 1999; Schmidt et al., 1996)or optic nerve (Veraart et al., 2003). Aside from the risks associated with surgical procedures these methods are expensive, provide a low functional acuity, and require extensive training to re-establish existing, or stimulate new, neural connections.

Non-invasive methods rely on the plasticity of the brain to transmit information usually attributed to the impaired visual system via an unimpaired modality. This method is termed sensory substitution with the prosthesis the sensory substitution device (SSD). Generally the substituted modality is vision with the substituting modalities touch (Arnoldussen & Fletcher, 2012; Bach-y-Rita, 2004; Bach-y-Rita, Collins, White, et al., 1969; Danilov et al., 2007)or audition (Abboud et al., 2014; Capelle et al., 1998; P. Meijer, 1992). For the former, acuity is low, dictated by density of touch receptors (representative of up to 400 functional pixels) but adequate for tasks such as object recognition, localisation and navigation. VA devices exploit the wide frequency resolution of the cochlea and large dynamic range of the auditory nerve to provide a higher theoretical and functional acuity (Haigh et al., 2013; Striem-Amit, Guendelman, et al., 2012).The effectiveness of VA devices has been demonstrated for both of the primary facets of visual perception -object recognition and localisation - in sighted (blindfolded), congenital and late blind users (D. J. Brown et al., 2011; Kim & Zatorre, 2008, 2010; Poirier et al., 2007; Proulx et al., 2008)

The conversion principles of one VA SSD, The vOICe (P. Meijer, 1992) utilise natural crossmodal correspondences (Spence, 2011) to inform the algorithm and therefore it is unsurprising that simple 2D and 3D objects can be recognised by naïve users with minimal training, or even when just the algorithm is explained to the listener (D. J. Brown et al., 2011; Kim & Zatorre, 2008; Proulx et al., 2008) However, as with all perceptual learning, increased device use should facilitate an increase in levels of performance emphasising the importance of developing effective training protocols.

With imagery being converted to sound, recognition of objects using one VA SSD, The vOICe requires the matching of visual features with corresponding auditory features in the output signal. The computational algorithm is informed by natural crossmodal correspondences, such as the associations between visual elevation and auditory pitch, or visual brightness and auditory loudness (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Marks, 1987; Stevens & Marks, 1965) and therefore it is unsurprising that naïve users perform above chance in object recognition and localisation tasks even with no training or knowledge of the algorithm (D. J. Brown et al., 2011; Kim & Zatorre, 2008).It would be interesting however to test whether these innate crossmodal understandings would also be found in the CB, where a lack of visual experience should hinder the formation of these correspondences.. In recognition of an objects' basic visual shape, if devoid of confounds such as texture, the outline boundaries of the shape are salient. This is also applicable to discrimination of object features in the soundscape of The vOICe. However, unless using the edge enhancement toggle on the device which uses a Sobel operator to enhance the outer edges of objects, the default setting is for the outline shape to be filled with pixels/auditory noise dependent on the brightness and density of pixels. While this additional information may be advantageous if conveying feature information such as shading or texture, if the

object is consistent in such features, then this additional information may be regarded as noise, impacting on the signal-to-noise ratio and hindering performance.

In SSD training naïve users are often presented with basic shapes at maximum contrast (white on black) to further the understanding of the algorithm. These are advantageous as they provide clear object information without ecological considerations such as changing light, movement and shading. However, it is interesting to note how learned objects can be misidentified when a novel but similar object is introduced into training. For example, a novel filled circle is misidentified as a learned semi-circle. .

This is understandable if we consider how the image is sonified. If the circle is divided in two at the midpoint of the y-axis to create two semicircles, the frequency pattern of the top semicircle rises from the y-axis midpoint left, peaks at the x-axis midpoint and falls to y-axis midpoint right. The opposite is found for the bottom semicircle. Frequency drops from y-axis midpoint left, troughs at the x-axis midpoint and then rises again to y-axis midpoint right. Both of these signals play overlaid simultaneously for the full circle. Consistent misrepresentation of a sonified semi-circle as a circle can be viewed from two perspectives: first there is a capacity limit on the amount of perceptual processing that can be carried out in a given time, and second there is an attentional preference for specific auditory frequencies.

Perception involves the extraction of information from the environment which is input into sensory memory, filtered for relevance to the goal or task, with relevant information subjected to higher-order processes for goal directed action and task-irrelevant material discarded. The early stages of the process are posited to have capacity limitations in both duration and number of pieces of information. For example visual short term memory can retain around 3 or 4 pieces of information for around 10 seconds prior to them being subjected to decay and forgotten, while haptic memory shows a similar duration and set size

(Bliss, Crane, Mansfield, & Townsend, 1966; Jiang, Olson, & Chun, 2000; Luck & Vogel, 1997; Pashler, 1988) Capacity limitations in sensory perception have been demonstrated in tasks such as the attentional blink, in which a second target may not be fully processed if presented within a certain temporal interval (Shen & Mondor, 2006; Tremblay, Vachon, & Jones, 2005), illustrating a consolidation bottleneck in the flow from sensation to action. Other bottlenecks have been shown for set size in change direction paradigms and the psychological refractory period in which the time to make a correct response to one stimulus delays processing on a second. (Pashler, 1994).

The selection of task-relevant information is crucial in perception and modulated by the amount of attentional resources dedicated to processing the stimuli, relevant to goal directed behaviour. How these attentional mechanisms are prioritised has been an area of research for many decades with early attentional models of perception based on ‘early’ and ‘late’ selection. Broadbent (Broadbent, 1958) proposed a limited-capacity model, later advanced by Treisman (Treisman & Geffen, 1967; Treisman & Riley, 1969) in which stimuli are filtered early in the process based on low-level features such as shape, colour and pitch. This pre-attentive filtering allowed the passing of information with similar characteristics for to high-order processing, while discarding task-irrelevant information.. Late selection models (Deutsch, Deutsch, Lindsay, & Treisman, 1967; Norman, 1968) posited that capacity is unlimited and all sensory information is automatically attended to equally until higher-order semantic coding selects task- relevant information. As empirical support was found for both models (Miller, 1987; Snyder, 1972; Treisman & Riley, 1969) Lavie proposed that the contrasting results on the locus of selection could be explained by perceptual load (Lavie, 1995; Lavie & Tsal, 1994).

The perceptual load theory conceives of perception as a limited-capacity process, as in early selection models, but which proceeds automatically, as in late selection models, until resources are utilised. The locus of attention is therefore modulated by the perceptual load of the task. When the perceptual load of the task is high, processing is dedicated to task-relevant information and therefore task-irrelevant information is not perceived. Conversely, if perceptual load is low, processing is not exhausted by the task-relevant information permitting resources for processing task-irrelevant information. Thus an increase in perceptual load in task-relevant processing should reduce the extent of interference from irrelevant stimuli (Lavie, 2006; Macdonald & Lavie, 2008).

Typical paradigms present a target with one (low-load) or many (high-load) nearby distractors with the requirement to respond to the target and ignore the distractor. Increased reaction times to congruent distractors indicate these distractors have been processed implying low perceptual load. Evidence to support the model has been shown in visual perception demonstrating both inattentional blindness (Cartwright-Finch & Lavie, 2007) and inattentional deafness (Macdonald & Lavie, 2011; Raveh & Lavie, 2015). For a review of visual perceptual load research see (Lavie, 2011)

Research into whether the perceptual load theory applies to other modalities has been less successful. While Santangelo and colleagues (2007) found peripheral auditory cuing effects reduced when the listener was directed to a central auditory stream, (offering support for the perceptual load theory in audition)(Santangelo, Olivetti Belardinelli, & Spence, 2007), Murphy failed to find any support for this (Murphy, Fraenkel, & Dalton, 2013).

The idea of limited-capacity perception brings posits explanations for the misidentification of objects in sensory substitution, as described above. If attentional resources are driven to one aspect of the soundscape, for example high pitch, then resources may be insufficient to attend

to the rest of the object. In the circle example, attention to high pitch codes the top semi-circle but through depletion of resources neglects the bottom. The resulting perception is a misidentified top semicircle. While most perceptual load paradigms utilise the number of e distractors and reaction times as dependent measures, it is questionable whether in the recognition of auditory objects in sensory substitution there is there is redundant 'distractor' information, (although this is implied to a certain degree in Chapter 3). In initial training accuracy is the main measure, although response times are salient for more advanced users dues to ecological validity, and therefore to evaluate perceptual load in sensory substitution a method was required that assessed the impact of information density on recognition accuracy and, moreover one that could be applied in alleviating the problem in training.

A novel paradigm, developed to test processing capacity in visual search tasks, uses accuracy as the dependent measure offers a framework for assessing capacity limits in sonified object recognition (Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972) In the visual search method an example trial would include, for example, 16 visual objects presented on screen either all at once (SIMultaneously) or as two SUCCessive eight item displays (the SIM/SUCC paradigm). If there is a limit to processing capacity then performance on the SUCC condition, where attention is focused on half of the items at a time, should be superior to the SIM condition where attention has to be spread over the entire item set. Numerous unimodal studies have used this design to test for limits in capacity in various perceptual tasks such as , visual search, mirror symmetry, perceptual surface completion, attentional blink, and 2D and 3D object shape perception (Attarha, Moore, Scharff, & Palmer, 2014; Huang & Pashler, 2005; Huang, Pashler, & Junge, 2004; Scharff, Palmer, & Moore, 2013)) with capacity limits dependent on task. For example, for 3D shape recognition a fixed capacity limit was proposed, while for 2D shapes at consistent angles there was no limit on capacity (Scharff et

al., 2013). The latter is used in the present study, although discrimination is via the auditory signal.

The paradigm also allows us to direct attention to particular features of the object and soundscape. If the global object is constituted by a conjunction of local features then halving the object reduces the set size of the features, demonstrated to limit perceptual capacity in visual tasks (Huang & Pashler, 2005). Successive presentation of top and bottom with only the corresponding soundscape for that half should thus facilitate successful perception based on smaller set sizes.

A final consideration was the modality of the object. All participants in the present study were sighted giving us the option of a visual-to-auditory match. The assumption is that the familiarity of object recognition in the visual modality should facilitate superior performance compared to modality of touch – we evaluate object shape more frequently with our eyes than hands. However, I also repeated the experiment with the same participants using a similar design and a SIM/SUCC tactile matching task for extension of the results for the target user groups of SSDs.

Based on the literature I make two hypotheses Firstly, that the presentation of information in the SIM condition will elicit inferior results in the recognition of sonified objects compared to the SUCC condition, implying potential capacity limits. Secondly, due to this increased load there would be slower reaction times in the SIM condition for both modalities of input.

5.2.0. Method.

Listeners.

I recruited 18 undergraduate and postgraduate students from 19 to 31 years of age (M=23.06, SD=3.44) from Queen Mary University of London. All participants reported normal or corrected vision and normal hearing. 15 listeners self-reported as right handed. The study was approved by the Queen Mary University of London Ethics Committee (REC/2009) and the University of Bath Psychology Ethics Committee with all listeners giving written consent prior to commencement of the study. Remuneration was £12 for completion of all sessions.

Materials and stimulus design.

Visual images for sonification and tactile object creation were obtained from the EST 80 image set (Max Planck Institute, Germany) and Clipart. Stimulus sonification used The vOICe image sonification feature at default settings, and Adobe Audition 3.0. Stimulus presentation was via E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) on a Windows 7 desktop PC. Auditory signals were listened to on Sennheiser HD555 headphones. The blindfold was the Mindfold (Mindfold Inc. Tucsan, AZ).

Stimulus design.

White images on a black background were sonified using The vOICe's sonification feature at default settings (1 second scan, normal contrast, foveal view off). Each soundscape's total duration (x axis) was 1000ms with a total frequency range (y axis) of 500-5000Hz. Bitmap images from The vOICe sonification (keeping relative dimensions) were printed and used as templates for the 5mm foam board tactile shapes. The foam board cut outs were attached to a

background card. Therefore all images presented on screen, tactile objects and associated sonifications were dimensionally consistent.

Stimuli for the SUCC conditions were made by obscuring half the digital image with a black oblong, with the top or bottom edge on the y axis midpoint. Sonifications were made of these 'half' objects with the bottom half representing frequencies 500-2499Hz and the top half frequencies 2500-5000Hz. In the tactile matching task card 'masks' were used to obscure the top and bottom of the full tactile objects.

Procedure.

Session 1: Listeners watched a PowerPoint presentation describing how The vOICe algorithm converts images to sound including audio-visual explanation of the conversion and eight sample shapes – (none from the test set). The second section of the presentation explained the experimental procedure with four example trials.

Visual matching task (VMT).

Figure 5.1. shows an example trial from the VMT. For each trial the listener was presented with a four alternative forced choice procedure (4AFC) visual/soundscape association task. Listeners viewed four numbered images on the PC monitor while listening to 1000ms soundscapes, each repeated twice with a 500ms inter-stimulus gap. In the SIM condition each of the two soundscapes and four images were of the 'full' objects. For the two SUCC conditions the soundscape and images were presented one half at a time; In SUCC1 the top half of the object and soundscape was presented followed by the bottom half, with this reversed for SUCC2. The listener's task was to indicate which image the soundscape had been created from by responding 1-4 on the keyboard. Soundscapes and images were repeated twice by default. Each block consisted to 32 trials with 3 blocks per condition (SIM, SUCC1, SUCC2). While accuracy was stressed as the primary objective reaction times (RT)

were also measured from offset of final soundscape (not including self-initiated repeats) to keyboard response. Accuracy feedback was given via a post-trial auditory tone indicating a correct response.

Tactile matching task (TMT).

The basic procedure was similar to the VMT except four tactile, rather than visual, objects were presented to the blindfolded listener to explore haptically while listening to soundscapes. Verbal responses 1-4 were directly inputted by the experimenter who gave tactile accuracy feedback (a tap on the shoulder for correct). For the SUCC conditions a card mask was used to obscure the irrelevant half of the tactile object. Due to the much longer trial time (set up and response) the TMT consisted of 2 x 32 trial blocks per condition. RT response was recorded by the experimenter immediately on verbal response.

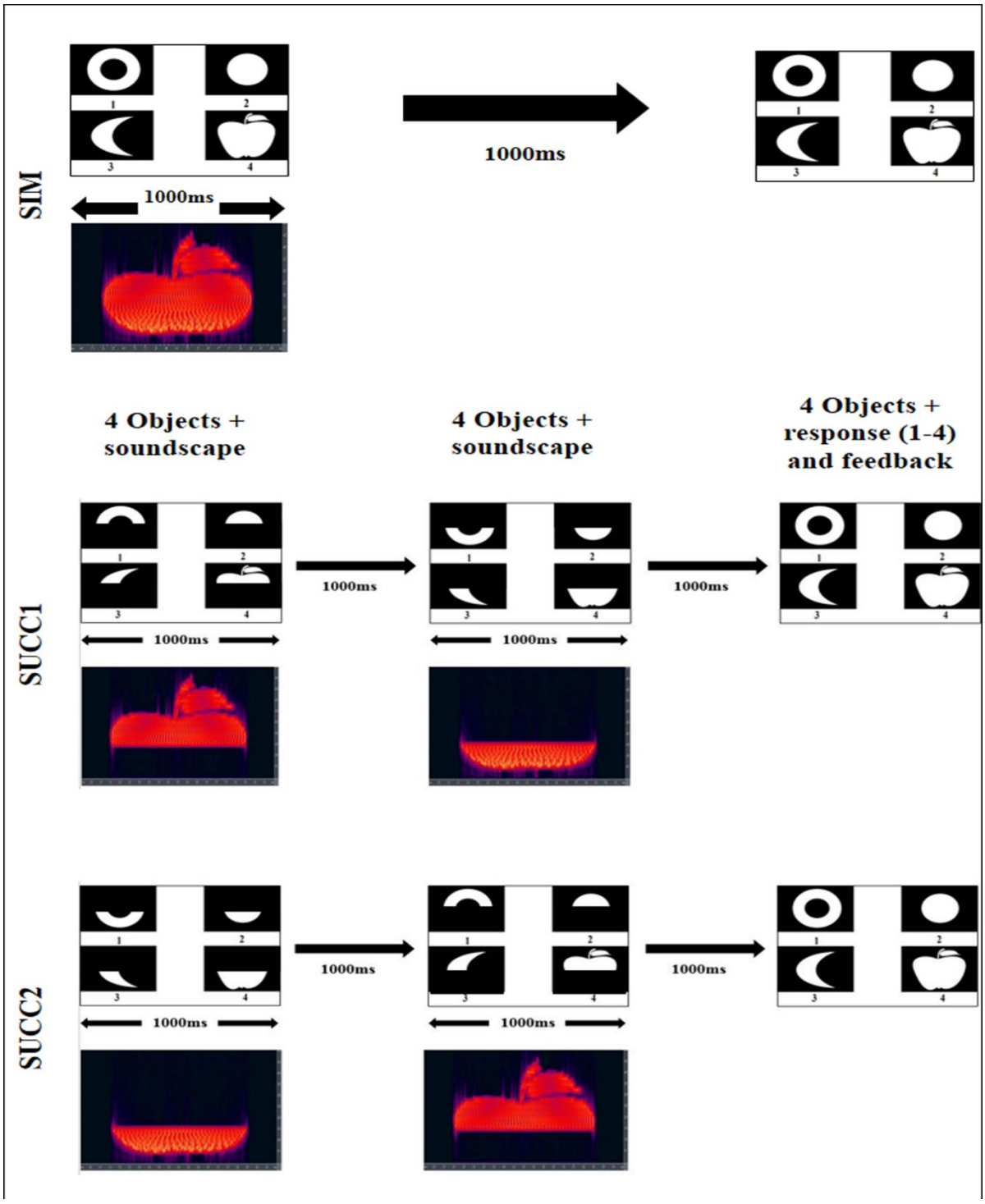

Figure 5.1.: Example trial showing visual/tactile presentation and spectrogram of the soundscape for each frame in the SIM and SUCC conditions.

5.3.0. Results.

5.3.1. Visual matching accuracy.

Figure 5.2. and Table 5.1. display the results for the first task in the experiment. Listeners were required to match visual shapes with the associated soundscape in a 4AFC. In the SIM condition full visual objects and soundscapes were presented simultaneously. In the SUCC1 condition the top half of the visual objects and soundscapes were presented first followed by the bottom half of the object and soundscape (vice versa for SUCC2) in a sequential format.

Table 5.1.: Mean accuracy and reaction times for SIM and SUCC conditions in both the visual and tactile matching tasks.

Condition	Accuracy %	S.D.	RT (ms)	S.D.
VISUAL				
SIM	38.67	10.30	1117	612.27
SUCC1	48.76	13.13	703	265.93
SUCC2	53.78	9.70	612	302.25
TACTILE				
SIM	45.41	8.14	7872	1545.06
SUCC1	58.98	7.52	7075	977.44
SUCC2	55.76	6.27	6955	1166.95
TOTAL				
SIM	42.04	8.66	4495	978.32
SUCC	54.32	7.63	3836	539.84

Analysis on the three conditions (SIM, SUCC1, SUCC2) showed a main effect of type of presentation with successful recognition in the SIM condition (M=38.67%, SD=10.30) being inferior to SUCC1 (M=48.76%, SD=13.13), and SUCC2 (M=53.78%, SD=9.70) conditions ($F(2,30)=19.432, p<0.001, \eta p^2=0.564$) although all were still above chance level of 25%. Planned contrasts with Bonferroni corrections showed significant differences in performance

between SUCC1 and SIM (M=10.09%, 95% CI[2.63,17.55],p=0.007) and SUCC2 and SIM (M=15.10%, 95% CI[9.33,20.87],p<0.001) demonstrating that the splitting of the signal and sequential presentation of the two 'halves' independently elicited better recognition than if the 'total' signal was presented. Within the SUCC condition, presentation of the bottom half (SUCC2) before the top (SUCC1) resulted in superior performance but not at a level that reached significance (M=5.01%, 95% CI[-1.60, 11.63],p=0.177). Collapsing the two SUCC conditions (M=51.27%, SD=10.45) into one and contrasting with the SIM condition (M=38.67%, SD=10.30) showed an overall improved level of performance for the former (t(15)=5.862,p<0.001, d=1.46).

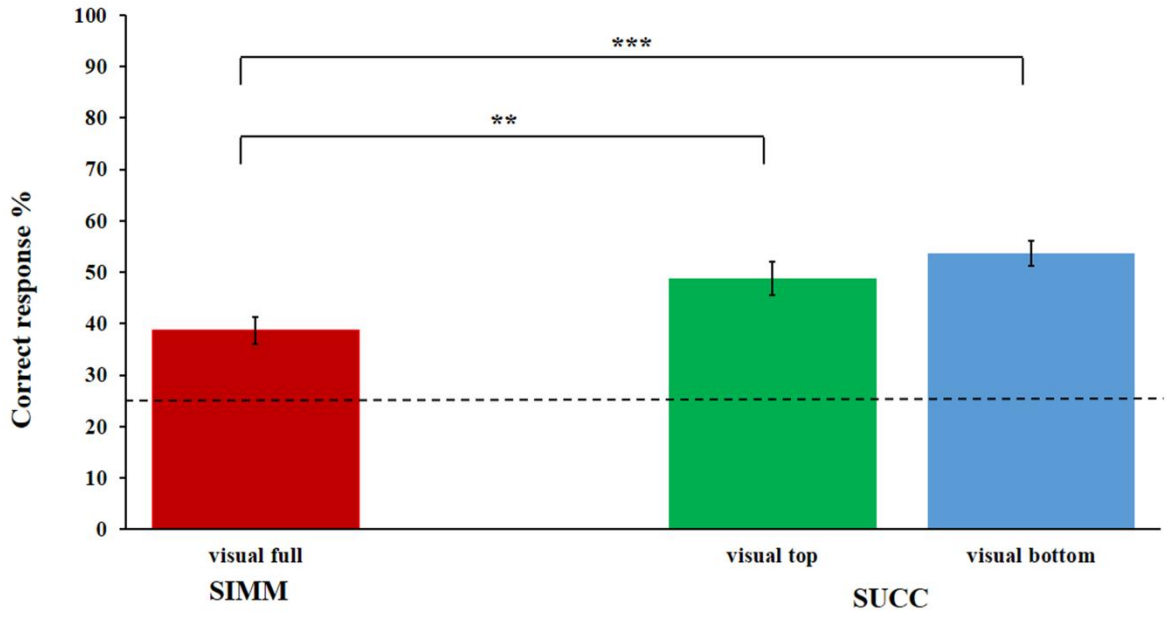

Figure 5.2.: Accuracy in the visual matching task for SIM and SUCC conditions. The dashed line represents chance and so all conditions were above this level. There was a significant difference between the SIM condition and both SUCC conditions but not between the two SUCC conditions. Error bars represent ±1 SEM . ** p<.01, *** p<.001

5.3.2. Tactile matching accuracy

Figure 5.3. and Table 5.1. show the results for the tactile/soundscape matching task. Presentation order of the two SUCC conditions was counterbalanced as in the visual matching task.

Analysis of variance demonstrated that, as in the VMT, there was a performance difference in the SIM condition (M=45.41%, SD=8.14) versus SUCC1 (M=58.98%, SD=7.52), and SUCC2 (M=55.76%, SD=6.27) conditions ($F(2,30)=50.274, p<0.001, \eta p^2=0.770$). Planned contrasts to assess where this difference lay showed significant differences between SUCC1 and SIM (M=13.57%, 95% CI[10.53,16.62],p<0.001), and between SUCC2 and SIM (M=10.35%, 95% CI[6.69,14.01],p<0.001) ,but not between SUCC1 and SUCC2 (M=3.22%, 95% CI[-1.35,7.79],p=0.231)

When the two SUCC conditions were collapsed (M=57.37%, SD=6.03) they demonstrated better matching than in the SIM condition (M=45.41%,SD=8.14), ($t(15)=14.052, p<0.001$, d=3.51) although both were still above chance, with poorer results in the SIM condition implying a limit in capacity for this presentation type.

Comparisons between the two modes of object display (tactile vs visual) demonstrated that overall matching the soundscapes with the tactile objects (M=53.39%, SD=6.58) was superior to the soundscape/visual object (M=47.07%, SD=9.58) matching ($t(15)=4.272, p=0.001$, d=1.07). This trend for superior performance in the tactile condition was significant for both SIM ($t(15)=6.738, p=0.001, d=1.00$) and SUCC conditions ($t(15)=6.104, p=0.001, d=0.80$).

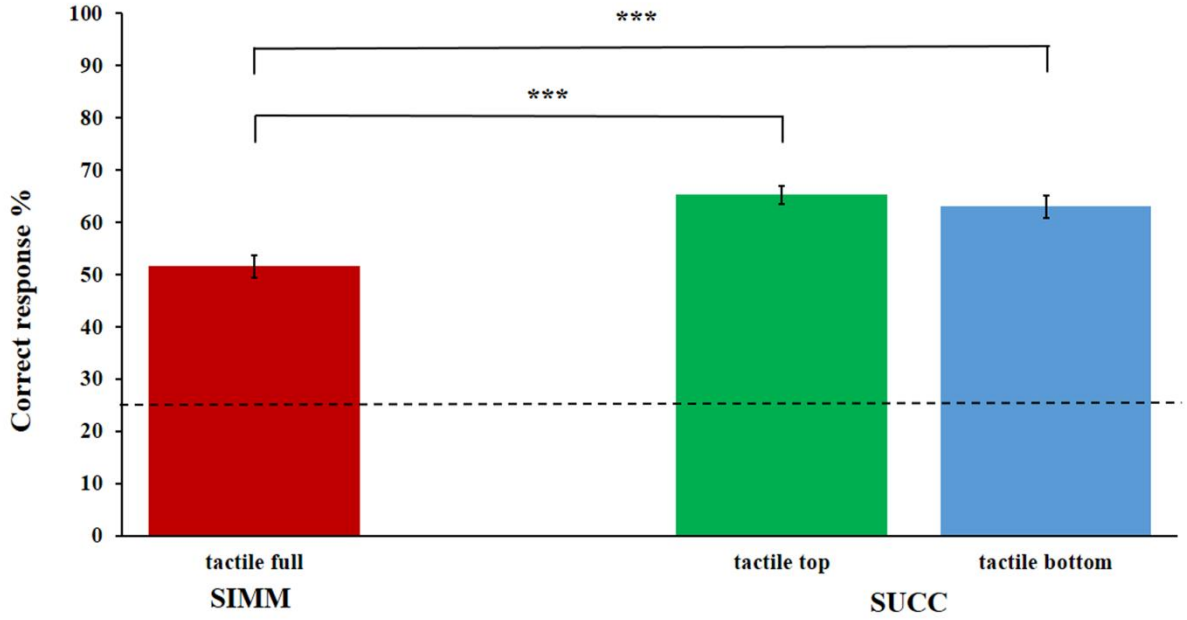

Figure 5.3.: Accuracy in the tactile matching task for SIM and SUCC conditions. The dashed line represents chance and so all conditions were above this level. There was a significant difference between the SIM condition and both SUCC conditions but not between the two SUCC conditions. Error bars represent ±1 SEM **p<.01, *** p<.001

5.3.3. Visual and tactile matching: reaction times.

Reaction times (RT) were taken from the offset of the auditory stimulus until the final response key press. In the visual object/soundscape matching tasks mean trial response time was 811 milliseconds (ms). For the tactile/soundscape matching task this was significantly longer at 7301ms. As object recognition times are much slower in the tactile modality compared to visual there was no real interest in comparisons between the two modalities. I was curious however about the reaction times for each type of presentation within modality, that is, SIM vs SUCC. Reaction times for both visual and tactile matching tasks are found in Figures 5.4. and 5.5. and Table 5.1.

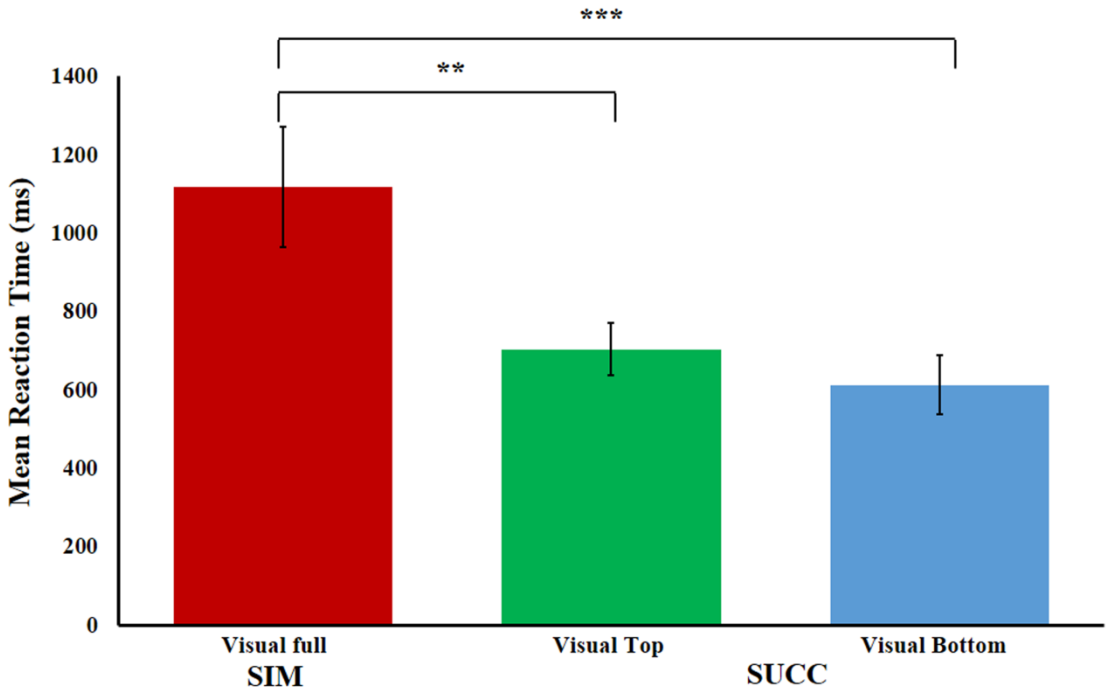

Figure 5.4.: Reaction times for responses in the visual matching task for the SIM and two SUCC conditions. Responses to the SIM condition were significantly longer than both SUCC conditions. Error bars represent±1 SEM . **p<.01 ***p<.001

For the VMT an ANOVA, with Bonferroni corrections, demonstrated that the type of presentation had a significant influence on trial speed ($F(1.21,18.16)=11.974, p=0.002, \eta p^2=0.444$). Objects in the SIM condition (M=1117ms, SD=612.27) were recognised much less rapidly than in the SUCC1 (M=703ms, SD=265.93), (M=414ms, 95% CI [133.12,695.79],p=0.004) and the SUCC2 (M=612ms ,SD=302.46) conditions (M=505ms, 95% CI [113.60,896.47],p=0.010). There was no difference in the speed of response between the two SUCC conditions (M=90.58ms, 95% CI[-86.25,267.41],p=.564).

Analysis of RT's in the TMT also demonstrated an influence of presentation type ($F(2,30)=3.994, p=0.029, \eta p^2=0.210$). As in the visual condition, contrasts between groups showed that object recognition in the SIM condition (M=7872.ms,SD=1545.06) was inferior compared to the SUCC1 (M=7075ms, SD=977.44) condition, although not quite reaching

significance (p=0.055), and the SUCC2 condition (M=6955ms, SD=6333.44), (M=917ms, 95% CI [67.04,1768.26],p=0.036). There was no significant difference between the SUCC1 and SUCC2 conditions (p=0.650).

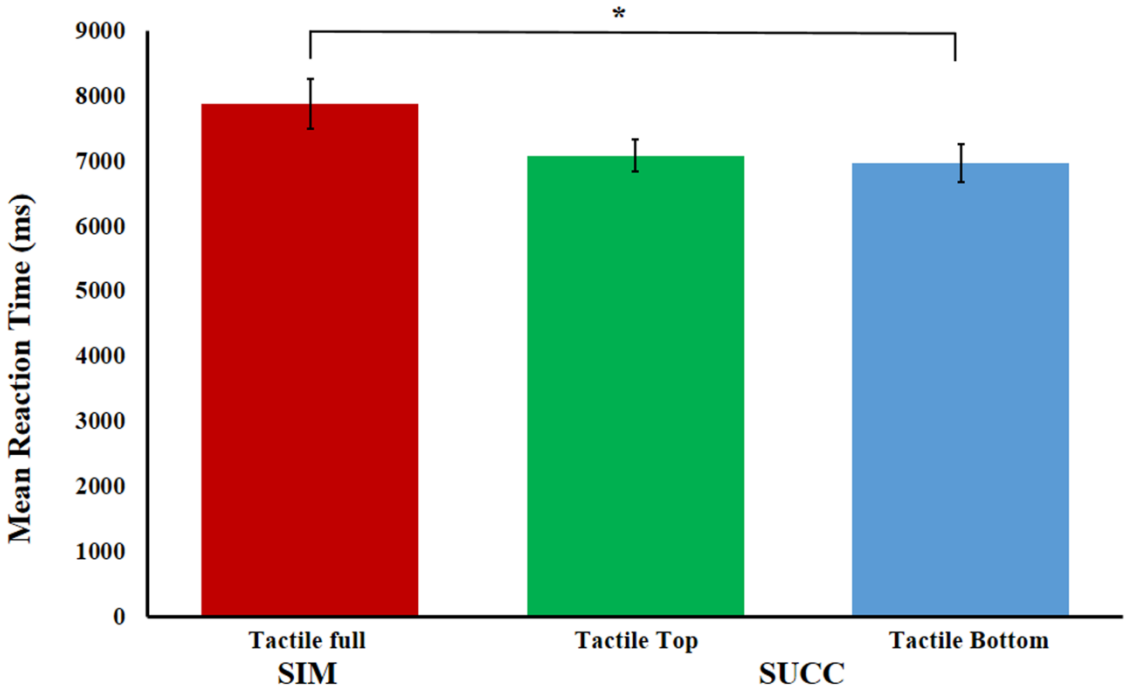

Figure 5.5.: Reaction times for responses in the tactile matching task for the SIM and two SUCC conditions. Responses to the SIM condition were longer than the SUCC2 condition but this only just reached significance. Error bars represent ±1 SEM . *p<.05

5.3.4. Correlation analysis of speed-accuracy trade-off.

To assess any speed/accuracy trade- off bivariate correlation analysis was performed on all group contrasts within modality. For the VMT there was no overall correlation but a significant correlation was found for the SUCC2 condition in which longer RT's equated to superior accuracy ($r_s(14)=0.557,p=0.013$). In the TMT there was an overall correlation in that an increase in accuracy again equated to slower RTs overall ($r_s(14)=0.438,p=0.045$) but there were no significant contrasts between SIM and SUCC conditions.

5.3.5. Results summary.

The results demonstrated that if the soundscape is split into two, (based on frequency components), and played to the listener successively correct matching is significantly better than if the whole soundscape is played simultaneously, irrespective of modality of the object to be matched (visual or tactile). As far as the speed of matching is concerned, the SIM condition was significantly slower than the SUCC conditions in both modalities although there was no significant speed-accuracy trade off. Overall these results suggest that matching in the SIM condition was more difficult, and less rapid, potentially due to limitations in processing capacity.

5.4.0. Discussion.

In this experiment I evaluated the effect of manipulating the level of information in an object recognition task using a visual-to-auditory sensory substitution device. The SIM presentation was designated as having high-perceptual load as it contained all stimulus information (auditory and visual) in a one-shot presentation while in the SUCC conditions the information was compartmentalised into two successive presentations each with a comparatively reduced perceptual load. As a second consideration I was interested in whether this performance would be affected by the modality of input on the 'non-auditory' (e.g. visual or tactile) stimulus feature.

In both auditory-visual and auditory-tactile matching tasks inferior performance in the SIM condition implies that either the processing of concurrent information is capacity-limited or the strategy induced by the paradigm is effective irrespective of perceptual load. While

caution must be taken in comparing the results of this experiment with the literature on perceptual load theory (due to major differences in the paradigm), it posits interesting questions. One of the primary differences in the experimental design was the use of crossmodal sensory information. Perceptual load theory has been applied to vision and audition as described, but recently has also looked at crossmodal interference. In an fMRI study using a one-back working memory design Klemen et al (2009) evaluated whether different levels of auditory perceptual load, manipulated by pitch discrimination, would differentially interfere with processing of task-irrelevant visual images shown at different visibility levels (Klemen, Buchel, & Rose, 2009). As visual object recognition is cortically represented by activation in the lateral occipital cortex (LOC) (Grill-Spector, Kourtzi, & Kanwisher, 2001), and incidentally also in object recognition using visual-to-auditory sensory substitution with The vOICe (Amedi et al., 2007), the authors hypothesised that activation in LOC would increase relative to object visibility and if attention was spread across modalities it would be reduced by high vs low auditory load. Conversely, modality-specific resources would result in LOC activity being unaffected by auditory load. Results demonstrated bidirectional interference with the processing of task-irrelevant visual stimuli in the LOC and a reduction of visual processing under high auditory load. In another fMRI study to evaluate whether load dependent effects cross modalities Weissman and colleagues found increased processing of auditory or visual targets in the high load condition compared to the low (Weissman, Warner, & Woldorff, 2004).

The demonstration of crossmodal inference in perceptual load theory allows us to loosely tie this to our paradigm and evaluate whether it is strategy or load that facilitates differential performance due to presentation type. The 4AFC paradigm in which a single soundscape has to be matched to its visual correspondent in essence has two targets (the soundscape + one visual object), and three visual ‘distractors’ i.e. the three incorrect objects Direct real-time

comparison of soundscape and visual object should elicit superior matching as object shapes can be viewed, haptically explored as the soundscape plays. This is not possible in the SIM condition as in the single presentation there are four visual objects but only two 1000ms soundscapes. Whilst it may not strictly be high perceptual load it is certainly a higher cognitive load involving memory. In the SUCC conditions there are sufficient presentations of the soundscape to make direct comparisons with the visual object. Frame 1 presents four visual objects and two soundscapes. Therefore two direct pairings can be made (e.g. the top two objects with the soundscape) and a 'no-no' or 'no-possibly' decision made. Frame 2 then provides two soundscapes and the visual images to corroborate the initial decision. Furthermore, in the SUCC condition there is a decrease in object 'set size', less information per frame, directed attention to the top and bottom edges of the shape, and a decreasing of noise from unattended 'filled' pixels. This gives the matching task a comparatively lower cognitive load.

The difference in direct matching mediated by presentation type may be further confounded by the use of audio-visual stimuli. While the literature suggests that three or four visual events can be processed at a time (R. D. Wright, 1994; Yantis & Johnson, 1990), this may be further limited in auditory-visual presentation. It is known that synchronous auditory stimulation can drive attention to a visual event making it salient (Ngo & Spence, 2010; Van der Burg, Cass, Olivers, Theeuwes, & Alais, 2010)even if irrelevant to the task (Matusz & Eimer, 2011) however the majority of multisensory integration paradigms uses a combination of sensory signals. Colonius et al (2004) postulated that multiple visual events may bind to one sound source if presented within a temporal window (Colonius & Diederich, 2004) with further evaluation by Van der burg et al (2013). In a pop-pin orienting paradigm participants could reliably detect a single synchronous audio-visual event, however performance declined significantly when more than one visual object was paired with the sound. The authors

posited that, based on ecological validity, there were different capacity limits in audio-visual processing, generally restricted to one event (Alais & Burr, 2004b; Van der Burg, Awh, & Olivers, 2013). If presented with multiple visual options which one does the sound bind to? Sound-vision matching in the SIM condition required a partial scan of the visual objects to assess all four, to contrast with two soundscapes. Binding may be due to the salience of the object features (e.g. high pitch auditory spike and sharp local visual feature) or simply by the object which is in view at the onset of the sound. The binding to a specific object may then restrict attention to the other objects during the second soundscape. While this would also occur in the SUCC conditions, the number of auditory presentations, and direct matching opportunities are still doubled.

Differentials in reaction times could also be a signifier of cognitive load as they were longer for the high-load SIM condition. However, it is more likely that this is due to the type of presentation. In the SUCC conditions there were four soundscape presentations in total for object/soundscape matching. If a definite match is made prior to the final soundscape then the reaction time is based on how quickly, post final soundscape offset, the response key can be hit, contrasted to the SIM condition with less presentations. Reaction times for the TMT were considerably slower than for the VMT and can potentially be attributed to the relative speeds of visual and tactile processing. Visual perception is rapid and often works in parallel (Cave & Wolfe, 1990) and thus in the VMT all four visual objects could be assessed quickly whilst on screen. Tactile exploration however is primarily serial and considerably slower (Craddock & Lawson, 2008; Overvliet, Smeets, & Brenner, 2007a; Overvliet et al., 2007b)and thus realistically only one object soundscape match could be made at a time.

It is difficult to distinguish whether the comparatively poor performance in the SIM condition is down to limitations from a high cognitive load or strategies allowed by the paradigm. If

performance is mediated by cognitive load then the density of pixels in an object should be influential, irrespective of condition, and analysis of individual objects within-condition should indicate this. For example, even in the lower-load SUCC conditions objects with a larger number of pixels (i.e. a high density image) would be recognised less easily than low density images. Unfortunately for this study individual object comparisons were not made. As for strategy, the requirements of the SUCC task drives attention to smaller subsets of object features and correspondent soundscape features, simplifying the stimuli and also offering double the amount of chances to make a crossmodal feature association. One way to differentiate between advantages conveyed by the paradigm and load factors would be to ensure number of soundscape/object presentations were consistent over SIM and SUCC conditions. If performance was still significantly better for the SUCC conditions then this points more towards the level of load being salient rather than the methodology. While not initially intended to be a part of this thesis supplementary Experiment 2, described below *(5.5.0.)* offers interesting data on this idea.

A final consideration is that attention could be driven to high or low frequency feature subsets based on attentional preferences. That is, why would people naturally attend to, for example, high frequencies and subsequently only recognise that set of shape features (such as the top semicircle in the circle soundscape)? There is little evidence in the literature to suggest frequency preferentialism although in adult speech perception males show a preference for high-pitched, rather than low-pitched, female voices with females generally showing the opposite (Re, O'Connor, Bennett, & Feinberg, 2012) and infants prefer infant-directed ‘motherese’ speech (IDS), characterized by higher pitched intonation and greater pitch variation, compared to adult-directed speech (ADS) (Cooper & Aslin, 1990; Fernald, 1985; Fernald & Kuhl, 1987). Considering the imbalance of participants as a function of

gender we would expect a preference for the low frequency presentations and with no significant differences between SUCC1 and SUCC2 this explanation seems unlikely.

5.5.0. Experiment 2 – Supplementary.

This experiment followed the basic procedure as of that in this chapter aside from three methodological differences: 1) there was no haptic condition, 2) there was only 1 SUCC condition (top half first), 3) 2 x 32 trial blocks instead of 3 x 32 trial blocks, 4) use of bone conduction headphones for half of the trials. The results below report firstly the SIM/SUCC minus the bone conduction results, as this is almost a direct replication of the VMT in this chapter and then inclusive of the bone conduction data.

5.5.1. Method.

I recruited 24 listeners (19 female) via an undergraduate Research Assistant module at the University of Bath. Age ranged between (18-23) with a mean age of (M=19.79, SD=1.06). All listeners reported normal or corrected eyesight and normal hearing. 4 self-reported as left handed. The study was approved by the University of Bath ethics panel (13-204#) and required informed consent prior to onset.

Materials.

The procedure used the same materials and design as in the experiment reported in this chapter aside from the additional use of Aftershokz Bone Conduction headphones for half the trials.

Procedure.

This follows the VMT aside from there was only one SUCC condition and two rather than three blocks for each condition. Therefore listeners performed the SIM condition (2 x 32 trials) followed by the SUCC condition (2 x 32 trials) using over-ear headphones followed by a repeat session using the bone conduction headphones. This was counterbalanced using a latin square to account for order effects.

5.5.2. Results.

Figure 5.6. Shows the results for the SIM vs SUCC comparison for Experiment 2 (left) where presentation times between conditions are equal. Considering the accuracy scores for the VMT (over ear headphone data only). A paired sample t-test contrasting types of presentation showed that accuracy for the SUCC condition (M=56.84%, SD=14.31) was significantly better than for the SIM condition (M=48.83%, SD=12.17) even when the presentation times were standardised across conditions (t(23)=3.670, p=0.001, d=0.75).

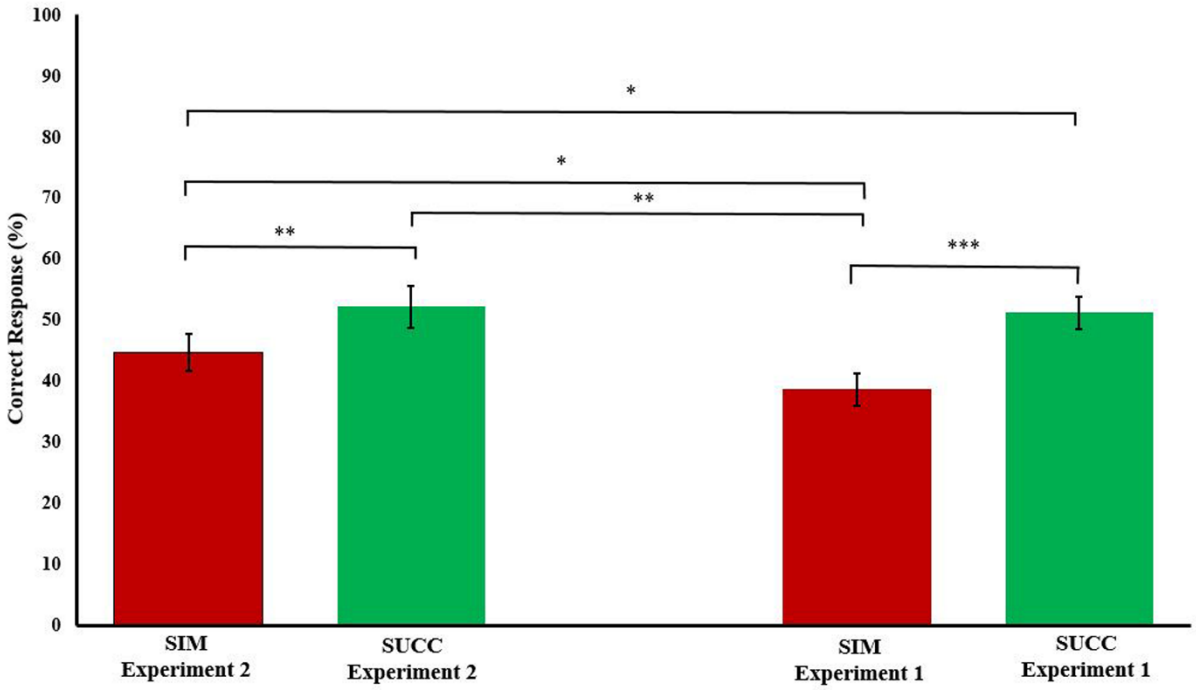

Figure 5.6.: Accuracy for SIM vs SUCC conditions for Experiment 2 (left) where the presentation time for each condition is equal, and Experiment 1 (right) where the SIM condition was presented for half the SUCC condition. . Error bars represent ±1 SEM *p<.05 **p<.01, *** p<.001

Even though the bone conduction headphones is a different type of input device (i.e. through the skull rather than the outer ears) and therefore we should be wary of direct comparisons, the advantage for the SUCC condition (M=56.35%, SD=10.53) over the SIM condition (M=50.39%, SD=9.60) was still apparent ($t(23)=3.570$, $p=0.002$, $d=0.73$) again with a relatively large effect size.

How does this data contrast with that reported in the main part of the chapter?

When contrasting the two experiments I would expect to find little difference between the two SUCC conditions, as this is basically a repeat procedure, and this was confirmed by the analysis with no significant difference between the two ($t(17)=0.291$, $p=0.775$, $d=0.07$). However for the SIM condition in the supplementary data, in which there were an equal number of presentations as the SUCC conditions, performance was better than for the SIM condition in the original experiment ($t(17)=2.118$, $p=0.051$, $d=0.53$), while still being significantly worse than either of the SUCC conditions.

Supplementary experiment results summary.

Overall the results imply a distinct advantage in object recognition using the successive presentation type even when the presentation of information across conditions is standardised i.e. the soundscapes/visual stimuli are presented for an equal amount of time. However, the standardisation of presentation times in the supplementary experiment SIM condition demonstrated superior performance compared to its counterpart in the original experiment implying that while the strategy the paradigm utilised explained some of the variance between the SIM and SUCC there is still solid evidence that the SUCC presentation is advantageous.

Application.

While the results overall have posited some ideas on load capacity in sensory substitution the behavioural data offers solid methods for how this can be applied in training for SSD use. For the primary target user group, the visually impaired, the visual aspect of the paradigm is redundant in training. The tactile task could be directly utilised however for early level training in basic shape recognition. From an applied sense it is logical to do this by manipulating the soundscape rather than the object. Thus the tactile object is always presented in its entirety but the soundscape is presented as two successive frequency based presentations – 2500-5000Hz first followed by 500-2499Hz – directing attention to the top and bottom edges of the shape successively. This could be mapped to the tactile shape by directing the user to explore the top and bottom edges of the object synchronously with the soundscape. Training in this using basic shapes may allow for attentional strategies for real world use outside the lab, negating the need to actually filter the output. For example, when scanning a scene to self-direct attention to the top frequency bands, and to 'picture' the top of the shape, the user could then repeat with low frequency for the bottom, in order to build up a representation of the object as a whole based on its component parts.

In conclusion, the experiment demonstrated that conditions with a high cognitive load are more problematic for recognition. This could be due to informational capacity limits, strategies allowed by the adapted SIM/SUCC paradigm, or a combination of both. The experimental paradigm employed here shows potential to be adapted to training due to the significant behavioural results.

Chapter 6.

6.1.0 General discussion.

This thesis, while concentrating on naïve users of The vOICe, had two interconnected routes of evaluation: firstly, in naïve users how does the initial auditory processing of the output of The vOICe impact its use as a substitution for vision? and secondly, how does the complexity of the signal influence processing and perceptual learning? Ideally the results of the experiments should feed forward into the methodology of developing training protocols for use of the device while also adding to established theory. In this final brief summary I will evaluate the results of each experiment with regard to the thesis questions before discussing limitations of the research and ideas on how to move the research forward.

Considering availability, devices such as The vOICe are underused in the visually impaired community. The vOICe is a free application and uses hardware that is relatively inexpensive and easily available (PC + webcam + headphones) and thus underuse is likely down to other reasons. Motivation could certainly be one of these. As shown in the literature, the crossmodal correspondences that inform the algorithm allow for impressive performance in simple tasks. However when moving away from the controlled conditions and limited-complexity tasks in the lab successful device use becomes more difficult. Hence, if the first use of the device is in a complex environment, without the solid grounding of the algorithm facilitated by training, then the difficulty of using the device may negatively impact on motivation. Make no mistake, seeing with sound to a level that is functionally effective is a difficult endeavour. A firm understanding of how the algorithm allows, for example, simple shape recognition can be viewed as crucial in extension to real-time use in the environment. Much like vision, not all sensory information in a scene is task relevant and thus needs to be

filtered to extract the salient information. The gradual building of complexity (stimulus and task) in training, particularly in feature recognition, should facilitate the extraction of task-relevant information allowing an ‘easier’ passage to real-world environments (Maidenbaum, Levy-Tzedek, Chebat, & Amedi, 2013).

Of course motivation may be degraded in other ways, especially through expectation. While phenomenological and sensory motor theories may illustrate a ‘visual’ experience and brain imaging research shows activation in visual areas of cortex, at the level of the layman the experience of sensory substitution is not of vision. This may be inconsequential to the congenitally blind with no memory of visual experience, but in late blind individuals the phenomenological comparison is stark. Researchers therefore need to be clear in explain what sensory substitution is, or more importantly, what it isn’t. That however is another area of research (Auvray & Myin, 2009).

Returning to Experiment 1 (reported in Chapter 2), the focus was on perceptual learning in a low-level sensory task using a complex signal from a sonified object. Compared to a unimodal equivalent the results demonstrated more rapid learning of the trained stimulus and an increase in the breadth of generalization to novel temporal stimuli. This contrasts with the auditory literature that almost exclusively show no generalization to temporal features, although Lapid et al found temporal generalization across stimulus types -interval to duration (Lapid et al., 2009). The demonstration of temporal generalization within stimulus type (duration to duration) as shown in Chapter 2 is, as far as I am aware, novel in the literature.

Consider first the breadth of generalization. In the auditory paradigm, employed previously by Wright and colleagues (B. A. Wright et al., 2010), generalized learning was found for untrained spectral features, such as frequency but not to temporal features, such as interval,

within the 10 day time course. The results of the experiment discussed in Chapter 2 demonstrate not only a more rapid generalization to untrained frequencies compared to the unimodal task but also a significant generalization to an untrained temporal duration, in the 10 day training group. The study did raise questions however. What facilitates this enhancement of specific and generalised perceptual learning? Is it the complexity of the signal per se, of the crossmodal nature of it? Also, within the signal, is it spectral or temporal features that are driving generalization? To a certain degree it is inconsequential as the behavioural performance is superior anyway, that is, the outcome for training is positive as sensory substitution deals with complex spectro-temporal stimuli.

The performance in the stereo condition was interesting. Naturally there is a practical advantage to using only one headphone as this allows environmental noise to be processed at the same time. However, with results better at pre-test for the stereo condition than after 12 days of training on the monaural condition there is a huge advantage to stereo input. This conflict may be counteracted by the use of bone conduction headphones which transmit the signal via the skull rather than the outer ear and thus leaving the ears open for input of sounds from the natural environment. Of course this may bring forth further issues regarding the interaction between the sonifications and environmental sounds, and how attention may shift between the two depending on salience of the signals. This needs to be tested further and indeed this is underway in the lab presently.

The final measure in the study was the long term maintenance of perceptual learning. In a follow up task, either three or six months later with no additional training, discrimination performance was maintained in both generalized and specific learning.

The application of the results in applied training are threefold; firstly, wear two headphones to fully benefit from the algorithm, secondly, the use of complex stimuli appears to drive

processing to higher levels facilitating broader generalization and thus more novel stimuli can be introduced quickly into training, thirdly long term maintenance with no further training allows training regimes to be halted and restarted with no deficit to performance.

One question not answered was, if in naïve users the signal is processed as an auditory, rather than crossmodal input, do the principles that govern the auditory system impinge on potential performance? This was evaluated in Chapter 3 where assessment was made on the amount of information required for successful object recognition in sensory substitution and whether the formation of object-based representations was limited by the auditory system.

The resolution of SSDs vary greatly between devices and yet behavioural performance on some simple tasks show a relative level of equivalence. How can this be explained? The logical theory is that for tasks such as object recognition high levels of acuity are not a requirement. From an applied perspective I assessed this in Chapter 3 by reducing the pixel resolution in the image prior to sonification and pairing the variously degraded soundscapes with full resolution tactile and visual stimuli. The ceiling effect at 8x8 pixels, after which performance stabilized, demonstrated that simple objects can be recognised in naïve users with limited feature information. This is interesting as far as retinal implants are concerned. Simulations for implants, described in Chapter 1, suggest for simple object recognition a 30x30 resolution (Weiland et al., 2005). If the results in Experiment 2 translate from sensory substitution to invasive interventions then this resolution may be overstated. This is informative in both the development of implant technology per se and the use of multiple devices in visual rehabilitation. Two devices (VA SSD + VT SSD or VA SSD + implant) should permit the transmission of more types of information (e.g. to code for depth when the appropriate auditory features (amplitude) are used for another mapping.) Further research is required to assess whether it would be advantageous to have a consistent resolution across

devices or to use the relative acuity for specific task requirements. If the former, the results imply a downgrading of the SSD to acuity levels of the implant while still retaining functionality, while for the latter device acuity is paired to task specifics, i.e. low acuity (implant) for object detection and high acuity (SSD) for fine grained feature recognition.

Theoretically, the experiment illustrated potential limitations in object formation based on phase locking at levels of the auditory hierarchy. As the pixel resolution is lowered size increases with a subsequent widening of bandwidth in the auditory signal. Different levels of the auditory hierarchy are limited in the frequencies they can process thus limiting where high fidelity objects are formed in cortex (Griffiths & Warren, 2004). However training may allow for object recognition to be occur at primary cortical levels circumventing the limitations of the higher cortical areas.

How does this impact on the literature? It is curious when reading brain imaging studies using SSDs that activation in the areas typically associated with the modality of input (e.g. auditory cortex) are given scant mention , with the focus on typical visual areas (e.g.occipital cortex). This is understandable considering 'the goal' and yet the literature also suggests that in naïve users, prior to training, cortical activation is in unimodal areas only, and thus it is learning to use the device that drives processing to visual and multisensory areas. This approach appears to give little credit to the idea that in VA sensory substitution auditory signal processing is crucial in the filtering of information prior to transmission to multisensory cortical areas and yet the present research demonstrates that, in naïve users, this is a salient factor. Of course, that trained users can successfully carry out 'visual' tasks, demonstrates that at some point the brain learns to counteract or over-ride the limitations in the modality of input to give a multisensory or 'visual' percept.

The magnitude of the potential limitations in auditory object formation were exemplified in Chapter 4, where I further explored the relevance of auditory scene analysis to sensory substitution. With horizontal lines in The vOICe coded as consistent frequencies across the time course, concurrent lines require frequency discrimination for segregation into separate objects. Considering the fine resolution of the auditory system, exemplified by JND research, grouping by proximity shouldn't be difficult and yet segregation into separate objects was problematic for parallel lines that showed auditory consonance. The results imply that in the rendering of objects, early auditory processes of harmonic grouping dominate grouping and segregation resulting in potential misidentification of objects. In the second task congruent audio-visual information increased performance, as would be expected in a spatial task, but primarily only for dissonant stimuli. This demonstrates the strength of the effect of harmonicity in feature grouping into objects (A. S. Bregman et al., 1990).

The experiment was devised to highlight a problem found in basic shape training using SSDs. If training uses simple static 2D objects the soundscape is consistent for each presentation and thus any harmonic conflicts are retained. This can easily be accounted for in two ways. If using virtual objects, provide multiple presentations of the object at various sizes as this will negate or shift consonant frequency based interference. Secondly, if using the device in real time, head movement or zoom will achieve the same effect.

While the effect of harmonicity was shown in a simple paradigm, unlikely to be encountered many times in real world use, future research can extend these findings to see if they apply in more complex objects. For example, consonant frequencies can be removed from sonified objects and compared to the full object to evaluate whether this impacts on complex object formation. This could be taken further by splitting the consonant and dissonant frequency components and presenting them successively to see if this further aids object recognition.

The results can be extrapolated to paradigms not using sensory substitution but utilising sonifications in a static environment, e.g. the sonification of flow charts and graphs in digital formats as discussed in Chapter 4.

The splitting of the soundscape was used in Experiment 5 to again evaluate the effect of signal complexity in sensory substitution and draws comparisons with what was found in the resolution experiment in Chapter 3. In the former experiment it was demonstrated that simple object recognition can be successful with a degraded level of input, while in the final study the other end of the continuum was investigated: At what point does additional information not just become superfluous but actively negate behavioural performance? Of all the experiments in the thesis the results of this one are most directly applicable to training. The type of presentation was used to evaluate cognitive load in an object recognition task, with either a high density presentation where all information was in a 'one shot' format, or a lower load presentation where information was split in two based on frequency. In both the visual and tactile matching tasks performance was significantly better for the low-load condition. Theoretically this suggests processing limitations, although imbalance in the paradigm (e.g. the total duration of presentation of a single object) may be influential. The results of the supplementary experiment reported at the end of Chapter 5 imply that while the paradigm was influential in Chapter 5 (main) there is still a significant effect of cognitive load when the paradigms are balanced. Irrespective, the strength of the results show that the paradigm can be directly applied to training. This can be done as in the experiment by varying the presentation of online stimuli, or by filtering the output of the device to attenuate high frequencies (2500-5000hz) in the first scan and low frequencies (500-2499Hz) in the second scan. It would be interesting to evaluate whether this could also be achieved be getting the user to self-direct attention to the high frequencies for the first scan of the full signal, and low frequencies for the second, thus re-training the brain rather than adapting the technology.

Aside from the ideas for future research described for each experiment, general future research in all tasks should be extended to different populations. For example, would the impact of auditory signal processing be found in the blind, or a trained sighted user group, or even in trained musicians? The latter group have been shown to be superior at basic tasks using SSDs (Haigh et al., 2013), most likely down to superior frequency discrimination, and this can be applied to training on VA SSDs. Simple musical tasks such as adaptive frequency discrimination, or temporal pattern identification should build general auditory ability and transfer to better performance using SSDs. This type of training task may also to serve to break potential monotony of training, maintaining motivation while at the same time training the brain to be a more effective processor of sensory information.

Prior to a final summary, limitations of the research should be addressed. The most obvious limitation is the demographic of the participants. In all four experiments sighted, normally blindfolded, listeners were used rather than the target user group, the visually impaired. There are negative implications of this but also justifications. Blindness is generally a degenerative disorder over time, hence most sufferers are elderly. Ideally we would use blind populations in each experiment or match the sighted participants on demographics such as age and gender. However, due to the locale of lab based experiments and methodological design this was problematic. Multiple-stage experiments require the participant to attend on a number of days. In a large conurbation like London this would require the blind participants to navigate in unfamiliar environments for little monetary benefit. Availability of sighted students however in a university setting is higher but also suffers from limitations. Firstly: the student population tends to be younger than the blind population and therefore we have to consider differences in neural development based on age. Secondly, of course, they are all sighted. Is there justification for using this population beyond availability?

Behavioural and neural differences between sighted, early blind, and congenitally blind have been demonstrated on a number of tasks. However, when contrasting across participant groups, performance in the early blind is often more similar to that found in blindfolded sighted populations, than the congenitally blind. This not only implies a heavy weighting to the benefit of previous visual experience, but also that similarities between late blind and sighted gives validates for the choice of participants. Secondly, most of the research was 'proof of concept' in naïve users. It is important to establish whether the theory and methods hold ground before extending the research to the visually impaired populations, whose time and availability may be more limited. Obviously it is important that any proof of concept displays methods of application, hence I tried to incorporate an auditory-tactile equivalent for audio-visual tasks, aside from when there was no visual stimuli involved (e.g. Chapter 2).

From a research perspective, SSDs provide a valuable tool for assessing crossmodal perceptual processes in the general population and differentials based on visual impairment. It would be advantageous to have a trained sighted user group who regularly use the device to contrast with blind users to assess the development in these groups. This need not be non-applied to the sighted either. The techniques used in sensory substitution can be utilised for sensory augmentation in which the extra information provided by the device extends the capabilities of an unimpaired sensory system rather than substitute for an impairment. For example, to provide 360^0 'vision'.

Finally there is inclusivity. Most technology designed for the blind population is designed to allow functioning in the visual world, that is, assisting in tasks that 'require' vision. The white cane to avoid obstacles, Braille for reading, SSDs for object recognition etc. Interestingly they are also used for fun, for example one long term congenitally blind user PL uses The vOICe to line up scenes to photograph. If we are to work, and play, together then

the adaptation doesn't have to be unidirectional. If a blind person is using technology for a task then an understanding of how the technology is being used would be beneficial to a sighted co-worker, friend (J. L. Gomez, Langdon, Bichard, & Clarkson, 2014).

Another limitation is the small sample used in the experiments, as this dictates the power and precision of the result estimations, that is, how much can the main effects be attributed to the manipulated variable rather than the randomness of the sample. Naturally as the sample size increases so does confidence in the estimations; we have more data points have to estimate less, and therefore have more power to detect differences. For this thesis, sample size was constrained by logistical factors such as costs on time and money, and availability of facilities, hence large effect sizes were of primary interest in analysis. This is not to say significant results with medium and small effect sizes are irrelevant as they may point to trends which can be confirmed or refuted in further research.

The simple solution to increase the confidence in the results is to increase the sample size. Given more time and resources this would have been applied here. Interestingly, limitations on sample size are also pertinent in the target population, the visually impaired. Not down to money however, but down to availability. Even in large cities like London, the prevalence of blindness is low, especially congenitally blind. Furthermore, this small population is massively reduced if only considering those who are willing and able to attend the lab, often for multiple sessions.

Increasing the availability of visually impaired participants for research is a present topic of discussion in the field of research. Labs conducting research into non-invasive methods of visual rehabilitation are dispersed across the globe making physical sharing of participants a non-starter. However, it is possible to design paradigms that translate easily across virtual

space and thus can be utilized in all labs. For example, sharing code and design for lab or computer based tasks to run in all labs, thus building a network of visually impaired participants. Another simple method of increasing participant numbers is to capitalise on the popularity of smartphones. Simple training techniques can be built into apps, perhaps in the form of games in which the visually impaired community can 'compete' against each other. Hopefully this can be not only effective training, but a source of data, and ultimately an enjoyable experience for the user (Prensky, 2005). In our lab at present we are developing a multiplatform app for just this purpose. More directly, experiments can be designed to be portable. If computer based the training can be easily taken to the participant rather than vice versa.

Final Summary and take home message.

In a series of four experiments I evaluated performance in simple tasks in naïve users of a visual-to-auditory sensory substitution device. The results demonstrated that first, and most importantly, naive users performed above chance in all tasks implying an advantage of using the device. Second, in naïve users auditory characteristics are crucial in object recognition and that signal complexity can drive processing to higher-order cortical areas that increase the breadth of generalized learning. Thirdly, if the signal is being processed as an auditory stimulus it is subjected to limitations in processing found in auditory scene analysis, which may limit object formation and feature segregation, and that this may be circumvented by perceptual training. Fourthly, there are apparent limitations in capacity that can be accounted for by the type of presentation used in training.

I hope the work adds to the body of literature in an interesting and worthwhile area of research, and encourages researchers to think before diving straight into the occipital cortex, there are potentially important processes happening elsewhere!

Encouragingly, even with these potential limitations at the early stage, the brain seems to adapt quickly, shown by successful behavioural results. This can only be good news for the visually impaired community.

7.0.0 References.

Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S., & Amedi, A. (2014). EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restor Neurol Neurosci, 32*(2), 247-257. doi: Doi 10.3233/Rnn-130338

Ahissar, M. (2001). Perceptual training: a tool for both modifying the brain and exploring it. *Proc Natl Acad Sci U S A, 98*(21), 11842-11843. doi: 10.1073/pnas.221461598

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci, 8*(10), 457-464. doi: 10.1016/j.tics.2004.08.011

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos Trans R Soc Lond B Biol Sci, 364*(1515), 285-299. doi: 10.1098/rstb.2008.0253

Alain, C. (2007). Breaking the wave: effects of attention and learning on concurrent sound perception. *Hear Res, 229*(1-2), 225-236. doi: 10.1016/j.heares.2007.01.011

Alais, D., & Burr, D. (2004a). No direction-specific bimodal facilitation for audiovisual motion detection. *Brain Res Cogn Brain Res, 19*(2), 185-194. doi: 10.1016/j.cogbrainres.2003.11.011

Alais, D., & Burr, D. (2004b). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol, 14*(3), 257-262. doi: 10.1016/j.cub.2004.01.029

Altes, R. A. (1976). Sonar for generalized target description and its similarity to animal echolocation systems. *J Acoust Soc Am, 59*(1), 97-105.

Amedi, A., Floel, A., Knecht, S., Zohary, E., & Cohen, L. G. (2004). Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nat Neurosci, 7*(11), 1266-1270. doi: 10.1038/nn1328

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., . . . Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci, 10*(6), 687-689. doi: Doi 10.1038/Nn1912

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Percept Psychophys, 67*(4), 691-698.

Arno, P., Capelle, C., Wanet-Defalque, M. C., Catalan-Ahumada, M., & Veraart, C. (1999). Auditory coding of visual patterns for the blind. *Perception, 28*(8), 1013-1029.

Arno, P., De Volder, A. G., Vanlierde, A., Wanet-Defalque, M. C., Streel, E., Robert, A., . . . Veraart, C. (2001). Occipital activation by pattern recognition in the early blind using auditory substitution for vision. *Neuroimage, 13*(4), 632-645. doi: 10.1006/nimg.2000.0731

Arno, P., Vanlierde, A., Streel, E., Wanet-Defalque, M. C., Sanabria-Bohorquez, S., & Veraart, C. (2001). Auditory substitution of vision: Pattern recognition by the blind. *Applied Cognitive Psychology, 15*(5), 509-519.

Arnoldussen, A., & Fletcher, D. C. (2012). Visual Perception for the Blind: The BrainPort Vision Device. *Retinal Physician, 9*(1), 32-34.

Attarha, M., Moore, C. M., Scharff, A., & Palmer, J. (2014). Evidence of unlimited-capacity surface completion. *J Exp Psychol Hum Percept Perform, 40*(2), 556-565. doi: 10.1037/a0034594

Auvray, M., Hanneton, S., Lenay, C., & O'Regan, K. (2005). There is something out there: distal attribution in sensory substitution, twenty years later. *J Integr Neurosci, 4*(4), 505-521.

Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: localisation and object recognition with 'the vOICe'. *Perception, 36*(3), 416-430.

Auvray, M., & Myin, E. (2009). Perception with compensatory devices: from sensory substitution to sensorimotor extension. *Cogn Sci, 33*(6), 1036-1058. doi: 10.1111/j.1551-6709.2009.01040.x

Bach-y-Rita, P. (1968). [Various neurophysiological findings of sensory substitution]. *Acta Neurol Latinoam, 14*(1), 125-131.

Bach-y-Rita, P. (2002). Sensory substitution and qualia. *Vision and mind*, 497-514.

Bach-y-Rita, P. (2004). Tactile sensory substitution studies. *Ann N Y Acad Sci, 1013*, 83-91.

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature, 221*(5184), 963-964.

Bach-y-Rita, P., Collins, C. C., White, B., Saunders, F. A., Scadden, L., & Blomberg, R. (1969). A tactile vision substitution system. *Am J Optom Arch Am Acad Optom, 46*(2), 109-111.

Bach-y-Rita, P., Kaczmarek, K. A., Tyler, M. E., & Garcia-Lara, J. (1998). Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *J Rehabil Res Dev, 35*(4), 427-430.

Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences, 7*(12), 541-546. doi: DOI 10.1016/j.tics.2003.10.013

Bach-y-Rita, P., & S, W. K. (2003). Sensory substitution and the human-machine interface. *Trends Cogn Sci, 7*(12), 541-546.

Bach-y-Rita, P., & Tyler, M. E. (2000). Tongue man-machine interface. *Stud Health Technol Inform, 70*, 17-19.

Ball, G. F., & Hulse, S. H. (1998). Birdsong. *American Psychologist, 53*(1), 37.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Res, 27*(6), 953-965.

Bartolo, R., & Merchant, H. (2009). Learning and generalization of time production in humans: rules of transfer across modalities and interval durations. *Exp Brain Res, 197*(1), 91-100. doi: 10.1007/s00221-009-1895-1

Barton, B., Venezia, J. H., Saberi, K., Hickok, G., & Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proc Natl Acad Sci U S A, 109*(50), 20738-20743. doi: DOI 10.1073/pnas.1213381109

Battaglia-Mayer, A., Ferraina, S., Genovesio, A., Marconi, B., Squatrito, S., Molinari, M., . . . Caminiti, R. (2001). Eye-hand coordination during reaching. II. An analysis of the relationships between visuomanual signals in parietal cortex and parieto-frontal association projections. *Cereb Cortex, 11*(6), 528-544.

Baumann, S., Griffiths, T. D., Sun, L., Petkov, C. I., Thiele, A., & Rees, A. (2011). Orthogonal representation of sound dimensions in the primate midbrain. *Nat Neurosci, 14*(4), 423-425. doi: Doi 10.1038/Nn.2771

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept Psychophys, 57*(8), 1151-1162.

Ben-Av, M. B., Sagi, D., & Braun, J. (1992). Visual attention and perceptual grouping. *Percept Psychophys, 52*(3), 277-294.

Benav, H., Bartz-Schmidt, K. U., Besch, D., Bruckmann, A., Gekeler, F., Greppmaier, U., . . . Zrenner, E. (2010). Restoration of useful vision up to letter recognition capabilities using subretinal microphotodiodes. *Conf Proc IEEE Eng Med Biol Soc, 2010*, 5919-5922. doi: 10.1109/IEMBS.2010.5627549

Bermant, R. I., & Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Percept Mot Skills, 42*(43), 487-493. doi: 10.2466/pms.1976.42.2.487

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J Exp Psychol, 87*(2), 241-247.

Berry, R. N. (1948). Quantitative relations among vernier, real depth, and stereoscopic depth acuities. *J Exp Psychol, 38*(6), 708-721.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychon Bull Rev, 5*(3), 482-489.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept Psychophys, 29*(6), 578-584.

Bliss, J. C., Crane, H. D., Mansfield, P. K., & Townsend, J. T. (1966). Information available in brief tactile presentations. *Perception & Psychophysics, 1*(4), 273-283.

Bliss, J. C., Katcher, M. H., Rogers, C. H., & Shepard, R. P. (1970). Optical-to-tactile image conversion for the blind. *Man-Machine Systems, IEEE Transactions on, 11*(1), 58-65.

Borisoff, J. F., Elliott, S. L., Hocaloski, S., & Birch, G. E. (2010). The development of a sensory substitution system for the sexual rehabilitation of men with chronic spinal cord injury. *J Sex Med, 7*(11), 3647-3658. doi: 10.1111/j.1743-6109.2010.01997.x

Boroojerdi, B., Bushara, K. O., Corwell, B., Immisch, I., Battaglia, F., Muellbacher, W., & Cohen, L. G. (2000). Enhanced excitability of the human visual cortex induced by short-term light deprivation. *Cereb Cortex, 10*(5), 529-534.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Percept Psychophys, 61*(5), 977-985.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*: MIT press.

Bregman, A. S., Levitan, R., & Liao, C. (1990). Fusion of auditory components: effects of the frequency of amplitude modulation. *Percept Psychophys, 47*(1), 68-73.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Can J Psychol, 44*(3), 400-413.

Brelen, M. E., Duret, F., Gerard, B., Delbeke, J., & Veraart, C. (2005). Creating a meaningful visual perception in blind volunteers by optic nerve stimulation. *J Neural Eng, 2*(1), S22-28. doi: 10.1088/1741-2560/2/1/004

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition, 126*(2), 165-172.

Brindley, G. S., & Lewin, W. S. (1968a). The sensations produced by electrical stimulation of the visual cortex. *J Physiol, 196*(2), 479-493.

Brindley, G. S., & Lewin, W. S. (1968b). The visual sensations produced by electrical stimulation of the medial occipital cortex. *J Physiol, 194*(2), 54-55P.

Broadbent, D. E. (1958). Perception and communication: London: Pergamon Press.

Brown, D. J., Macpherson, T., & Ward, J. (2011). Seeing with sound? exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception, 40*(9), 1120-1135.

Brown, D. J., & Proulx, M. J. (2013). Increased Signal Complexity Improves the Breadth of Generalization in Auditory Perceptual Learning. *Neural Plasticity*. doi: Artn 879047

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Exp Brain Res, 198*(1), 49-57. doi: 10.1007/s00221-009-1933-z

Burton, H., Snyder, A. Z., Conturo, T. E., Akbudak, E., Ollinger, J. M., & Raichle, M. E. (2002). Adaptive changes in early and late blind: a fMRI study of Braille reading. *J Neurophysiol, 87*(1), 589-607.

Burton, H., Snyder, A. Z., Diamond, J. B., & Raichle, M. E. (2002). Adaptive changes in early and late blind: a FMRI study of verb generation to heard nouns. *J Neurophysiol, 88*(6), 3359-3371. doi: 10.1152/jn.00129.2002

Capalbo, Z., & Glenney, B. (2009). *Hearing color: radical plurastic realism and SSDs.* Paper presented at the Proceedings of the Fifth Asia-Pacific Computing and Philosophy Conference (AP-CAP 2009), Tokyo, Japan.

Capelle, C., Trullemans, C., Arno, P., & Veraart, C. (1998). A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution. *IEEE Trans Biomed Eng, 45*(10), 1279-1293. doi: 10.1109/10.720206

Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition, 102*(3), 321-340. doi: 10.1016/j.cognition.2006.01.002

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cogn Psychol, 22*(2), 225-271.

Chader, G. J., Weiland, J., & Humayun, M. S. (2009). Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis. *Prog Brain Res, 175*, 317-332. doi: 10.1016/S0079-6123(09)17522-2

Chai, X., Li, L., Wu, K., Zhou, C., Cao, P., & Ren, Q. (2008). C-sight visual prostheses for the blind. *IEEE Eng Med Biol Mag, 27*(5), 20-28. doi: 10.1109/MEMB.2008.923959

Chai, X., Yu, W., Wang, J., Zhao, Y., Cai, C. S., & Ren, Q. S. (2007). Recognition of pixelized chinese characters using simulated prosthetic vision. *Artificial Organs, 31*(3), 175-182. doi: DOI 10.1111/j.1525-1594.2007.00362.x

Chai, X., Zhang, L., Li, W., Shao, F., Yang, K., & Ren, Q. (2008). Study of tactile perception based on phosphene positioning using simulated prosthetic vision. *Artif Organs, 32*(2), 110-115. doi: 10.1111/j.1525-1594.2007.00469.x

Chebat, D. R., Rainville, C., Kupers, R., & Ptito, M. (2007). Tactile-'visual' acuity of the tongue in early blind individuals. *Neuroreport, 18*(18), 1901-1904.

Chebat, D. R., Schneider, F. C., Kupers, R., & Ptito, M. (2011). Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport, 22*(7), 342-347. doi: 10.1097/WNR.0b013e3283462def

Chekhchoukh, A., Goumidi, M., Vuillerme, N., Payan, Y., & Glade, N. (2013). *Electrotactile vision substitution for 3D trajectory following.* Paper presented at the Engineering in

Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE.

Cherry, C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *J Acoust Soc Am, 25*(5), 975-979. doi: citeulike-article-id:3057226

Chuang, A. T., Margo, C. E., & Greenberg, P. B. (2014). Retinal implants: a systematic review. *Br J Ophthalmol, 98*(7), 852-856. doi: 10.1136/bjophthalmol-2013-303708

Clark, A. (2004). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*: Oxford University Press.

Clark, H. H., & Brownell, H. H. (1976). Position, direction, and their perceptual integrality. *Perception & Psychophysics, 19*(4), 328-334.

Cohen, L. G., Weeks, R. A., Sadato, N., Celnik, P., Ishii, K., & Hallett, M. (1999). Period of susceptibility for cross-modal plasticity in the blind. *Ann Neurol, 45*(4), 451-460.

Collignon, O., Lassonde, M., Lepore, F., Bastien, D., & Veraart, C. (2007). Functional cerebral reorganization for auditory spatial processing and auditory substitution of vision in early blind subjects. *Cereb Cortex, 17*(2), 457-465. doi: 10.1093/cercor/bhj162

Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Exp Brain Res, 192*(3), 343-358. doi: 10.1007/s00221-008-1553-z

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J Cogn Neurosci, 16*(6), 1000-1009. doi: 10.1162/0898929041502733

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Dev, 61*(5), 1584-1595.

Craddock, M., & Lawson, R. (2008). Repetition priming and the haptic recognition of familiar and unfamiliar objects. *Percept Psychophys, 70*(7), 1350-1365. doi: 10.3758/PP.70.7.1350

Craig, J. C. (1981). Tactile letter recognition: Pattern duration and modes of pattern generation. *Perception & Psychophysics, 30*(6), 540-546.

Craig, J. C. (1983). Some factors affecting tactile pattern recognition. *International Journal of Neuroscience, 19*(1-4), 47-57.

Crandell, J. M., & Wallace, D. H. (1974). Speed reading in braille: An empirical study. *New Outlook for the Blind*.

Cronly-Dillon, J., Persaud, K., & Gregory, R. P. (1999). The perception of visual images encoded in musical form: a study in cross-modality information transfer. *Proc Biol Sci, 266*(1436), 2427-2433. doi: 10.1098/rspb.1999.0942

Dagnelie, G., Keane, P., Narla, V., Yang, L., Weiland, J., & Humayun, M. (2007). Real and virtual mobility performance in simulated prosthetic vision. *J Neural Eng, 4*(1), S92-101. doi: 10.1088/1741-2560/4/1/S11

Danilov, Y., & Tyler, M. (2005). Brainport: an alternative input to the brain. *J Integr Neurosci, 4*(4), 537-550.

Danilov, Y. P., & Tyler, M. (2005). Brainport: An Alternative Input to the Brain. *Journal of Integrative Neuroscience, 04*(04), 537-550. doi: doi:10.1142/S0219635205000914

Danilov, Y. P., Tyler, M. E., Skinner, K. L., & Bach-y-Rita, P. (2006). Efficacy of electrotactile vestibular substitution in patients with bilateral vestibular and central balance loss. *Conf Proc IEEE Eng Med Biol Soc, Suppl*, 6605-6609. doi: 10.1109/IEMBS.2006.260899

Danilov, Y. P., Tyler, M. E., Skinner, K. L., Hogle, R. A., & Bach-y-Rita, P. (2007). Efficacy of electrotactile vestibular substitution in patients with peripheral and central vestibular loss. *J Vestib Res, 17*(2-3), 119-130.

Das, A., & Huxlin, K. R. (2010). New approaches to visual rehabilitation for cortical blindness: outcomes and putative mechanisms. *Neuroscientist, 16*(4), 374-387. doi: 10.1177/1073858409356112

Davies, J. B., & Davies, J. B. (1978). *The psychology of music*: Hutchinson London.

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *Br J Psychol, 52*, 259-268.

De Volder, A. G., Bol, A., Blin, J., Robert, A., Arno, P., Grandin, C., . . . Veraart, C. (1997). Brain energy metabolism in early blind subjects: neural activity in the visual cortex. *Brain Res, 750*(1-2), 235-244.

Delbeke, J., Oozeer, M., & Veraart, C. (2003). Position, size and luminosity of phosphenes generated by direct optic nerve stimulation. *Vision Res, 43*(9), 1091-1102.

Delbeke, J., Pins, D., Michaux, G., Wanet-Defalque, M. C., Parrini, S., & Veraart, C. (2001). Electrical stimulation of anterior visual pathways in retinitis pigmentosa. *Invest Ophthalmol Vis Sci, 42*(1), 291-297.

Delbeke, J., Wanet-Defalque, M. C., Gérard, B., Troosters, M., Michaux, G., & Veraart, C. (2002). The Microsystems Based Visual Prosthesis for Optic Nerve Stimulation. *Artificial Organs, 26*(3), 232-234. doi: 10.1046/j.1525-1594.2002.06939.x

Delhommeau, K., Micheyl, C., Jouvent, R., & Collet, L. (2002). Transfer of learning across durations and ears in auditory frequency discrimination. *Percept Psychophys, 64*(3), 426-436.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *J Acoust Soc Am, 111*(3), 1377-1388.

Deutsch, J. A., Deutsch, D., Lindsay, P. H., & Treisman, A. (1967). Comments and reply on "Selective attention: perception or response? *Q J Exp Psychol, 19*(4), 362-367. doi: 10.1080/14640746708400117

DeWitt, L. A., & Crowder, R. G. (1987). Tonal fusion of consonant musical intervals: The oomph in Stumpf. *Percept Psychophys, 41*(1), 73-84.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci U S A, 109*(29), 11854-11859. doi: DOI 10.1073/pnas.1205381109

Dobelle, W. H., & Mladejovsky, M. G. (1974). Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind. *J Physiol, 243*(2), 553-576.

Dobelle, W. H., Mladejovsky, M. G., & Girvin, J. P. (1974). Artifical vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis. *Science, 183*(4123), 440-444.

Doucet, M. E., Guillemot, J. P., Lassonde, M., Gagne, J. P., Leclerc, C., & Lepore, F. (2005). Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Exp Brain Res, 160*(2), 194-202. doi: 10.1007/s00221-004-2000-4

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience, 22*(1), 567-631.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: the spotlight metaphor breaks down. *J Exp Psychol Hum Percept Perform, 15*(3), 448-456.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends Cogn Sci, 2*(7), 254-262. doi: 10.1016/S1364-6613(98)01188-7

Durette, B., Louveton, N., Alleysson, D., & Hérault, J. (2008). *Visuo-auditory sensory substitution for mobility assistance: testing TheVIBE.* Paper presented at the Workshop on Computer Vision Applications for the Visually Impaired.

Eickenscheidt, M., Jenkner, M., Thewes, R., Fromherz, P., & Zeck, G. (2012). Electrical stimulation of retinal neurons in epiretinal and subretinal configuration using a multicapacitor array. *J Neurophysiol, 107*(10), 2742-2755. doi: 10.1152/jn.00909.2011

Elbert, T., Sterr, A., Rockstroh, B., Pantev, C., Muller, M. M., & Taub, E. (2002). Expansion of the tonotopic area in the auditory cortex of the blind. *J Neurosci, 22*(22), 9941-9944.

Epstein, W. (1985). Amodal information and transmodal perception. . In D. H. Warren & E. R. Strelow (Eds.), *Electronic spatial sensing for the blind* (pp. 421-430). Dordrecht, The Netherlands: Matinus Nijhoff.

Epstein, W., Hughes, B., Schneider, S., & Bach-y-Rita, P. (1986). Is there anything out there? A study of distal attribution in response to vibrotactile stimulation. *Perception, 15*(3), 275-284.

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: some results and methodological considerations. *J Exp Psychol, 79*(2), 1-16.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1). doi: Artn 6

Facchini, S., & Aglioti, S. M. (2003). Short term light deprivation increases tactile spatial acuity in humans. *Neurology, 60*(12), 1998-1999.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Res, 35*(21), 3003-3013.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci, 22*(13), 5749-5759. doi: 20026562

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Percept Psychophys, 63*(4), 719-725.

Fenn, K. M., Gallo, D. A., Margoliash, D., Roediger, H. L., 3rd, & Nusbaum, H. C. (2009). Reduced false memory after sleep. *Learn Mem, 16*(9), 509-513. doi: 10.1101/lm.1500808

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant behavior and development, 8*(2), 181-195.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant behavior and development, 10*(3), 279-293.

Fernandes, R. A., Diniz, B., Ribeiro, R., & Humayun, M. (2012). Artificial vision through neuronal stimulation. *Neurosci Lett, 519*(2), 122-128. doi: 10.1016/j.neulet.2012.01.063

Fiorentini, A., & Berardi, N. (1980). Perceptual-Learning Specific for Orientation and Spatial-Frequency. *Nature, 287*(5777), 43-44. doi: Doi 10.1038/287043a0

Fiorentini, A., & Berardi, N. (1981). Learning in Grating Waveform Discrimination - Specificity for Orientation and Spatial-Frequency. *Vision Res, 21*(7), 1149-1158. doi: Doi 10.1016/0042-6989(81)90017-1

Fitch, R. H., Miller, S., & Tallal, P. (1997). Neurobiology of speech perception. *Annual review of neuroscience, 20*(1), 331-353.

Fitzgerald, M. B., & Wright, B. A. (2005). A perceptual learning investigation of the pitch elicited by amplitude-modulated noise. *Journal of the Acoustical Society of America, 118*(6), 3794-3803. doi: Doi 10.1121/1.2074687

Foerster, O. (1929). Beitrage zur Pathophysiologie der Sehbahn und der Sehsphare. *Journal fur Psychologie und Neurologie, 39*, 463-485.

Fornos, A. P., Sommerhalder, J., & Pelizzone, M. (2011). Reading with a simulated 60-channel implant. *Front Neurosci, 5*, 57. doi: 10.3389/fnins.2011.00057

Fortin, C., & Rousseau, R. (1998). Interference from short-term memory processing on encoding and reproducing brief durations. *Psychol Res, 61*(4), 269-276.

Foulke, E. (1982). Reading braille. *Tactual perception: A sourcebook*, 168.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res, 147*(3), 332-343. doi: 10.1007/s00221-002-1262-y

Frisken-Gibson, S. F., Bach-y-Rita, P., Tompkins, W. J., & Webster, J. G. (1987). A 64-solenoid, four-level fingertip search display for the blind. *IEEE Trans Biomed Eng, 34*(12), 963-965.

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic–auditory associations in the "Bouba–Kiki" effect. *Cognition, 132*(2), 164-173.

Fujikado, T., Kamei, M., Sakaguchi, H., Kanda, H., Morimoto, T., Ikuno, Y., . . . Nishida, K. (2011). Testing of semichronically implanted retinal prosthesis by suprachoroidal-transretinal stimulation in patients with retinitis pigmentosa. *Invest Ophthalmol Vis Sci, 52*(7), 4726-4733. doi: 10.1167/iovs.10-6836

Gagnon, L., Schneider, F. C., Siebner, H. R., Paulson, O. B., Kupers, R., & Ptito, M. (2012). Activation of the hippocampal complex during tactile maze solving in congenitally blind subjects. *Neuropsychologia, 50*(7), 1663-1671. doi: 10.1016/j.neuropsychologia.2012.03.022

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Percept Psychophys, 68*(7), 1191-1203.

Gibby, R. G., Jr., Gibby, R. G., Sr., & Townsend, J. C. (1970). Short-term visual restriction in visual and auditory discrimination. *Percept Mot Skills, 30*(1), 15-21. doi: 10.2466/pms.1970.30.1.15

Goertz, Y. H. H., van Lierop, A. G., Houkes, I., & Nijhuis, F. J. N. (2010). Factors related to the employment of visually impaired persons: A systematic literature review. *Journal of Visual Impairment & Blindness, 104*(7), 404-418.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12*(6), 878-892.

Gomez, J. D., Bologna, G., & Pun, T. (2014). See ColOr: an extended sensory substitution device for the visually impaired. *Journal of Assistive Technologies, 8*(2), 77-94.

Gomez, J. L., Langdon, P. M., Bichard, J. A., & Clarkson, P. J. (2014). Designing Accessible Workplaces for Visually Impaired People *Inclusive Designing* (pp. 269-279): Springer.

Gougoux, F., Belin, P., Voss, P., Lepore, F., Lassonde, M., & Zatorre, R. J. (2009). Voice perception in blind persons: a functional magnetic resonance imaging study. *Neuropsychologia, 47*(13), 2967-2974. doi: 10.1016/j.neuropsychologia.2009.06.027

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., & Belin, P. (2004). Neuropsychology: pitch discrimination in the early blind. *Nature, 430*(6997), 309. doi: 10.1038/430309a

Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., & Lepore, F. (2005). A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS Biol, 3*(2), e27. doi: 10.1371/journal.pbio.0030027

Gould, E., Negus, N. C., & Novick, A. (1964). Evidence for echolocation in shrews. *Journal of Experimental Zoology, 156*(1), 19-37.

Grant, A. C., Thiagarajah, M. C., & Sathian, K. (2000). Tactile perception in blind Braille readers: a psychophysical study of acuity and hyperacuity using gratings and dot patterns. *Percept Psychophys, 62*(2), 301-312.

Graulty, C., Papaioannou, O., Bauer, P., Pitts, M., & Canseco-Gonzalez, E. (2014). Electrophysiological Dynamics of Auditory-Visual Sensory Substitution. *Journal of Vision, 14*(10), 438. doi: 10.1167/14.10.438

Grice, H. P. (1962). *Some remarks about the senses*: The Senses, cit.

Griffin, D. R., & Thompson, D. (1982). Echolocation by cave swiftlets. *Behavioral Ecology and Sociobiology, 10*(2), 119-123.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nat Rev Neurosci, 5*(11), 887-892. doi: 10.1038/nrn1538

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res, 41*(10-11), 1409-1422.

Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., & Collet, L. (2003). Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination. *Hear Res, 184*(1-2), 41-50.

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Percept Psychophys, 54*(3), 383-394.

Haigh, A., Brown, D. J., Meijer, P., & Proulx, M. J. (2013). How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Front Psychol, 4*, 330. doi: 10.3389/fpsyg.2013.00330

Hanneton, S., Auvray, M., & Durette, B. (2010). The Vibe: a versatile vision-to-audition sensory substitution device. *Applied Bionics and Biomechanics, 7*(4), 269-276.

Hertz, U., & Amedi, A. (2014). Flexibility and Stability in Sensory Processing Revealed Using Visual-to-Auditory Sensory Substitution. *Cereb Cortex*. doi: 10.1093/cercor/bhu010

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *J Exp Psychol Hum Percept Perform, 37*(1), 283-295. doi: 10.1037/a0020772

Heyes, A. D. (1984). Sonic Pathfinder - a Programmable Guidance Aid for the Blind. *Electronics & Wireless World, 90*(1579), 26-&.

Hinton, L., Nichols, J., & Ohala, J. J. (2006). *Sound symbolism*: Cambridge University Press.

Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition, 94*(3), B101-111. doi: 10.1016/j.cognition.2004.06.006

Huang, L., Pashler, H., & Junge, J. A. (2004). Are there capacity limitations in symmetry perception? *Psychon Bull Rev, 11*(5), 862-869.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *J Physiol, 206*(2), 419-436.

Hughes, B. (2001). Active artificial echolocation and the nonvisual perception of aperture passability. *Hum Mov Sci, 20*(4-5), 371-400.

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage, 50*(3), 1202-1211. doi: 10.1016/j.neuroimage.2010.01.046

Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychon Bull Rev, 15*(3), 548-554.

Irvine, D. R., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *J Acoust Soc Am, 108*(6), 2964-2968.

Jain, A., Fuller, S., & Backus, B. T. (2010). Absence of cue-recruitment for extrinsic signals: sounds, spots, and swirling dots fail to influence perceived 3D rotation direction after training. *PLoS One, 5*(10), e13295. doi: 10.1371/journal.pone.0013295

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 683.

Jones, G., & Teeling, E. C. (2006). The evolution of echolocation in bats. *Trends in Ecology & Evolution, 21*(3), 149-156.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews, 84*(2), 541-577. doi: DOI 10.1152/physrev.00029.2003

Kaas, J. H. (2000). The reorganization of somatosensory and motor cortex after peripheral nerve or spinal cord injury in primates. *Prog Brain Res, 128*, 173-179. doi: 10.1016/S0079-6123(00)28015-1

Kaczmarek, K. A. (2011). The tongue display unit (TDU) for electrotactile spatiotemporal pattern presentation. *Scientia Iranica, 18*(6), 1476-1485.

Kane, D., Grassi, W., Sturrock, R., & Balint, P. (2004). A brief history of musculoskeletal ultrasound:'From bats and ships to babies and hips'. *Rheumatology, 43*(7), 931-933.

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learn Mem, 10*(2), 141-147. doi: 10.1101/lm.55503

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc Natl Acad Sci U S A, 88*(11), 4966-4970.

Kauffman, T., Theoret, H., & Pascual-Leone, A. (2002). Braille character discrimination in blindfolded human subjects. *Neuroreport, 13*(5), 571-574.

Kay, L. (1964). An ultrasonic sensing probe as a mobility aid for the Blind. Ultrasonics. *Ultrasonics, 2*(2), 53-59. doi: 10.1016/0041-624X(64)90382-8

Kay, L. (1985). Sensory aids to spatial perception for blind persons: Their design and evaluation. In D. H. Warren & E. R. Strelow (Eds.), *Electronic spatial sensing for the blind* (pp. 125-139). Dordrecht, The Netherlands: Martinus Nijhoff.

Keseru, M., Feucht, M., Bornfeld, N., Laube, T., Walter, P., Rossler, G., . . . Richard, G. (2012). Acute electrical stimulation of the human retina with an epiretinal electrode array. *Acta Ophthalmol, 90*(1), e1-8. doi: 10.1111/j.1755-3768.2011.02288.x

Kim, J. K., & Zatorre, R. J. (2008). Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Res, 1242*, 263-275. doi: 10.1016/j.brainres.2008.06.038

Kim, J. K., & Zatorre, R. J. (2010). Can you hear shapes you touch? *Exp Brain Res, 202*(4), 747-754. doi: 10.1007/s00221-010-2178-6

Kim, J. K., & Zatorre, R. J. (2011). Tactile-auditory shape learning engages the lateral occipital complex. *J Neurosci, 31*(21), 7848-7856. doi: 10.1523/JNEUROSCI.3399-10.2011

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416*(6877), 172-174.

Klauke, S., Goertz, M., Rein, S., Hoehl, D., Thomas, U., Eckhorn, R., . . . Wachtler, T. (2011). Stimulation with a wireless intraocular epiretinal implant elicits visual percepts in blind humans. *Invest Ophthalmol Vis Sci, 52*(1), 449-455. doi: 10.1167/iovs.09-4410

Klein, S. A., Casson, E., & Carney, T. (1990). Vernier acuity as line and dipole detection. *Vision Res, 30*(11), 1703-1719.

Klemen, J., Buchel, C., & Rose, M. (2009). Perceptual load interacts with stimulus processing across sensory modalities. *Eur J Neurosci, 29*(12), 2426-2434. doi: 10.1111/j.1460-9568.2009.06774.x

Knott, S. T., & Hersey, J. B. (1958). Interpretation of high-resolution echo-sounding techniques and their use in bathymetry, marine geophysics, and biology. *Deep Sea Research (1953), 4*, 36-44.

Kohler, W. (1947). Gestalt Psychology (1929). *Liveright, New York*.

Kollmeier, B., Brand, T., & Meyer, B. (2008). Perception of speech and sound, volume Springer: Berlin,. Springer Handbook of Speech Processing (Benesty, Sondhi and Huang Eds.).

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychon Bull Rev, 13*(2), 262-268.

Krieg, W. J. S. (1953). *Functional Neuroanatomy*: Blakiston Comp.

Kujala, T., Huotilainen, M., Sinkkonen, J., Ahonen, A. I., Alho, K., Hamalainen, M. S., . . . et al. (1995). Visual cortex activation in blind humans during sound discrimination. *Neurosci Lett, 183*(1-2), 143-146.

Kumar, S., Forster, H. M., Bailey, P., & Griffiths, T. D. (2008). Mapping unpleasantness of sounds to their auditory representation. *J Acoust Soc Am, 124*(6), 3810-3817. doi: 10.1121/1.3006380

Kupers, R., Chebat, D. R., Madsen, K. H., Paulson, O. B., & Ptito, M. (2010). Neural correlates of virtual route recognition in congenital blindness. *Proc Natl Acad Sci U S A, 107*(28), 12716-12721. doi: 10.1073/pnas.1006199107

Landry, S. P., Shiller, D. M., & Champoux, F. (2013). Short-term visual deprivation improves the perception of harmonicity. *J Exp Psychol Hum Percept Perform, 39*(6), 1503-1507. doi: 10.1037/a0034015

Lapid, E., Ulrich, R., & Rammsayer, T. (2009). Perceptual learning in auditory temporal discrimination: no evidence for a cross-modal transfer to the visual modality. *Psychon Bull Rev, 16*(2), 382-389. doi: 10.3758/PBR.16.2.382

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res, 158*(4), 405-414. doi: 10.1007/s00221-004-1913-2

Lauritzen, T. Z., Harris, J., Mohand-Said, S., Sahel, J. A., Dorn, J. D., McClure, K., & Greenberg, R. J. (2012). Reading visual braille with a retinal prosthesis. *Front Neurosci, 6*, 168. doi: 10.3389/fnins.2012.00168

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *J Exp Psychol Hum Percept Perform, 21*(3), 451-468.

Lavie, N. (2006). The role of perceptual load in visual awareness. *Brain Res, 1080*(1), 91-100. doi: 10.1016/j.brainres.2005.10.023

Lavie, N. (2011). 'Load theory' of attention. Nilli Lavie. *Curr Biol, 21*(17), R645-647.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Percept Psychophys, 56*(2), 183-197.

Leclerc, C., Saint-Amour, D., Lavoie, M. E., Lassonde, M., & Lepore, F. (2000). Brain functional reorganization in early blind humans revealed by auditory event-related potentials. *Neuroreport, 11*(3), 545-550.

Leclerc, C., Segalowitz, S. J., Desjardins, J., Lassonde, M., & Lepore, F. (2005). EEG coherence in early-blind humans during sound localization. *Neurosci Lett, 376*(3), 154-159. doi: 10.1016/j.neulet.2004.11.046

Lenay, C., Gapenne, O., Hanneton, S., Marque, C., & Genouelle, C. (2003). Sensory substitution: limits and perspectives *Touching for knowing, Cognitive psychology of haptic manual perception* (pp. 275-292): John Benjamins Publishing Company.

Levy-Tzedek, S., Hanassy, S., Abboud, S., Maidenbaum, S., & Amedi, A. (2012). Fast, accurate reaching movements with a visual-to-auditory sensory substitution device. *Restor Neurol Neurosci, 30*(4), 313-323. doi: 10.3233/RNN-2012-110219

Levy-Tzedek, S., Riemer, D., & Amedi, A. (2014). Color improves "visual" acuity via sound. *Front Neurosci, 8*, 358. doi: 10.3389/fnins.2014.00358

Lewald, J. (2007). More accurate sound localization induced by short-term light deprivation. *Neuropsychologia, 45*(6), 1215-1222. doi: 10.1016/j.neuropsychologia.2006.10.006

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory--visual integration. *Behav Brain Res, 121*(1-2), 69-79.

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex, 10*(9), 873-888.

Li, S., Hu, J., Chai, X., & Peng, Y. H. (2012). Image Recognition With a Limited Number of Pixels for Visual Prostheses Design. *Artificial Organs, 36*(3), 266-274. doi: DOI 10.1111/j.1525-1594.2011.01347.x

Li, S., Wang, K., Wang, D., & Akamatsu, T. (2005). Echolocation signals of the free-ranging Yangtze finless porpoise (Neophocaena phocaenoides asiaeorientialis). *J Acoust Soc Am, 117*(5), 3288-3296.

Lloyd, D., Morrison, I., & Roberts, N. (2006). Role for human posterior parietal cortex in visual processing of aversive objects in peripersonal space. *J Neurophysiol, 95*(1), 205-214. doi: 10.1152/jn.00614.2005

Lobo, L., Travieso, D., Barrientos, A., & Jacobs, D. M. (2014). Stepping on obstacles with a sensory substitution device on the lower leg: practice without vision is more beneficial than practice with vision. *PLoS One, 9*(6), e98801. doi: 10.1371/journal.pone.0098801

Loomis, J. M. (1974). Tactile letter recognition under different modes of stimulus presentation. *Perception & Psychophysics, 16*(2), 401-408.

Loomis, J. M. (1981a). On the tangibility of letters and braille. *Perception & Psychophysics, 29*(1), 37-46.

Loomis, J. M. (1981b). Tactile pattern perception. *Perception, 10*(1), 5-27.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci, 28*(5), 264-271. doi: 10.1016/j.tins.2005.03.008

Macdonald, J. S., & Lavie, N. (2008). Load induced blindness. *J Exp Psychol Hum Percept Perform, 34*(5), 1078-1091. doi: 10.1037/0096-1523.34.5.1078

Macdonald, J. S., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Atten Percept Psychophys, 73*(6), 1780-1789. doi: 10.3758/s13414-011-0144-4

Maidenbaum, S., Hannasi, S., Abboud, S., Arbel, R., Shipuznikov, A., Levy-Tzedek, S., . . . Amedi, A. (2012). The EyeCane - Distance information for the blind. *Journal of Molecular Neuroscience, 48*, S75-S76.

Maidenbaum, S., Levy-Tzedek, S., Chebat, D. R., & Amedi, A. (2013). Increasing Accessibility to the Blind of Virtual Environments, Using a Virtual Mobility Aid Based On the "EyeCane":

Marconi, B., Genovesio, A., Battaglia-Mayer, A., Ferraina, S., Squatrito, S., Molinari, M., . . . Caminiti, R. (2001). Eye-hand coordination during reaching. I. Anatomical relationships between parietal and frontal cortex. *Cereb Cortex, 11*(6), 513-527.

Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness. *Am J Psychol, 87*(1-2), 173-188.

Marks, L. E. (1987). On Cross-Modal Similarity - Auditory Visual Interactions in Speeded Discrimination. *Journal of Experimental Psychology-Human Perception and Performance, 13*(3), 384-394. doi: Doi 10.1037/0096-1523.13.3.384

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. Calvert, C. Spence & B. E. Stein (Eds.), *Handbook of Multisensory Processes.* (pp. 85-106). Cambridge, Massachusetts.: MIT Press.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nat Rev Neurosci, 5*(3), 229-240. doi: 10.1038/nrn1348

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: tests of the semantic coding hypothesis. *Perception, 28*(7), 903-923.

Martinovic, J., Gruber, T., Hantsch, A., & Muller, M. M. (2008). Induced gamma-band activity is related to the time point of object identification. *Brain Res, 1198*, 93-106. doi: 10.1016/j.brainres.2007.12.050

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychon Bull Rev, 18*(5), 904-909. doi: 10.3758/s13423-011-0131-8

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Dev Sci, 9*(3), 316-322. doi: 10.1111/j.1467-7687.2006.00495.x

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Dev Sci, 11*(1), 122-134. doi: 10.1111/j.1467-7687.2007.00653.x

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

McKee, S. P., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Percept Psychophys, 24*(3), 258-262.

Meijer, P. (1992). An Experimental System for Auditory Image Representations. *Ieee Transactions on Biomedical Engineering, 39*(2), 112-121. doi: Doi 10.1109/10.121642

Merabet, L. B., Battelli, L., Obretenova, S., Maguire, S., Meijer, P., & Pascual-Leone, A. (2009). Functional recruitment of visual cortex for sound encoded object identification in the blind. *Neuroreport, 20*(2), 132-138. doi: 10.1097/WNR.0b013e32832104dc

Merabet, L. B., Hamilton, R., Schlaug, G., Swisher, J. D., Kiriakopoulos, E. T., Pitskel, N. B., . . . Pascual-Leone, A. (2008). Rapid and reversible recruitment of early visual cortex for touch. *PLoS One, 3*(8), e3046. doi: 10.1371/journal.pone.0003046

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*(7397), 233-U118. doi: Doi 10.1038/Nature11020

Micheyl, C., Bernstein, J. G., & Oxenham, A. J. (2006). Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise. *J Acoust Soc Am, 120*(3), 1493-1505.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res, 219*(1-2), 36-47. doi: 10.1016/j.heares.2006.05.004

Miller, J. (1987). Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters. *Perception & Psychophysics, 41*(5), 419-434.

Milne, J. L., Arnott, S. R., Kish, D., Goodale, M. A., & Thaler, L. (2014). Parahippocampal cortex is involved in material processing via echoes in blind echolocation experts. *Vision Res*. doi: 10.1016/j.visres.2014.07.004

Milne, J. L., Goodale, M. A., & Thaler, L. (2014). The role of head movements in the discrimination of 2-D shape by blind echolocation experts. *Atten Percept Psychophys, 76*(6), 1828-1837. doi: 10.3758/s13414-014-0695-2

Morgan, M. J. (1977). *Molyneux's question: Vision, touch and the philosophy of perception*: Cambridge U Press.

Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., & Wright, B. A. (2006). Perceptual-learning evidence for separate processing of asynchrony and order tasks. *J Neurosci, 26*(49), 12708-12716. doi: 10.1523/JNEUROSCI.2254-06.2006

Mossbridge, J. A., Scissors, B. N., & Wright, B. A. (2008). Learning and generalization on asynchrony and order tasks at sound offset: implications for underlying neural circuitry. *Learn Mem, 15*(1), 13-20. doi: 10.1101/lm.573608

Murata, K., Cramer, H., & Bach-y-Rita, P. (1965). Neuronal convergence of noxious, acoustic, and visual stimuli in the visual cortex of the cat. *J Neurophysiol, 28*(6), 1223-1239.

Murphy, S., Fraenkel, N., & Dalton, P. (2013). Perceptual load does not modulate auditory distractor processing. *Cognition, 129*(2), 345-355. doi: 10.1016/j.cognition.2013.07.014

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nat Rev Neurosci, 9*(11), 856-869. doi: 10.1038/nrn2478

Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *J Neurosci, 18*(4), 1559-1570.

Nau, A. C., Pintar, C., Arnoldussen, A., & Fisher, C. (2015). Acquisition of Visual Perception in Blind Adults Using the BrainPort Artificial Vision Device. *American Journal of Occupational Therapy, 69*(1), 6901290010p6901290011-6901290010p6901290018.

Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Atten Percept Psychophys, 72*(6), 1654-1665. doi: 10.3758/APP.72.6.1654

Noorsal, E., Sooksood, K., Xu, H. C., Hornig, R., Becker, J., & Ortmanns, M. (2012). A Neural Stimulator Frontend With High-Voltage Compliance and Programmable Pulse Shape for Epiretinal Implants. *Ieee Journal of Solid-State Circuits, 47*(1), 244-256. doi: Doi 10.1109/Jssc.2011.2164667

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological review, 75*(6), 522.

Normann, R. A., Maynard, E. M., Rousche, P. J., & Warren, D. J. (1999). A neural interface for a cortical vision prosthesis. *Vision Res, 39*(15), 2577-2587.

O'Regan, J. K., Myin, E., & Noe, A. (2005). Skill, corporality and alerting capacity in an account of sensory consciousness. *Prog Brain Res, 150*, 55-68. doi: 10.1016/S0079-6123(05)50005-0

O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav Brain Sci, 24*(5), 939-973; discussion 973-1031.

Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Percept Psychophys, 65*(1), 123-132.

Ortega, L., Guzman-Martinez, E., Grabowecky, M., & Suzuki, S. (2014). Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Atten Percept Psychophys, 76*(5), 1485-1502. doi: 10.3758/s13414-014-0663-x

Ortiz, T., Poch, J., Santos, J. M., Requena, C., Martinez, A. M., OrtizTeran, L., . . . Pascual-Leone, A. (2011). Recruitment of occipital cortex during sensory substitution training linked to subjective experience of seeing in people with blindness. *PLoS One, 6*(8), e23264. doi: 10.1371/journal.pone.0023264

Overvliet, K. E., Smeets, J. B., & Brenner, E. (2007a). Haptic search with finger movements: using more fingers does not necessarily reduce search times. *Exp Brain Res, 182*(3), 427-434. doi: 10.1007/s00221-007-0998-9

Overvliet, K. E., Smeets, J. B., & Brenner, E. (2007b). Parallel and serial search in haptics. *Percept Psychophys, 69*(7), 1059-1069.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811-814. doi: 10.1038/33918

Pascolini, D., & Mariotti, S. P. (2012). Global estimates of visual impairment: 2010. *Br J Ophthalmol, 96*(5), 614-618. doi: 10.1136/bjophthalmol-2011-300539

Pascual-Leone, A., Cammarota, A., Wassermann, E. M., Brasil-Neto, J. P., Cohen, L. G., & Hallett, M. (1993). Modulation of motor cortical outputs to the reading hand of braille readers. *Ann Neurol, 34*(1), 33-37. doi: 10.1002/ana.410340108

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Prog Brain Res, 134*, 427-445.

Pascual-Leone, A., Hamilton, R., Tormos, J. M., Keenan, J. P., & Catala, M. D. (1999). Neuroplasticity in the adjustment to blindness *Neuronal plasticity: Building a bridge from the laboratory to the clinic* (pp. 93-108): Springer.

Pascual-Leone, A., & Torres, F. (1993). Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. *Brain, 116 (Pt 1)*, 39-52.

Pashler, H. (1988). Familiarity and visual change detection. *Percept Psychophys, 44*(4), 369-378.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol Bull, 116*(2), 220-244.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., . . . Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol, 10*(1), e1001251. doi: 10.1371/journal.pbio.1001251

Pasqualotto, A., Lam, J. S., & Proulx, M. J. (2013). Congenital blindness improves semantic and episodic memory. *Behav Brain Res, 244*, 162-165. doi: 10.1016/j.bbr.2013.02.005

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neurosci Biobehav Rev, 36*(4), 1179-1187. doi: 10.1016/j.neubiorev.2012.01.008

Pasqualotto, A., Spiller, M. J., Jansari, A. S., & Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behav Brain Res, 236*(1), 175-179. doi: 10.1016/j.bbr.2012.08.042

Pasqualotto, A., Taya, S., & Proulx, M. J. (2014). Sensory deprivation: visual experience alters the mental number line. *Behav Brain Res, 261*, 110-113. doi: 10.1016/j.bbr.2013.12.017

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H., Cohen, L., . . . Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proc Natl Acad Sci U S A, 101*(15), 5658-5663. doi: 10.1073/pnas.0400707101

Piyathaisere, D. V., Margalit, E., Chen, S. J., Shyu, J. S., D'Anna, S. A., Weiland, J. D., . . . Humayun, M. S. (2003). Heat effects on the retina. *Ophthalmic Surg Lasers Imaging, 34*(2), 114-120.

Planetta, P. J., & Servos, P. (2008). Somatosensory temporal discrimination learning generalizes to motor interval production. *Brain Res, 1233*, 51-57. doi: 10.1016/j.brainres.2008.07.081

Plaza, P., Cuevas, I., Grandin, C., De Volder, A. G., & Renier, L. (2012). Looking into Task-Specific Activation Using a Prosthesis Substituting Vision with Audition. *ISRN Rehabilitation, 2012*, 15. doi: 10.5402/2012/490950

Poirier, C., Collignon, O., Devolder, A. G., Renier, L., Vanlierde, A., Tranduy, D., & Scheiber, C. (2005). Specific activation of the V5 brain area by auditory motion processing: an fMRI study. *Brain Res Cogn Brain Res, 25*(3), 650-658. doi: 10.1016/j.cogbrainres.2005.08.015

Poirier, C., Collignon, O., Scheiber, C., Renier, L., Vanlierde, A., Tranduy, D., . . . De Volder, A. G. (2006). Auditory motion perception activates visual motion areas in early blind subjects. *Neuroimage, 31*(1), 279-285. doi: 10.1016/j.neuroimage.2005.11.036

Poirier, C., De Volder, A., Tranduy, D., & Scheiber, C. (2007). Pattern recognition using a device substituting audition for vision in blindfolded sighted subjects. *Neuropsychologia, 45*(5), 1108-1121. doi: DOI 10.1016/j.neuropsychologia.2006.09.018

Poirier, C., De Volder, A. G., Tranduy, D., & Scheiber, C. (2006). Neural changes in the ventral and dorsal visual streams during pattern recognition learning. *Neurobiol Learn Mem, 85*(1), 36-43. doi: 10.1016/j.nlm.2005.08.006

Prensky, M. (2005). Computer games and learning: Digital game-based learning. *Handbook of computer game studies, 18*, 97-122.

Prokofyeva, E., & Zrenner, E. (2012). Epidemiology of major eye diseases leading to blindness in Europe: a literature review. *Ophthalmic Res, 47*(4), 171-188. doi: 10.1159/000329603

Proulx, M. J., Brown, D. J., Pasqualotto, A., & Meijer, P. (2014). Multisensory perceptual learning and sensory substitution. *Neurosci Biobehav Rev, 41*, 16-25. doi: 10.1016/j.neubiorev.2012.11.017

Proulx, M. J., Stoerig, P., Ludowig, E., & Knoll, I. (2008). Seeing 'Where' through the Ears: Effects of Learning-by-Doing and Long-Term Sensory Deprivation on Localization Based on Image-to-Sound Substitution. *PLoS One, 3*(3). doi: Artn E1840

Ptito, M., Moesgaard, S. M., Gjedde, A., & Kupers, R. (2005). Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind. *Brain, 128*(Pt 3), 606-614. doi: 10.1093/brain/awh380

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited. *Psychol Res, 49*(1), 17-22.

Ramachandran, V. S., & Hubbard, E. M. (2003). Hearing colors, tasting shapes. *Sci Am, 288*(5), 52-59.

Rammsayer, T. H., & Leutner, D. (1996). Temporal discrimination as a function of marker duration. *Percept Psychophys, 58*(8), 1213-1223.

Rauschecker, J. P. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci, 18*(1), 36-43.

Rauschecker, J. P. (2008). Plasticity of cortical maps in visual deprivation. *Blindness and brain plasticity in navigation and object perception. Taylor & Francis, New York*, 43-66.

Rauschecker, J. P., Tian, B., Korte, M., & Egert, U. (1992). Crossmodal changes in the somatosensory vibrissa/barrel system of visually deprived animals. *Proc Natl Acad Sci U S A, 89*(11), 5063-5067.

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Atten Percept Psychophys, 77*(2), 483-492. doi: 10.3758/s13414-014-0776-2

Re, D. E., O'Connor, J. J. M., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PLoS One, 7*(3), e32719.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *J Neurophysiol, 89*(2), 1078-1093. doi: 10.1152/jn.00706.2002

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci, 13*(1), 87-103.

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Curr Biol, 21*(5), 363-368. doi: 10.1016/j.cub.2011.01.040

Renier, L., Bruyer, R., & De Volder, A. G. (2006). Vertical-horizontal illusion present for sighted but not early blind humans using auditory substitution of vision. *Percept Psychophys, 68*(4), 535-542.

Renier, L., Collignon, O., Poirier, C., Tranduy, D., Vanlierde, A., Bol, A., . . . De Volder, A. G. (2005). Cross-modal activation of visual cortex during depth perception using auditory substitution of vision. *Neuroimage, 26*(2), 573-580. doi: 10.1016/j.neuroimage.2005.01.047

Renier, L., & De Volder, A. G. (2010). Vision substitution and depth perception: early blind subjects experience visual perspective through their ears. *Disabil Rehabil Assist Technol, 5*(3), 175-183. doi: 10.3109/17483100903253936

Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., & De Volder, A. G. (2005). The Ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception, 34*(7), 857-867.

Reuschel, J., Rosler, F., Henriques, D. Y. P., & Fiehler, K. (2012). Spatial Updating Depends on Gaze Direction Even after Loss of Vision. *Journal of Neuroscience, 32*(7), 2422-2429. doi: Doi 10.1523/Jneurosci.2714-11.2012

Ricciardi, E., & Pietrini, P. (2011). New light from the dark: what blindness can teach us about brain function. *Curr Opin Neurol, 24*(4), 357-363. doi: 10.1097/WCO.0b013e328348bdbf

Ricciardi, E., Vanello, N., Sani, L., Gentili, C., Scilingo, E. P., Landini, L., . . . Pietrini, P. (2007). The effect of visual experience on the development of functional architecture in hMT+. *Cereb Cortex, 17*(12), 2933-2939. doi: 10.1093/cercor/bhm018

Rizzo, J. F., 3rd, Wyatt, J., Loewenstein, J., Kelly, S., & Shire, D. (2003a). Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays. *Invest Ophthalmol Vis Sci, 44*(12), 5355-5361.

Rizzo, J. F., 3rd, Wyatt, J., Loewenstein, J., Kelly, S., & Shire, D. (2003b). Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials. *Invest Ophthalmol Vis Sci, 44*(12), 5362-5369.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Int J Psychophysiol, 50*(1-2), 19-26.

Roder, B., Teder-Salejarvi, W., Sterr, A., Rosler, F., Hillyard, S. A., & Neville, H. J. (1999). Improved auditory spatial tuning in blind humans. *Nature, 400*(6740), 162-166. doi: 10.1038/22106

Rojas, J. A. M., Hermosilla, J. A., Montero, R. S., & Espí, P. L. L. (2009). Physical Analysis of Several Organic Signals for Human Echolocation: Oral Vacuum Pulses. *Acta Acustica united with Acustica, 95*(2), 325-330. doi: 10.3813/AAA.918155

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory Stimulus Timing Influences Perceived duration of Co-Occurring Visual Stimuli. *Front Psychol, 2*, 215. doi: 10.3389/fpsyg.2011.00215

Rush, A. D., & Troyk, P. R. (2012). A power and data link for a wireless-implanted neural recording system. *IEEE Trans Biomed Eng, 59*(11), 3255-3262. doi: 10.1109/TBME.2012.2214385

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: what is learned? *Vision Res, 35*(4), 519-527.

Sadato, N., Okada, T., Honda, M., & Yonekura, Y. (2002). Critical period for cross-modal plasticity in blind humans: a functional MRI study. *Neuroimage, 16*(2), 389-400. doi: 10.1006/nimg.2002.1111

Sadato, N., Pascual-Leone, A., Grafman, J., Deiber, M. P., Ibanez, V., & Hallett, M. (1998). Neural networks for Braille reading by the blind. *Brain, 121 (Pt 7)*, 1213-1229.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380*(6574), 526-528. doi: 10.1038/380526a0

Saenz, M., Lewis, L. B., Huth, A. G., Fine, I., & Koch, C. (2008). Visual Motion Area MT+/V5 Responds to Auditory Motion in Human Sight-Recovery Subjects. *J Neurosci, 28*(20), 5141-5148. doi: 10.1523/JNEUROSCI.0803-08.2008

Sampaio, E., Maris, S., & Bach-y-Rita, P. (2001). Brain plasticity: 'visual' acuity of blind persons via the tongue. *Brain Res, 908*(2), 204-207.

Santangelo, V., Olivetti Belardinelli, M., & Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *J Exp Psychol Hum Percept Perform, 33*(1), 137-148. doi: 10.1037/0096-1523.33.1.137

Scharff, A., Palmer, J., & Moore, C. M. (2013). Divided attention limits perception of 3-D object shapes. *J Vis, 13*(2), 18. doi: 10.1167/13.2.18

Schmidt, E. M., Bak, M. J., Hambrecht, F. T., Kufta, C. V., O'Rourke, D. K., & Vallabhanath, P. (1996). Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. *Brain, 119 (Pt 2)*, 507-522.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *J Physiol, 483 (Pt 3)*, 797-810.

Seitz, A. R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Curr Biol, 16*(14), 1422-1427. doi: 10.1016/j.cub.2006.05.048

Seitz, A. R., & Watanabe, T. (2009). The phenomenon of task-irrelevant perceptual learning. *Vision Res, 49*(21), 2604-2610. doi: 10.1016/j.visres.2009.08.003

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends Neurosci, 34*(3), 114-123. doi: 10.1016/j.tins.2010.11.002

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408*(6814), 788. doi: 10.1038/35048669

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends Cogn Sci, 12*(11), 411-417. doi: 10.1016/j.tics.2008.07.006

Sharpee, T. O., Atencio, C. A., & Schreiner, C. E. (2011). Hierarchical representations in the auditory cortex. *Current opinion in neurobiology, 21*(5), 761-767.

Shen, D., & Mondor, T. A. (2006). Effect of distractor sounds on the auditory attentional blink. *Percept Psychophys, 68*(2), 228-243.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *J Exp Psychol, 93*(1), 72-82.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends Cogn Sci, 12*(5), 182-186. doi: 10.1016/j.tics.2008.02.003

Shipley, T. (1964). Auditory Flutter-Driving of Visual Flicker. *Science, 145*(3638), 1328-1330.

Shower, E., & Biddulph, R. (1931). Differential pitch sensitivity of the ear. *J Acoust Soc Am, 3*(1A), 7-7.

Simmons, J. A., & Stein, R. A. (1980). Acoustic imaging in bat sonar: echolocation signals and the evolution of echolocation. *Journal of Comparative Physiology, 135*(1), 61-84.

Simon, J. Z., & Ding, N. (2010). Magnetoencephalography and Auditory Neural Representations. *26th Southern Biomedical Engineering Conference: Sbec 2010, 32*, 45-48.

Simpson, A. J., & Reiss, J. D. (2013). The dynamic range paradox: a central auditory model of intensity change detection. *PLoS One, 8*(2), e57497. doi: 10.1371/journal.pone.0057497

Simpson, A. J., Reiss, J. D., & McAlpine, D. (2013). Tuning of human modulation filters is carrier-frequency dependent. *PLoS One, 8*(8), e73590. doi: 10.1371/journal.pone.0073590

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex, 17*(10), 2387-2399. doi: 10.1093/cercor/bhl147

Snyder, C. R. (1972). Selection, inspection, and naming in visual search. *J Exp Psychol, 92*(3), 428.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Brain Res Cogn Brain Res, 14*(1), 139-146.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004a). Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture. *Cogn Affect Behav Neurosci, 4*(2), 208-217.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004b). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *J Exp Psychol Hum Percept Perform, 30*(2), 330-345. doi: 10.1037/0096-1523.30.2.330

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten Percept Psychophys, 73*(4), 971-995. doi: 10.3758/s13414-010-0073-7

Srivastava, N. R., Troyk, P. R., & Dagnelie, G. (2009). Detection, eye-hand coordination and virtual mobility performance in simulated vision for a cortical visual prosthesis device. *J Neural Eng, 6*(3), 035008. doi: 10.1088/1741-2560/6/3/035008

Srivastava, N. R., Troyk, P. R., Towle, V. L., Curry, D., Schmidt, E., Kufta, C., & Dagnelie, G. (2007). Estimating phosphene maps for psychophysical experiments used in testing a cortical visual prosthesis device. *2007 3rd International Ieee/Embs Conference on Neural Engineering, Vols 1 and 2*, 130-133. doi: Doi 10.1109/Cne.2007.369629

Stevens, J. C., Foulke, E., & Patterson, M. Q. (1996). Tactile acuity,aging, and braille reading in long term blindness. *J Exp Psychol, 2*(2), 91-106. doi: 10.1037/1076-898X.2.2.91

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proc Natl Acad Sci U S A, 54*(2), 407-411.

Stiles, N., Chib, V., & Shimojo, S. (2012). Behavioral and fMRI Measures of "Visual" Processing with a Sensory Substitution Device. *Journal of Vision, 12*(9), 703. doi: 10.1167/12.9.703

Stingl, K., Bach, M., Bartz-Schmidt, K. U., Braun, A., Bruckmann, A., Gekeler, F., . . . Zrenner, E. (2013). Safety and efficacy of subretinal visual implants in humans: methodological aspects. *Clin Exp Optom, 96*(1), 4-13. doi: 10.1111/j.1444-0938.2012.00816.x

Stingl, K., Greppmaier, U., Wilhelm, B., & Zrenner, E. (2010). [Subretinal visual implants]. *Klin Monbl Augenheilkd, 227*(12), 940-945. doi: 10.1055/s-0029-1245830

Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. *Curr Biol, 24*(6), 687-692. doi: 10.1016/j.cub.2014.02.010

Striem-Amit, E., Cohen, L., Dehaene, S., & Amedi, A. (2012). Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron, 76*(3), 640-652. doi: 10.1016/j.neuron.2012.08.026

Striem-Amit, E., Dakwar, O., Reich, L., & Amedi, A. (2012). The large-scale organization of "visual" streams emerges without visual experience. *Cereb Cortex, 22*(7), 1698-1709. doi: 10.1093/cercor/bhr253

Striem-Amit, E., Guendelman, M., & Amedi, A. (2012). 'Visual' Acuity of the Congenitally Blind Using Visual-to-Auditory Sensory Substitution. *PLoS One, 7*(3). doi: ARTN e33136

Stromeyer, C. F., 3rd., & Julesz, B. (1972). Spatial-frequency masking in vision: critical bands and spread of masking. *J Opt Soc Am, 62*(10), 1221-1232.

Sur, M., Pallas, S. L., & Roe, A. W. (1990). Cross-modal plasticity in cortical development: differentiation and specification of sensory neocortex. *Trends Neurosci, 13*(6), 227-233.

Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *Elife, 2*, e00699. doi: 10.7554/eLife.00699

Terhardt, E. (1974). Pitch, consonance, and harmony. *J Acoust Soc Am, 55*(5), 1061-1069.

Thaler, L., Arnott, S. R., & Goodale, M. A. (2011). Neural Correlates of Natural Human Echolocation in Early and Late Blind Echolocation Experts. *PLoS One, 6*(5). doi: ARTN e20162

Thaler, L., Milne, J. L., Arnott, S. R., Kish, D., & Goodale, M. A. (2014). Neural correlates of motion processing through echolocation, source hearing, and vision in blind echolocation experts and sighted echolocation novices. *J Neurophysiol, 111*(1), 112-127. doi: 10.1152/jn.00501.2013

Thaler, L., Wilson, R. C., & Gee, B. K. (2014). Correlation between vividness of visual imagery and echolocation ability in sighted, echo-na < ve people. *Exp Brain Res, 232*(6), 1915-1925. doi: DOI 10.1007/s00221-014-3883-3

Treisman, A., & Geffen, G. (1967). Selective attention: perception or response? *Q J Exp Psychol, 19*(1), 1-17. doi: 10.1080/14640746708400062

Treisman, A., & Riley, J. G. (1969). Is selective attention selective perception or selective response? A further test. *J Exp Psychol, 79*(1), 27-34.

Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Percept Psychophys, 67*(2), 195-208.

Troyk, P., Bak, M., Berg, J., Bradley, D., Cogan, S., Erickson, R., . . . Towle, V. (2003). A model for intracortical visual prosthesis research. *Artif Organs, 27*(11), 1005-1015.

Tyler, M., Danilov, Y., & Bach, Y. R. P. (2003). Closing an open-loop control system: vestibular substitution through the tongue. *J Integr Neurosci, 2*(2), 159-164.

Uneri, A., & Polat, S. (2009). Vestibular rehabilitation with electrotactile vestibular substitution: early effects. *Eur Arch Otorhinolaryngol, 266*(8), 1199-1203. doi: 10.1007/s00405-008-0886-3

Vaidya, A., Borgonovi, E., Taylor, R. S., Sahel, J. A., Rizzo, S., Stanga, P. E., . . . Walter, P. (2014). The cost-effectiveness of the Argus II retinal prosthesis in Retinitis Pigmentosa patients. *BMC Ophthalmol, 14*, 49. doi: 10.1186/1471-2415-14-49

Van Boven, R. W., Hamilton, R. H., Kauffman, T., Keenan, I. P., & Pascual-Leone, A. (2000). Tactile spatial resolution in blind braille readers. *Neurology, 54*, 2230-2236. doi: 10.1212/WNL.54.12.2230

Van der Burg, E., Awh, E., & Olivers, C. N. (2013). The capacity of audiovisual integration is limited to one item. *Psychol Sci, 24*(3), 345-351. doi: 10.1177/0956797612452865

Van der Burg, E., Cass, J., Olivers, C. N., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One, 5*(5), e10664. doi: 10.1371/journal.pone.0010664

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A, 102*(4), 1181-1186. doi: 10.1073/pnas.0408949102

Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions. *Brain Res, 1111*(1), 134-142.

Veraart, C., Raftopoulos, C., Mortimer, J. T., Delbeke, J., Pins, D., Michaux, G., . . . Wanet-Defalque, M. C. (1998). Visual sensations produced by optic nerve stimulation using an implanted self-sizing spiral cuff electrode. *Brain Res, 813*(1), 181-186.

Veraart, C., Wanet-Defalque, M.-C., Gérard, B., Vanlierde, A., & Delbeke, J. (2003). Pattern Recognition with the Optic Nerve Visual Prosthesis. *Artificial Organs, 27*(11), 996-1004. doi: 10.1046/j.1525-1594.2003.07305.x

Vincent, M., Tang, H., Zhu, Z., & Ro, T. (2014). Discrimination of Shapes and Line Orientations on the Tongue. *Journal of Vision, 14*(10), 1094-1094.

von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature, 404*(6780), 871-876. doi: 10.1038/35009102

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J. P., & Lepore, F. (2004). Early- and late-onset blind individuals show supra-normal auditory abilities in far-space. *Curr Biol, 14*(19), 1734-1738. doi: 10.1016/j.cub.2004.09.051

Vroomen, J., & de Gelder, B. (2003). Visual motion influences the contingent auditory motion aftereffect. *Psychol Sci, 14*(4), 357-361.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychol Sci, 21*(1), 21-25. doi: 10.1177/0956797609354734

Wallace, M. T., Perrault, T. J., Hairston, W. D., & Stein, B. E. (2004). Visual experience is necessary for the development of multisensory integration. *Journal of Neuroscience, 24*(43), 9580-9584. doi: Doi 10.1523/Jneurosci.2535-04.2004

Wanet-Defalque, M. C., Veraart, C., De Volder, A., Metz, R., Michel, C., Dooms, G., & Goffinet, A. (1988). High metabolic activity in the visual cortex of early blind human subjects. *Brain Res, 446*(2), 369-373.

Ward, J., & Meijer, P. (2010a). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition, 19*(1), 492-500. doi: DOI 10.1016/j.concog.2009.10.006

Ward, J., & Meijer, P. (2010b). Visual experiences in the blind induced by an auditory sensory substitution device. *Conscious Cogn, 19*(1), 492-500. doi: 10.1016/j.concog.2009.10.006

Watt, R. J. (1984). Towards a general theory of the visual acuities for shape and spatial arrangement. *Vision Res, 24*(10), 1377-1386.

Waugh, S. J., & Levi, D. M. (1993a). Visibility, luminance and vernier acuity. *Vision Res, 33*(4), 527-538.

Waugh, S. J., & Levi, D. M. (1993b). Visibility, timing and vernier acuity. *Vision Res, 33*(4), 505-526.

Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L. G., . . . Rauschecker, J. P. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *J Neurosci, 20*(7), 2664-2672.

Weiland, J. D., & Humayun, M. S. (2005). A biomimetic retinal stimulating array. *IEEE Eng Med Biol Mag, 24*(5), 14-21.

Weiland, J. D., Liu, W., & Humayun, M. S. (2005). Retinal prosthesis. *Annu Rev Biomed Eng, 7*, 361-401. doi: 10.1146/annurev.bioeng.7.060804.100435

Weiss, T., Miltner, W. H., Huonker, R., Friedel, R., Schmidt, I., & Taub, E. (2000). Rapid functional plasticity of the somatosensory cortex after finger amputation. *Exp Brain Res, 134*(2), 199-203.

Weiss, T., Miltner, W. H., Liepert, J., Meissner, W., & Taub, E. (2004). Rapid functional plasticity in the primary somatomotor cortex and perceptual changes after nerve block. *Eur J Neurosci, 20*(12), 3413-3423. doi: 10.1111/j.1460-9568.2004.03790.x

Weissman, D. H., Warner, L. M., & Woldorff, M. G. (2004). The neural mechanisms for minimizing cross-modal distraction. *J Neurosci, 24*(48), 10941-10949. doi: 10.1523/JNEUROSCI.3669-04.2004

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Res, 15*, 1137-1141.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Res, 17*(8), 941-947.

Wever, E., & Wedell, C. (1941). Pitch discrimination at high frequencies. *Psychol Bull, 38*, 727.

Williams, M. D., Ray, C. T., Griffith, J., & De l'Aune, W. (2011). The use of a tactile-vision sensory substitution system as an augmentative tool for individuals with visual impairments. *J. Vis Impairment Blindness, 105*(1), 45-50.

Wong, M., Gnanakumaran, V., & Goldreich, D. (2011). Tactile spatial acuity enhancement in blindness: evidence for experience-dependent mechanisms. *J Neurosci, 31*(19), 7028-7037. doi: 10.1523/JNEUROSCI.6461-10.2011

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci, 10*(4), 420-422. doi: Doi 10.1038/Nn1872

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *J Neurosci, 17*(10), 3956-3963.

Wright, B. A., & Fitzgerald, M. B. (2005). Learning and generalization on five basic auditory discrimination tasks as assessed by threshold changes. *Auditory Signal Processingp: Physiology, Psychoacoustics, and Models*, 510-516.

Wright, B. A., & Sabin, A. T. (2007). Perceptual learning: how much daily training is enough? *Exp Brain Res, 180*(4), 727-736. doi: DOI 10.1007/s00221-007-0898-z

Wright, B. A., Wilson, R. M., & Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task. *J Neurosci, 30*(35), 11635-11639. doi: 10.1523/JNEUROSCI.1441-10.2010

Wright, R. D. (1994). Shifts of visual attention to multiple simultaneous location cues. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 48*(2), 205.

Wright, T., & Ward, J. (2013). The evolution of a visual-to-auditory sensory substitution device using interactive genetic algorithms. *Q J Exp Psychol (Hove), 66*(8), 1620-1638. doi: 10.1080/17470218.2012.754911

Xiang, J. J., Poeppel, D., & Simon, J. Z. (2013). Physiological evidence for auditory modulation filterbanks: Cortical responses to concurrent modulations. *Journal of the Acoustical Society of America, 133*(1), El7-El12. doi: Doi 10.1121/1.4769400

Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *J Exp Psychol Hum Percept Perform, 16*(4), 812-825.

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception, 31*(8), 995-1003.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is 'where' in the human auditory cortex? *Nat Neurosci, 5*(9), 905-909. doi: 10.1038/nn904

Zrenner, E. (2002a). The subretinal implant: can microphotodiode arrays replace degenerated retinal photoreceptors to restore vision? *Ophthalmologica, 216 Suppl 1*, 8-20; discussion 52-23. doi: 64650

Zrenner, E. (2002b). Will retinal implants restore vision? *Science, 295*(5557), 1022-1025. doi: 10.1126/science.1067996

Zrenner, E., Bartz-Schmidt, K. U., Benav, H., Besch, D., Bruckmann, A., Gabel, V. P., . . . Wilke, R. (2011). Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proc Biol Sci, 278*(1711), 1489-1497. doi: 10.1098/rspb.2010.1747