Improving hearing ability in challenging conditions

By

Liping Zhang

A thesis submitted in partial fulfilment of the requirements for the degree

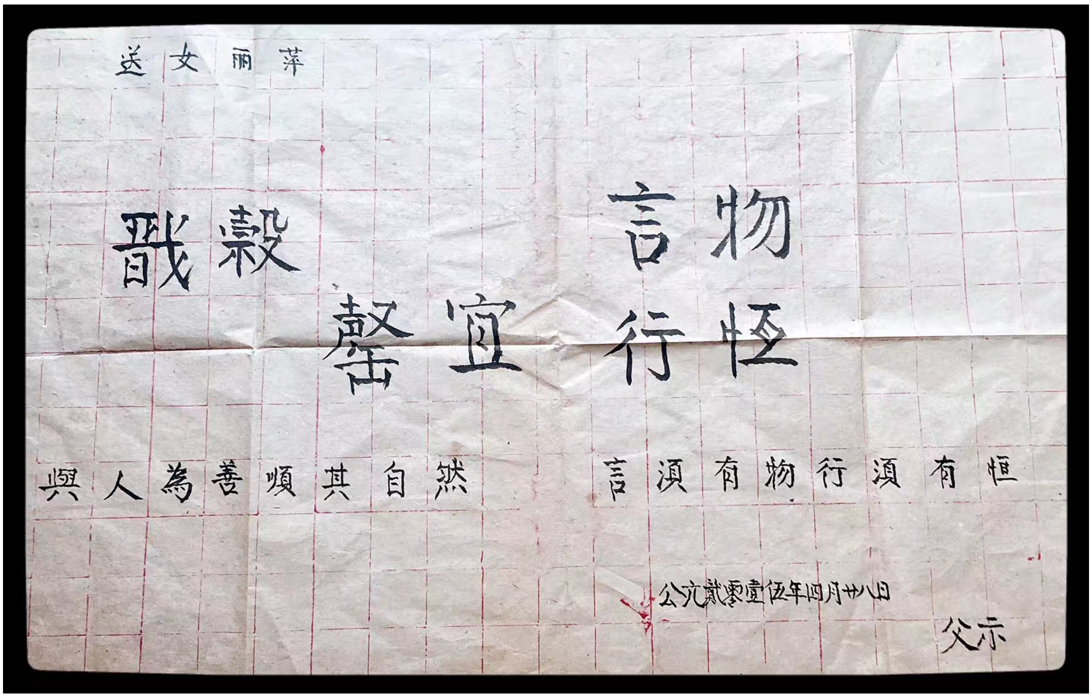
of

Doctor of Philosophy in Engineering

Warwick Manufacturing Group, The University of Warwick

June 2019

To all the up and down periods in this journey



送女雨萍

戩穀　　言物

罄宜　行恒

與人為善順其自然　言須有物行須有恒

公元貳零壹伍年四月廿八日

父示

Table of contents

List of figures

List of tables

Acknowledgements

Declaration

The author declares that this thesis and the work presented in it are my own work and has been generated by me as the result of my own original research.

All the research was undertaken independently at WMG, University of Warwick. All preparation, conduct, analysis and interpretation of the work was carried out by the author.

This thesis has not been submitted for a degree at any other university. This thesis is presented in accordance with the regulations of the University of Warwick.

Abstract

Although speech recognition using hearing aids and cochlear implants has improved significantly recently, most people with hearing impairment still have difficulty understanding speech in noisy environments. Improving the ability of the brain to learn how to make full use of prosthetic devices is as important as developments in the technology. Auditory perceptual training helps people to be more sensitive to target sounds. Therefore, auditory training programmes have the potential to optimise the performance of hearing-impaired users and help them get more benefit from their prosthetic devices. Better understanding of how and when auditory perceptual training generalises with normal hearing people may help in devising better training for people with hearing impairment. However, in literature, researchers have mainly focused on changing the target stimuli using amplitude modulated sounds or speech stimuli. Fewer researchers have explored the auditory learning and generalization effect of changing the background noise. It is not clear whether training generalizes to other types of noise, and in particular real-world environmental noises.

A novel element of this study is that it focuses on auditory training of people to pick up the target stimuli by changing the background noise. This project was divided into four stages. The first stage of this work looked at basic detection thresholds for amplitude modulation (AM) in sound stimuli, and found that training with AM-detection did not generalize to AM-rate discrimination, regardless of the modulation depths. For the second stage, two nonsense stimuli (Vowel Consonant Vowel VCV) training studies were carried out to explore auditory perceptual learning patterns with nonsense syllables across fixed and random background noise. It was motivated by visual research which showed that people can improve their detection performance by learning to ignore constant visual noise and that this skill transfers to new, random visual noise. Results showed that learning with random noise produced better identification performance than with fixed noise. There was no generalization from fixed noise training to random noise environments. These results were in contrast to the visual learning studies. Followed by the second stage, a short single session VCV study was conducted to investigate whether nonsense syllable adaption to fixed noise was different to random noise. Results showed that listeners' VCV identification was similar for fixed and random babble noise conditions. This was different from stage two that showed better nonsense recognition with random noise training than with fixed noise training. It is suggested that test method differences (multi-sessions vs single session) lead to performance differences between fixed and random noise conditions. The final stage of this work was to explore whether any learning effect from training with speech in random babble noise generalized to other environmental noises, such as car and rain. Results demonstrated that speech in babble noise training generalized to car and rain noise conditions, and part of the learning effect from speech in babble noise was sustained after several weeks.

This project investigated auditory perceptual learning performance of normal hearing people using AM stimuli, nonsense speech, and speech with various types of background noise (babble, car, rain). The learning outcomes provide important evidence for the use of background noises (fixed noise, random noise, and real-world environmental noises) in auditory perceptual training programmes, which can help to build up clinical guidelines for training people with hearing impairment to improve their hearing in challenging conditions.

List of publications

Parts of this thesis have been published or in the process by the author. All of them are listed below:

Zhang. L., Schlaghecken, F., Jennings, P., Harte, J. Robert, K. Generalization resulting from training of speech from babble noise to other background noises, submitted to ***Journal of speech, language and hearing research.*** *Corrections under review.*

Zhang. L., Schlaghecken, F., Robert, K., Harte, J. Auditory training of Vowel Consonant Vowel (VCV) recognition with fixed and random babble noise, supposed to submit to ***Hearing research***

Zhang. L., Schlaghecken, F., Harte, J., Jennings, P. (2016). Perceptual learning to ignore different background noises in speech test, ***the proceeding of the 2016 BSA Annual Conference***, UK, p 25-27, April 2016

Zhang. L., Schlaghecken, F., Harte, J., Jennings, P. (2016). Training normal hearing people in challenging conditions with no sense syllable stimuli, ***the Proceeding of WIN conference***, UK, p 45-46

Zhang. L., Schlaghecken, F., Harte, J., Jennings, P. (2015). Learning to ignore random babble noise with VCV tasks, ***the 3rd International Conference on Cognitive Hearing Science for Communication***, Sweden, 13-19 June 2015

Zhang. L., Schlaghecken, F., Harte, J., Jennings, P. (2014). Learning to ignore background noise in VCV test, ***the proceeding of the 2014 BSA Annual Conference***, UK, p29-30.

Zhang. L., Schlaghecken, F., Jennings, P. Harte, J. (2014). learning to ignore the interferer in frozen babble noise using VCV tasks, ***the proceeding of the 8th IEEE PG-Biomed Conference***, UK, p17-18.

Zhang, L. Schlaghecken, F Harte, J. (2013). Generalization resulting from training on a SAM detection task to a SAM-rate discrimination task with different depths, ***the 4th International Symposium on Audiology and Audiological Research (ISAAR)***, Denmark, p 61-69. **ISBN 978-87-990013-4-7**

Abbreviations

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
APD	Auditory Process Disorder
BKB	Bamford-Kowal-Bench
BM	Basilar Membrane
CI	Cochlear Implant
dB	Decibel
EEG	Electroencephalography
fMRI	functional Magnetic Resonance Imaging
HA	Hearing aid
HL	Hearing Level
Hz	Herz
IHC	Inner Hair Cell
IID	Interaural Intensity Difference
ILD	Interaural Level Difference
ITD	Interaural Time Difference
ms	Millisecond
OHC	Outer Hair Cell
SAM	Sinusoidal Amplitude Modulation
SNR	Signal to Noise Ratio
SPL	Sound Pressure Level
VCV	Vowel Consonant Vowel

Chapter 1 Thesis overview

1.1 Introduction

Hearing aids (HA) and Cochlear implants (CI) are widely used prosthetic devices to improve the auditory function of people with hearing impairments. With these modern devices, the ability of hearing-impaired people to understand speech in quiet conditions almost approaches that of the normal hearing community. Speech is one of the reliable tools for communication, even when it is degraded or masked by other competing sounds in daily life. A normal hearing person can make use of the context, rhythm, stress, and intonation of speech to understand another speaker. However, it is difficult for hearing-impaired people to make use of these cues. Although speech recognition by users of HAs and CIs has improved significantly over the past several years, the ability of most hearing-impaired people to understand speech in noisy environments is still quite poor (Dorman & Wilson, 2004; Ricketts & Hornsby, 2005).

The brain can process distorted and impoverished input sounds once the sound pattern-recognition system has been created. However, if complex sound patterns have not been learned during the normal language acquisition period, or if sound inputs are severely distorted, the central auditory system has to learn to process a new array of sound inputs. Hearing assistive devices only work well for hearing-impaired people who had hearing ability before their deafness, or received hearing assistance at a very young age. Rehabilitation and auditory-training programmes have the potential to improve the abilities of hearing-impaired people to make use of

HAs or CIs, and help them to obtain more benefit from their prosthetic devices. Therefore, the ability of the brain to learn how to make use of an assistive device is as important as technological development (Plomp, 1978; Moore & Shannon, 2009). This thesis focused on exploring how to improve human perceptual learning in noisy environments.

Auditory learning is defined as an improvement in the ability to detect, discriminate, or group sounds and speech information (Goldstone, 1998; Halliday *et al.*, 2012). Training of the auditory system may lead to long-lasting changes in a person's perceptual system in order to improve their ability to recognise sounds in their surrounding environment. There are two aspects of auditory training: one is the learning effect, where a listener's ability to perform an auditory task could be improved through practice of the same task or stimulus; the other is the generalization effect, where training in one task or stimulus leads to improvement in performance of another.

Perceptual learning studies are not new in hearing research. However, in clinical trials most speech rehabilitation programmes have attempted to train HA or CI users to understand speech material better in a quiet environment. Typically, auditory perceptual learning rehabilitation programmes present speech with no background noise or other competing environmental factors. Auditory training in a quiet environment may help listeners to detect subtle listening cues, but it is not certain whether these auditory cues can be made use of in noisy environments (Fu & Nogaki, 2005).

A novel element of this thesis is that it explores how people can learn to recognise target information by changing background noise. This approach is motivated by the work of Schubö, Schlaghecken, and Meinecke (2001) in vision research. From perceptual learning in the visual domain, it was found that people can improve detection performance by learning to ignore visual masks. Research in this thesis uses this approach to learn how to develop new training methods to help improve auditory speech perception in challenging conditions.

Speech and sound information are mainly carried by amplitude and frequency variations over time by the auditory system. Hearing-impaired people have a reduced ability to detect these cues, particularly in challenging auditory environments. The hearing dynamic range for hearing-impaired people is much narrower than it for normal hearing people. But it is likely that perceptual learning can help to improve hearing ability for both NH and HI people (Halliday *et al.*, 2008). Better understanding of how and when auditory perceptual training generalizes normal hearing people may help devise better training for people with hearing impairment. This project focuses on exploring methods of auditory training in an attempt to improve the performance of normal hearing subjects; that is, to improve their understanding of speech in noisy environments, and to provide evidence of auditory training for future clinical use. The first stage of this accesses detection thresholds for amplitude modulation (AM) in sound stimuli, and whether training of a simple detection task can be generalized to other stimulus conditions or across psychophysical tasks – i.e., from an AM detection task to an AM rate-discrimination task. Later stages of the work focus on how to train people for more complex speech sounds (such as vowels, consonants and words) within various background environmental noises (i.e. car noise, babble noise, rain noise).

1.2 Thesis structure

- Chapter 2. Literature review: this chapter provides an overview of normal hearing and hearing loss, and describes methods to improve hearing. It also includes an introduction to perceptual learning. Both non-speech and speech perceptual learning studies, and types of perceptual learning are critically reviewed. Finally, this chapter identifies where knowledge gaps exist and what gaps this thesis aimed to accomplish.

- Chapter 3. Research Methodology: this chapter explains the research approach taken, including: test procedure considerations, feedback considerations, experimental preparation and participant selection.

- Chapter 4. No generalization from training on a SAM detection task to a SAM-rate discrimination task with different depths: Practice can improve the detection threshold for AM in sound stimuli. A recent study (Fitzgerald & Wright, 2011) also demonstrated that AM detection learning generalizes from trained to untrained AM rates, but not to a new task (rate discrimination). This experiment investigated whether the lack of generalisation found by Fitzgerald and Wright (2011) was due to the use of 100% AM depth in the rate discrimination task, and aimed to investigate if it is possible to improve the generalization of AM detection rate discrimination by using lower AM depths, such as 70% and 40%, in the discrimination task. The study did not show a generalization effect from SAM detection to SAM-rate discrimination with any of the lower modulation depths.

- Chapter 5. Auditory training of nonsense stimuli recognition with fixed and random babble noise: This chapter describes two nonsense stimuli Vowel Consonant Vowel (VCV) training studies. They were carried out to explore if it is possible to improve the ability to process auditory stimuli by training a listener to recognise the stimuli sound with fixed or random background noise over time. The studies showed that participants' performance was significantly improved between pre and post VCV random when the tested babble noise was random for all test groups. Better identification performance (auditory learning) occurred against a random-noise background rather than a fixed noise background. However, it was noticed in the first VCV study, that VCV identification performance was highly similar across groups except for poorer pre-test identification performance in for a random-noise training group rather than a performance with for a group trained with fixed babble noise. So results in the first VCV study cannot confirm did not clearly show whether fixed training or random training is better. The second VCV study reduced the identification performance differences across groups in the pre-test by using a lower SNR than for the first VCV study. It confirmed that random babble noise training produced better identification performance against for a random-noise background (both pre-and post-test session is random noise) than contrasted against learning with a fixed sample of babble noise. It was also investigated, through VCV research, whether learning effects generalized from training normal hearing listeners in fixed babble noise to random background noises. The second study showed that improvement for the fixed training group did not significantly differ from the control (untrained) group. Therefore, there was no generalization of learning from fixed babble noise training to random babble noises.

- Chapter 6. Single session study of nonsense stimulus recognition with fixed and random babble noise: This chapter describes a follow-up study, which was carried out in a single session experiment with both fixed and random babble background noise using VCV stimuli to explore whether test method differences (multiple training sessions versus a single training session) would lead to different results in Chapter 5 (multiple training sessions) and the study in this chapter (a single training session). The other object of this experiment was to compare listeners' performance on VCV stimuli adaption to fixed babble and random babble noise. The results showed that listener's performance from fixed babble noise was similar to the VCV identification condition that with random babble noise. It confirmed that the test method difference leads to the results differences between Chapter 5 and this chapter. It is concluded that results differ for single and for multiple training sessions.

- Chapter 7. Generalization resulting from training of speech in babble noise to other background noises: the VCV experiment from Chapter 5 showed that VCV identification with a random-noise background produced better learning than against a fixed noise. Perceptual learning studies in hearing have demonstrated that training outcomes (improvements in the ability to identify the words in sentences) are better with word and sentence stimuli than with nonsense syllables (Stacey & Summerfield, 2008). The experiments presented in Chapter 7 investigated whether the same pattern that was observed from training with nonsense syllables (VCV stimuli) against random babble background noise would be obtained using sentences as stimuli. Chapter 7 was also intended to explore whether training with Bamford-Kowal-Bench (BKB), against a babble background generalized to other background noises such as traffic or rain. The results showed that participants' performance significantly improved between pre and post-test conditions for both test and control groups. Improvement for the trained group was significantly great than that for the control group. Therefore, a generalization effect was obtained from training that involved identifying BKB speech with babble noise training to BKB with car and rain environmental sounds. Part of the learning effect was also sustained after several weeks.

- Chapter 8. General Discussion: this chapter reviews the studies reported above, and critically analyses their merits and limitations, and suggests ideas for further research.

- Chapter 9. Conclusions: this chapter summarizes the main contributions of this thesis. The learning outcomes from the perceptual learning studies throughout this thesis suggest that concepts around using random noise as training background noise, and changing background noise in perceptual learning studies, can be used as baselines to develop better training methods for training people to be more sensitive to speech sounds within various noisy environments.

Chapter 2 Literature review

This literature review chapter is divided into four sections to understand human hearing, and to identify the current challenges of perceptual learning in hearing. Part 2.1 provides brief an overview of normal hearing and of hearing loss. Part 2.2 describes methods to improve human hearing; there are two main approaches described in this part: one concerns the use of assistive devices (i.e., hearing aids and cochlear implants); the other method consists of auditory perceptual training. Following this, section 2.3 provides a critical review of previous studies that investigated adults' perceptual learning (in hearing), and three main types of perceptual learning (stimulus learning, task learning, and procedural learning). Finally, part 2.4 discusses key gaps in the current knowledge about auditory perceptual learning, and lists the research questions of the thesis.

2.1 Introduction to human hearing

Hearing is the sense that acquires sound information from the environment in our daily life. Human ears have evolved to make use of the useful sound information and to be aware of the surroundings (Yost, 2007). The following part will give a brief introduction to sound information processing via our auditory system, including what are the auditory mechanisms differences between normal hearing and hearing impaired people.

2.1.1 Normal hearing

For normal hearing individuals, when a sound wave occurs it usually travels through the ear canal before it impinges on the tympanic membrane, which causes vibration that move the ossicles of the middle ear. The main components of the ossicles of the middle ear are the malleus, incus, and stapes. The "footplate" of these middle ear bones is attached to a flexible membrane in the cochlea named the oval window. The cochlea is a bony structured and filled with fluid. As shown in Fig. 2.1, the cochlea is divided by the basilar membrane (BM) into three chambers: the scala vestibuli, the scala media, and the scala tympani. The scala media is separated from the other two chambers by two membranes: the Reissiner's membrane and the BM. The oval window is in contact with the scala vestibuli and the ossicular chain, while the round window is the membrane-cover between the middle ear and the scala tympani. With the three bones as the ossicular chain, the vibration creates a movement of the cochlea fluids from the oval window to the round window. Internal and external movements of the windows induce pressure fluctuations in the cochlear fluids, which in turn initiate a travelling wave along the BM (Pickles, 1988 & Wilson *et al.*, 2008a).

The cochlea is tonotopically arranged (sensitive to different sound frequencies at different positions) and the BM has graded mechanical properties. The base of the cochlea, which is near the stapes and the oval window, is narrow and stiff. However, the top of the cochlea is wide and flexible. These features of the BM let it respond differently at certain places for different frequencies. For example, high frequency sounds create maxima (maximum movement) at the part of the BM that is near the base of the cochlea, whereas low frequency sounds create maxima at the part of the BM that is near the apex. There are two types of hair cells: inner hair cells (IHCs) and outer hair cells (OHCs). The IHCs are responsible for converting vibration from the BM into electrical activity, while the main task for the OHCs is to change the movement of the BM. The hair cells attached to the BM are bent according to the displacements of the BM. This bending activity releases an electrochemical element that can cause neurons to fire and leads to neuronal excitation at a certain site in the inner ear. These neurons interconnect with the central nervous system and transfer acoustic information to the brain (Loizou, 1999; Wilson *et al*., 2008a).

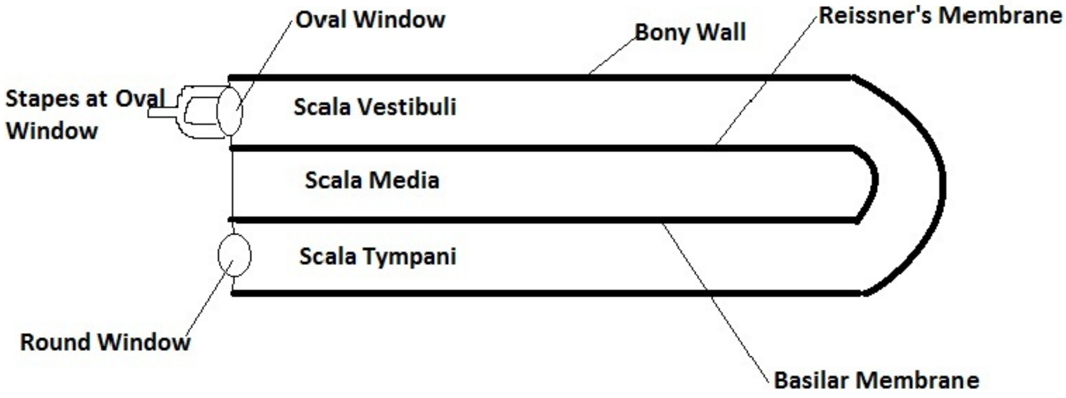


Fig. 2.1 The three chambers of the cochlea. The cochlea is separated by the Basilar and Reissner's membrane into three chambers: the scala vestibuli, the scala media, and the scala tympani. The oval window is in contact with the scala vestibuli, while the round window is the membrane-cover between the middle ear and the scala tympani.

2.1.2 Hearing loss

It has been reported that currently that 10 million people in the UK are affected by hearing loss, and by 2031 this figure will have increased to 14.5 million (Action on Hearing Loss, 2011). Mathers *et al*. (2000) stated that more than 250 million people suffered from hearing loss worldwide. According to a report from WHO (2004), it is estimated that hearing loss will be the world's top 10 burden of disease by 2030.

Hearing loss of an individual ear is defined as the mean of the hearing thresholds at 250, 500, 1000, 2000, 4000, 8000 Hz, rather than as the actual threshold at each frequency. According to BSA (2011), it can be categorized into four levels: mild (20 – 40 dB HL), moderate (41 – 70 dB HL), severe (71 – 94 dB HL) and profound (> 95 dB HL), depending on how well a person can hear the stimulus across the six frequencies. Hearing impairment can exist in an individual's one ear or both of their ears. Fig. 2.2 illustrates different hearing levels (BSA, 2011) and an example of mild to moderate high frequency hearing loss.

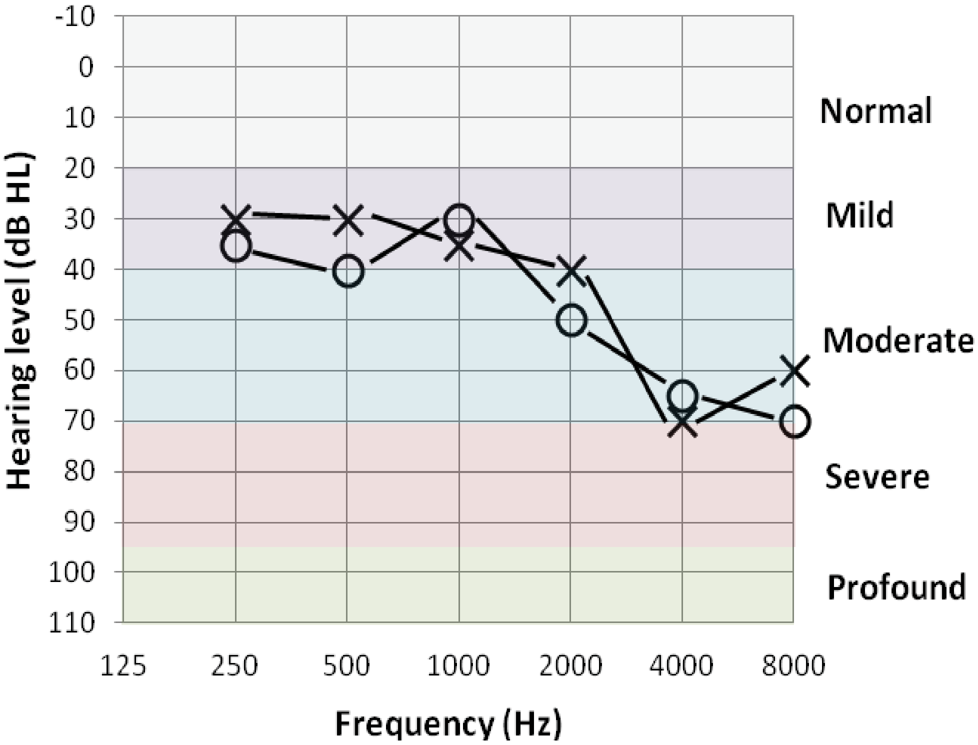


Fig. 2.2 The definition of hearing loss levels and an Audiogram of mild to moderate high frequencies hearing loss. 'O' indicates the right ear, while 'X' stands for the left ear

Hair cells are essential for neurons to transmit signal information to the brain. Unfortunately, hair cells are damaged easily due to various reasons. These include genetic defects, infectious disease (e.g., rubella, meningitis), certain drugs (mycin, streptomycin, and cisplatin), overexposure to loud noise, and ageing. Destruction of the OHCs elevates hearing thresholds and degrades frequency resolution. However, destruction of the IHCs produces more profound hearing impairment, such as total deafness. It is known that acoustic sound travels through the outer ear, the middle ear, and then the inner ear (Dorman and Wilson, 2004). If there are damaged hair cells in the inner ear, the auditory system cannot easily transform acoustic signals into a neural signal. Hinojosa and Marion (1983) indicate that the loss of hair cells is a common cause of deafness rather than a lack of auditory neurons, which provides the possibility that cochlear implants that stimulate remaining neurons can successfully convey acoustic information to the brain. Fig. 2.3 (a) indicates a simple diagram of the normal human ear, which consists of the tympanic membrane, the three bones of the middle ear, the oval window, the BM, the IHCs, and the adjacent neurons of the auditory nerve. Fig. 2.3 (b) shows a simplified diagram of the deafened human ear. Compared with a normal ear there is: (1) an entire absence of sensory hair cells, and (2) inadequate survival of spiral ganglion cells, but neural processes peripheral to cells are still viable.

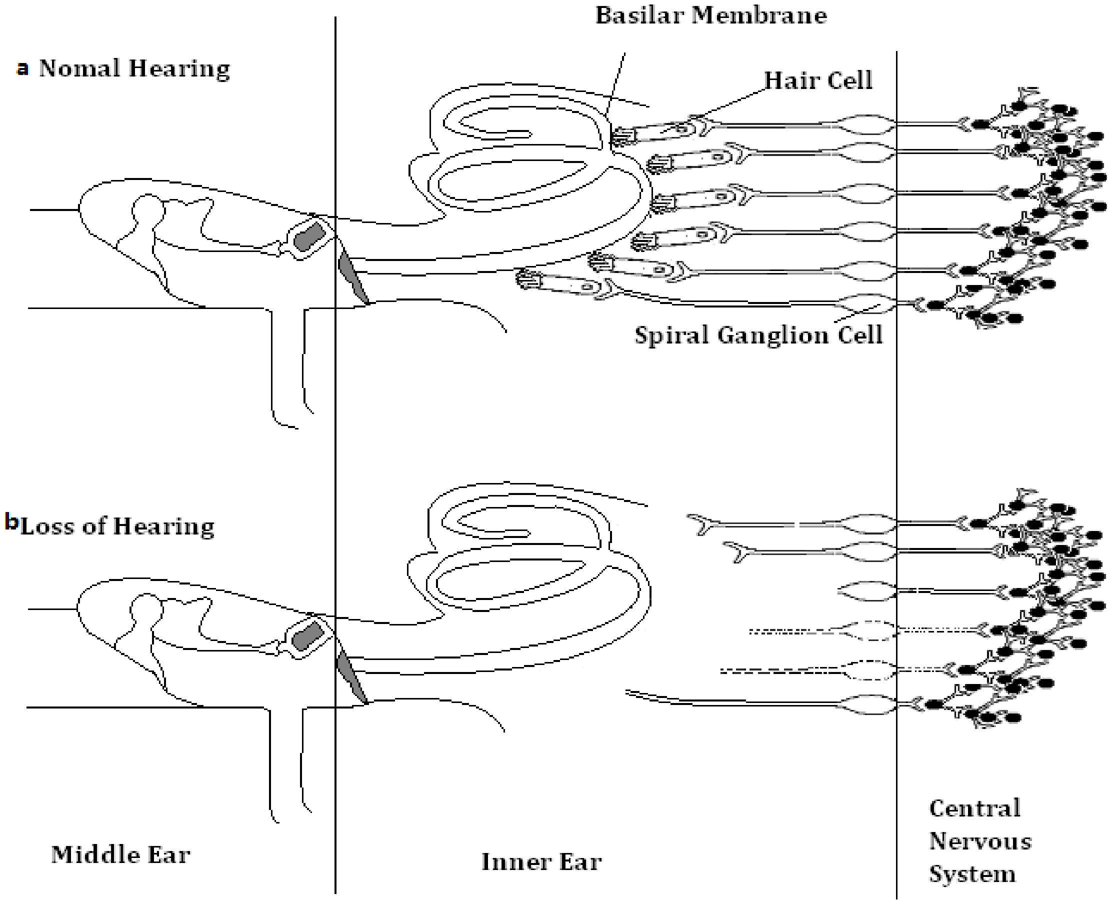


Fig. 2.3 (a) Normal hearing ear & (b) Loss of hearing ear. Reproduced from Dorman and Wilson (2004)

2.2 Methods to improve hearing

Hearing aids (HAs) and cochlear implants (CIs) are the main assistive devices to help people with hearing loss to improve their hearing ability in daily life. HAs are commonly used for hearing impaired people with mild, moderate or severe hearing losses. However, CIs can help profoundly hearing-impaired people restore partial hearing by ignoring the damaged hair cells and stimulating the spiral ganglion cells directly. Both HAs and CIs are widely used prosthetic devices to restore hearing-impaired people's auditory function. With these modern devices, the ability of hearing-impaired people to understand speech in quiet conditions almost approaches that for the normal hearing community. Although speech recognition by cochlear-implant and hearing-aid users has improved significantly over the past years, most users still experience major difficulties in noisy environments (Dorman & Wilson, 2004; Ricketts & Hornsby, 2005).

Moore and Shannon (2009) suggested that improving the ability of the brain to learn how to fully utilise prosthetic devices is as important as developments in the technology. The brain can process distorted and impoverished input sounds once a sound pattern-recognition system has been built up. However, if a complex sound pattern has never been learned during normal language acquisition or if the sound input is severely distorted then the central auditory system has to learn how to process a new array of sound inputs. Hearing assistive devices work well for hearing-impaired people who had previous hearing ability before deafness arose, or for those who received hearing assistance at very young age. Therefore, auditory training programmes have the potential to improve the performance of hearing-impaired users and help them obtain more benefit from their prosthetic device. Auditory training can also be useful to people who aren't hearing-impaired (e.g., learning phonemes in a second language) (Clarke & Garrett, 2004; Bradlow & Bent, 2008), or don't use a prosthetic device (e.g., Auditory Processing Disorder – APD) (Bradlow *et al.*, 2003; Ziegler *et al.*, 2005, 2009).

Auditory perceptual learning in hearing is defined as an improvement in the skill to detect, discriminate, or group sounds and speech information (Goldstone, 1998). Auditory learning can be obtained from a training task, the stimuli or the procedure (see details in section 2.3.4 Types of perceptual learning). Training in the auditory system may lead to long-lasting changes to the auditory perceptual system, which improve its ability to analyse environmental sounds. A human's perceptual ability to detect and discriminate sounds has been shown to improve after a certain amount of auditory training (Moore & Shannon, 2009). There are two aspects of such training. One is called the learning effect where listeners' ability to perform an auditory task or discriminate a stimuli can be improved through practice with the same task/stimuli. The other is the generalization effect, where training with one task/stimuli leads to improvement in another.

In the human auditory domain, there are various tasks that can be improved after auditory training, including those involving frequency discrimination (e.g. Irvine, Martin, Klimkeit, & Smith, 2000), temporal processing (e.g. Bao, Chang, Woods, & Merzenich, 2004; Karmarkar & Buonomano, 2003), localisation (e.g. Hofman, Rinswick, & Opstal, 1998; Wright & Fitzgerald, 2001), and speech discrimination (Aoyama *et al.*, 2004; Culter *et al.*, 2006). Wright and Zhang (2009) reviewed studies in perceptual learning and concluded that auditory learning ability generalizes across different frequencies, ears, stimulus durations, and different presentation styles (e.g. pure tone, amplitude modulated tone, narrow band tone). Auditory training is mainly affected by age and training duration (overtraining). The details of which are discussed below:

Age

The improvement of auditory learning ability is varied across different ages (Halliday *et al.*, 2008). After auditory frequency discrimination training, the mean frequency discrimination thresholds for the oldest people are slightly higher. Older people's frequency discrimination ability is not as good as for younger adults age between 18-40 years. Tremblay *et al.* (2002) also demonstrated that older adults have much more difficulty in processing time varying cues than younger adults. They compared older hearing-impaired adults with younger adults' performance on a consonant-vowel syllable discrimination task /ba/-/pa/, and found that older adults had difficulty to distinguish the voiced consonant /b/ from voiceless consonant /p/. They stated that older adults' difficulty in distinguishing consonant /b/-/p/ was due to synchronous responses to the onset of the vowel were delayed in older adults. Halliday *et al.* (2008) demonstrated that some children's frequency discrimination thresholds can achieve an adult's level, but the majority of children show a fluctuating pattern of learning. Auditory learning has a prolonged developmental time course. Human learning ability on temporal interval discrimination tasks is still immature at 14 years old (Sarro *et al.*, 2009; Wright, 2010).

Training duration (overtraining)

It is known that perceptual learning and generalization can be improved with increased training (Wright & Sabin, 2007). Wright and Sabin (2007) indicate that in order to obtain a perceptual learning and generalization effect after training, two requirements regarding training length have to be taken into account: 1) a critical amount of training has to be achieved per training day; and 2) the number of critical trials, for practice per day, is task and stimulus dependent.

The learning effect is based on the same training stimuli/task across training and post-test sessions, whilst the generalization of learning is obtained from the perceptual learning of a trained stimulus/task to an untrained stimulus/task in post-test session (which differs from the trained stimulus) (Wright, Wilson & Sabin, 2010; Henshaw & Ferguson, 2013). Therefore, increasing training duration might increase the possibility of obtaining generalization effects. However, the critical number of training trials is condition dependent, so both the task and the stimulus used in training can affect the critical number. If the number of training trials per day went beyond the minimum critical number of training trials, no additional improvement in the perceptual learning or generalization would be obtained. This might lead to overtraining. Overtraining occurs when there is an additional amount of training beyond the critical number of training trials. An additional amount of training is superfluous to learning and cannot lead to significant improvement performance for learning process) or retuning (retune) of the learning effect and also may inhibit the learning from the training task to be transferred to another task (Wright & Sabin, 2007). Wright and Sabin (2007) tested how varying the amount of training trials affected perceptual learning over multiple days with frequency discrimination and temporal-interval discrimination tasks. Twenty-seven listeners attended this experiment and were divided into four groups. They compared improvements in discrimination thresholds (through two tasks: frequency discrimination and temporal-interval discrimination), and between different training groups, which were trained with either 360 trials (frequency discrimination: n = 7; temporal-interval: n = 6) or 900 trials per day for 6 days (frequency discrimination: n = 8; temporal interval: n = 6). Results showed that perceptual learning on the frequency discrimination task required more than 360 trials per day. In contrast, perceptual learning on the temporal-interval discrimination task could occur with 360 trials per day, and additional daily training did not demonstrate any benefit to increasing the amount of performance improvement. Similar studies on a visual text-discrimination task, also demonstrated that the addition of two or more practice sessions in a single day did not provide greater improvement than one practice session (Karni & Saqi, 2003).

2.3 Perceptual learning in hearing

Many researchers have carried out studies to investigate the principles and biological mechanisms of auditory training with normal hearing people and hearing impaired people (e.g., Tremblay, 2007; Song *et al.*, 2011; Henshaw & Ferguson, 2013). Rhebergen and Versfeld (2005) demonstrated that hearing level differences may lead to different audibility and intelligibility. Therefore, it is necessary to take actions to help people improve their hearing ability in daily life. The following sections of the literature review critically evaluate previous research on auditory perceptual learning studies. They will be reviewed in two categories: one is about perceptual learning of non-speech sounds, such as pure tone or amplitude modulated sounds; the other one is about perceptual learning of speech-based stimuli.

2.3.1 Perceptual learning for non-speech sounds

The following sections divide research about perceptual learning with non-speech sounds into three categories: frequency discrimination, intensity discrimination, and spatial hearing. This review is restricted, as relevant for the thesis, to behavioural investigations of human auditory perceptual learning with adults.

2.3.1.1 Frequency discrimination

Frequency discrimination is the ability to distinguish between non-simultaneous signals due to their frequency differences. There are two main methods to measure the frequency discrimination between two signals. The first is named ‘the difference limen for frequency’. For this method there are two slightly different tones presented successively in a time. Participants are required to judge which one has the higher pitch. The order of the signals is presented randomly from trial to trial and the smallest frequency different limen will be taken as the frequency separation index for the two sounds. The second measurement method is called ‘the frequency modulation detection limen’. The sounds used in this method are commonly frequency-modulated tones with a low modulation rate. Two successive signals are presented (one is modulated and the other one is unmodulated), the participants are required to decide which one is modulated. The smallest amount of modulation that a participant can detect between the two sounds is recorded as the frequency modulation detection limen (Moore, 2004).

Amitay *et al*. (2005) reported that perceptual learning in a frequency discrimination task generalized to untrained frequencies. The generalization effect only occurred when the evaluated target frequency discrimination task shared the same frequency range with the training frequency discrimination task. Demany (1985) trained participants with frequency discrimination tasks at 0.2, 0.36, 2.5 and 6 kHz to determine whether there was a learning effect on the frequency difference limens at 0.2 kHz. This study found that training on the first three frequencies (0.2, 0.36 and 2.5 kHz) led to a similar amount of improvement for the frequency difference limens at 0.2 kHz, while training on 6 kHz led to the least improvement. This experiment suggested that generalization was limited to a certain frequency range. Psychophysical research in sound (not perceptual learning studies) indicated that the human auditory system processes tone frequencies below and above 5000 Hz differently (Attneave & Olson, 1971; Semal & Demany,1990). For tone frequencies below 5000 Hz, the auditory system processes sounds by using temporal frequency cues (Rose *et al.*, 1967; Moore, 1973; Moore & Glasberg, 1989). However, for tones above 5000 Hz, tones were coded purely by tonotopic cues. The explanation for the above results from Demany (1985) was that the first three frequencies (0.2, 0.36 and 2.5 kHz) were processed by using cues at the auditory system with 0.2 kHz, but that 6 kHz was too high to be coded by the same cues, purely tonotopic cues for 6 kHz were not effective for sounds at 0.2 kHz. A subsequent study by Demany and Semal (2002) indicated that although listeners' frequency discrimination performance was better after training than before training – to discriminate a pure tone of 3 kHz from pure tones with slightly different frequencies – this learning effect did not transfer from 3 kHz to tones of 1.2 kHz and 6.5 kHz. These results caused by the different amount of training trials used in their studies: 700 trials were used in the study of Demany (1985), while Demany and Semal (2002) used 11,000 trials. Demany and Semal (2002) demonstrated that participants in the study from Demany (1985) were not trained very extensively, so they used 11,000 trials for their studies. According to their results, listeners almost reached their asymptotic performance after test session 4 or 5 (440/550 trials) in the first ten test sessions. They concluded that the frequency-specific perceptual learning increases in the sequence of learning and leads to reducing participants' ability of generalising the learning to other frequencies. It could be considered that overtraining had occurred in the study of Demany and Semal (2002), and hence weakened the generalization effect being applied from the training frequency to the untrained one. Wright and Sabin (2007) also demonstrated that if learning on two tasks had modified different circuitries at physiological level, training on one of the task would inherit some features and then made that circuit less amenable to change.

2.3.1.2 Intensity discrimination

The smallest detectable difference in intensity between two stimuli is considered to be the threshold for intensity discrimination. The unit for detecting intensity changes is usually the decibel (dB). There are three methods for measuring the thresholds for intensity discrimination: 1) amplitude modulation detection; 2) increment detection; and, 3) intensity discrimination of gated or pulsed stimulus. Intensity comparison across frequency and time are crucial for the auditory system to identify sounds. The firing rates of the auditory fibres represent sound intensity. Intensity discrimination is degraded by non-simultaneous masking. The auditory system can improve intensity discrimination by coding a stimulus intensity reference signal to proximal signals over time and frequency (Plack & Carlyon, 1995).

Only one publication was found that deals with the effect of perceptual learning on intensity discrimination. Buss (2008) investigated the generalization of learning on intensity discrimination by training listeners (n=8) on intensity discrimination tasks with a target tone (948.7 Hz at 50 dB SPL) and masker tones (synchronously gated tones at 300 Hz and 3000 Hz, masker sound level roving from 42 to 58 dB SPL). The results indicated that intensity discrimination improved after 6 hours of training under masking conditions. However, learning on the task of intensity discrimination with masker did not generalize to intensity discrimination in quiet (during no mask sound conditions, intensity discrimination task in quiet conditions). A potential reason given for this result was that listeners fail to sustain their attention or motivation across hours of training.

2.3.1.3 Spatial hearing

The ability to determine the position of a sound source in the horizontal plane is crucial for the auditory system. Listeners can monitor environmental changes or sound direction via this ability, especially to discriminate certain sounds from noise in noisy situations (Wright & Zhang, 2006). There are two basic cues for detecting sound source location in the horizontal plane: interaural level differences (ILDs, also known as interaural intensity differences – IIDs, and interaural time differences – ITDs). Listeners can make use of spatial hearing differences between two ears to calculate the position of a sound source in space.

ITDs and ILDs are caused by sounds arriving at the further ear later, and a bit quieter, than at the nearer ear. ILDs are produced by the head shadow effect. The wavelengths of lower frequency sounds are longer than the size of the head, there is no or little diffraction effect caused by the head in such cases. Therefore, for low frequency sounds (below 1500 Hz), when sound sources are far away from the listener, the key factor for determining sound source position is the arrival time differences of the sound (ITDs) between two ears. For higher frequency sounds (above 1500 Hz), the wavelengths are shorter than the size of the head, and little diffraction occurs. In this situation, the main cue for determining sound source position involves sound arrival level differences (ILDs) between two ears (Rayleigh, 1907; Feddersen *et al.*, 1957).

i) Interaural level differences (ILDs)

Wright and Fitzgerald (2001) investigated the learning on for distinguishing small differences in interaural sounds presented over headphones. They trained one group (n=8) with a 4 kHz pure tone over one hour daily and over 9 days using an ILD discrimination task. Another group (n=8) was tested for one hour daily sessions over 10 days on an ITD discrimination task with a 0.5 kHz pure tone. Results showed that performance on the ILD task improved after two hours of training, and continued to improve with additional training. A generalization effect on ILDs was not observed for untrained frequencies such as 0.5 and 6 kH. It also indicated that training on the ILD task at 4 kHz showed no generalization effect to ITD discrimination task at 0.5 kHz. Comparing the amount of individuals' ILD improvement, the ones with the poorest initial performances demonstrated the largest improvements, while the reverse was also true.

Zhang *et al.* (2009) investigated the influence of amplitude modulation on the learning patterns of human adults during ILD tasks. They trained listeners on an ILD discrimination task with a sinusoidal amplitude modulated (SAM) tone (4 kHz carrier with modulated at 0.3 kHz). Results showed that a learning effect was observed across all of the test sessions for the ILD discrimination tasks using a sinusoidal amplitude modulated (SAM) tone (4 kHz modulated at 0.3 kHz). It was also found that training on ILD discrimination task with a 4 kHz carrier transferred to the untrained ILD discrimination task with a 6 kHz carrier, but did not transfer to an ILD discrimination task including an unmodulated 4 kHz tone. Zhang and Wright (2009) suggested that the neural system processed the ILD discrimination with SAM tones and ILD discrimination with pure tones in two different ways. For the ILD discrimination with pure tone task, the affected neural system processed different ILD values in a frequency specific manner. A stimulus-type-specific manner was used from the neural system to process sounds for the task of ILD discrimination with SAM tones. Moore (2004) also suggested that the pitch of pure tone tasks corresponded well with stimulus frequency, but the pitch for SAM sounds was close to the stimulus modulation rate and the stimulus with small changes of the carrier frequency. Therefore, although ILD discrimination with a pure tone (4 kHz) and ILD discrimination with the SAM sound (4 kHz modulated at 0.3 kHz) had the same central frequency (4 kHz), their pitches were significantly different, and that led to the learning pattern differences from pure tone and SAM tones in trained ILD discrimination tasks.

ii) Interaural time differences (ITDs)

Wright and Fitzgerald (2001) showed that learning variability on the ITD task was large. In addition, they demonstrated that if training participants obtained improvements after the initial two hours testing session, there were no improvements from additional training. However, a similar study by Rowan and Lutman (2005) suggested that a learning effect (performance improvement) on ITD tasks could be observed across test sessions, and the learning effect generalized to different stimuli, such as pure tones, transposed stimuli, and SAM stimuli. But Wight and Fitzgerald (2001) showed that there was no learning effect observed after the initial practice session (no significant performance improvement from the pre-training session to the training session). This was due to the different lengths of training per test session between these two studies. The first one was longer (720 trials by Wright and Fitzgerald 2001) than the second one (300 trials in tests by

Rowan and Lutman 2005). Furthermore, there was another study of ITD based discrimination tasks by Saberi and Perrott (1990), which supported the finding that generalization of learning could be observed from training on different ITD discrimination stimuli and sound levels. Therefore, based on the literature, both learning and generalization effects can result from training with ITD tasks.

Further studies, which specifically relate to the initial rapid learning effect of ITD discrimination were conducted by Ortiz and Wright. They highlighted three main points from the ITD discrimination test: 1) a significant learning effect could be observed within 20 minutes of testing; 2) the rapid learning effect may be caused by three kinds of learning: learning of the testing procedure, the lateralization tasks and the testing stimulus; 3) the proportion of learning was determined by the amount of training given, and the length of the rest time between test and training sessions (Ortiz & Wright 2003; Ortiz & Wright 2005).

2.3.2 Perceptual learning in speech

Human perceptual learning of speech can be affected by both the target speech sounds and background noise. Studies of auditory perceptual learning speech have developed in several directions and focuses on different aspects of this topic. In this section, for the purposes of the literature review, the perceptual learning of speech will be divided into five aspects: first, speech contrasted sounds; second, accented speech; third, talker variability transfer learning; fourth, lexical information induced perceptual learning; fifth, perceptual learning for speech in noise. All of these topics will be reviewed in relation to research studies on perceptual learning in adults with various speech target stimuli (such as nonsense syllables, word or sentence stimuli).

2.3.2.1 Speech contrasted sounds perceptual learning

Most perceptual learning research about speech contrasted sound tests was carried out at the phoneme level with various speech tasks, for example, using pseudo words (nonsense words that still sounded like a word) or nonsense syllables (consisting of vowel and/or consonant) sounds in real words or sentences. Previous studies showed that training listeners with ambiguous syllables embedded within typical word or sentence contexts led to changes of listeners' phoneme category boundaries (Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006; Maye *et al.,* 2008).

A perceptual learning study of Rochet (1995) showed that Mandarin speakers achieved more French-like voice onset time perceptual categorization after they were trained with a synthetic French /bu/-/pu/ continuum. In the same year, Flege (1995) conducted a similar perceptual learning study using Mandarin speakers. The results indicated that the ability of native Mandarin speakers to distinguish whether contrasted English consonants '/t/ and /d/' could be improved with evaluated tasks (pre-training and post-training tasks) or the same/different training tasks. As native Mandarin speakers have difficulty in discriminating the differences between /t/ and /d/ in the final position of English words, Flege (1995) trained native Mandarin speakers to identify stop sounds '/t/ and /d/' within consonant-vowel-consonant structured English words (e.g., beat-bead, bit-bid, bet-bed, bat-bad) with two different perceptual training procedures. Listeners were all Mandarin speakers living in the United States. Half of the participants were required to do identification tasks (report whether the test word's final stop sound was /t/or /d/). The other half of the listeners were trained with a task which included a categorical same (two /t/ or two /d/) or different discrimination (one /t/ and one /d/) tasks of final stop /t/ and /d/ consonants. The results from this study demonstrated that both training methods improved listeners' perceptual learning performances on the two stop consonants.

A related perceptual learning study has also been carried out with Japanese native speakers. Native Japanese speakers have difficulty in discriminating the English syllables /r/ and /l/ (Takagi and Mann, 1995). As the pronunciations of these two syllables are quite different in the Japanese and English language, neither of these two English syllables (/r/ and /l/) matched Japanese language segments. Japanese speech was phonetically closer to an English /l/ than to an English /r/ (Aoyama *et al*., 2004; Culter *et al*., 2006). However, an auditory perceptual training study of Bradlow *et al*. (1997) demonstrated that, after training, Japanese listeners identified the English syllable /r/ more accurately than /l/. It showed that Japanese leaners of English achieved more success in learning the English syllable /r/. In their study, Bradlow *et al*. (1997) trained adult Japanese speakers to identify the English contrasted sounds '/r/-/l/'. Apart from the training session, all the participants were required to pronounce the English contrasted syllables '/r/and /l/' before and after they attended their training session. Participants' post-performance was tested with old (familiar) speakers, novel speakers and novel tasks. The consequences of this training indicated that perceptual learning occurred after training participants with /r/-/l/ contrasted sounds, and in addition the learning effect was generalized to novel items by novel speakers.

Sheldon and Strange (1982) demonstrated that even though Japanese speakers could not identify the English syllables /r/ and /l/ reliably, they were still able to produce identifiable /r/ and /l/ sounds. Some Japanese listeners' production abilities exceeded their English syllable perceptual abilities, but the reverse was not true (their English syllable perceptual ability cannot exceed their syllable production abilities) (Yamada *et al.*, 1994). The perceptual training study from Bradlow *et al.* (1997) extended the results from Yamada *et al.* (1994), and confirmed the findings of Rochet (1995) that auditory perceptual training can alter listeners' perceptual speech abilities.

Apart from the studies described above, one perceptual learning study from Norris *et al.* (2003) provided further evidence of how human auditory mechanisms process speech perceptual tasks. They demonstrated that listeners use lexical knowledge to adjust their perceptual learning skills when perceiving ambiguous sounds. Listeners who heard ambiguous stimuli in the context of final word with /f/ attributed more choices on '/f/-/s/'contrasted sounds as /f/. While if participants heard a sound with final word /s/, they then categorized more choices on the same /f/-/s/ contrasted task as /s/. However, non-word ambiguous sound training led to a no sound categories shift. Results from this study also showed that perceptual training with acoustic phonetics (such as /f/ to /s/ syllables) led to phonological information remapping in the human auditory system (Greenspan, Nusbaum, & Pisoni, 1988; Eisner & McQueen, 2005).

2.3.2.2 Accented speech perceptual learning

Accented speech refers to a speech sound that has a non-pathological disorder, but there remain some noticeable pronunciation differences in the speech of native speakers (Munro & Derwing, 1995). As accented speech affects both segment and suprasegment aspects of speech signals, an accent increases the cognitive difficulty for listeners to understand, and usually leads to speech perceptual mapping failure in daily communication (Anderson-Hsieh *et al*., 1992; Van Wijngaarden, 2001). However, due to perceptual learning, on some occasions listeners can recalibrate speakers' phonemic and/or prosodic categories to adjust accidental mispronunciations between native speakers and accented speakers (Kraljic *et al*., 2008; Maye *et al*., 2008; Sidaras *et al*., 2009).

Earlier in the 1980s, researchers started to conduct experiments on accented speech perceptual learning. Gass and Varonis (1984) demonstrated that sentence transcription performance from native speakers improved after exposure to a story that was told by a non-native speaker. Another study from Wingstedt and Schulman (1987) presented Swedish utterances with a cryptic accent (one that was unusual or mysterious, and that native listeners were not familiar with) to native speakers, and found that listeners who had experienced repeated accented sentence exposure obtained higher word accuracy. The results indicated that participants could adjust the accented acoustic-phonetic/syllable into the native pronunciations of the intended phonemes.

Several other studies replicated the findings above concerned with perceptual understanding improvement with accented speech exposure. Clark (2000) trained two groups of native English speakers with accented voices for three days: one group was trained with Spanish-accented voices and non-accented voices (English). The other group was trained with Chinese-accented speech and non-accented speech

(English). Then their understanding was tested using a word intelligibility test using new sentences presented in noise. It included both trained and new Spanish- and Chinese-accented voices. The results demonstrated that both test groups obtained larger performance accuracy with the accented speech they were trained with than with the other different or new accented voices. These findings from Clark (2000) indicated that learning occurred with foreigner-accented speech sounds, but could not be transferred to new accented voices. Therefore, a lack of transferred learning suggested that speech perceptual learning is voice-specific.

A similar accented voice perceptual study from Weli (2001) trained participants with Marathi-accented words and sentences for four training days. This study demonstrated that the accuracy of transcription performance by trained participants was much higher than that of untrained participants. Bradlow and Bent (2003) also found that there was a learning effect observed after training native English speakers with two days of Chinese-accented sentence transcription tasks. In doing so they confirmed the findings of Clark (2000), from the previous paragraph.

Adank (2009) discovered that familiarity with a speaker's accented speech could provide benefits for listeners in noisy speech environments. Comparing non-native accented speech sounds with native speakers' pronunciation, the former ones' speech processing speed in noisy environments was slower than the later ones'. However, this disadvantage was reduced when listeners were exposed to accented speech for a certain amount of time (Clarke & Garrett, 2004; Bradlow & Bent, 2008). This was probably because listeners could extract certain speech rhythms, or speech syllables, once they had familiarised themselves with certain speech characteristics; they could make use of these cues to aid their perception in more complex contexts.

2.3.2.3 Talker variability induced perceptual learning

In section 2.3.2.2, listeners were able to learn new accented speech sounds which allowed them to improve their hearing ability with accented speech voices, meaning that listeners could adapt to unfamiliar speech contexts. However, considerable listener effort was still required to understand some of the foreign speakers' accented speech and demonstrating that perceptual learning with various talkers is important for hearing ability in our daily lives. The following section will focus on previous studies of talker variability induced perceptual learning with English accented speech.

The English language is recognized in 51 countries as their official language. Nowadays, there are more non-native English speakers than English native speakers. Therefore, people are required to communicate with more foreign accent language utterances than ever before (Graddol, 1997; Jenkins, 2000). For young listeners, even with no prior experience with an unfamiliar talker, they have the ability to recognize familiar words from any speaker (Hallé & Boysson-Bardies, 1994; Swingley, 2005).

The topic of perceptual learning with talker variability has been widely investigated in the context of English accented speech for English native talkers (Eisner & McQueen, 2005, 2006; Kraljic & Samuel, 2005, 2006, 2007; Maye *et al.,* 2003; Norris *et al.,* 2003). Studies in this area have also been carried out on special speakers with various types of speech sounds, such as speech from children with hearing impairments (McGarr, 1983), computer synthesized speech (Schwab *et al*., 1985; Greenspan *et al*., 1988), time compressed speech, and noise-vocoded speech (Dupoux & Green, 1997; Davis *et al*., 2005). Evidence from perceptual learning studies with several different talkers demonstrate that listeners adjust themselves to accented speech, and this ability could generalize to new utterances of the same sound (Kraljic & Samuel, 2006; McQueen, *et al*, 2006). In addition, listeners were able to adapt to novel speakers and accents with appropriate, accented, sound exposure (Clarke and Garrett, 2004; Bradlow & Bent, 2008; Sidaras *et al.,* 2009).

Norris *et al*. (2003) demonstrated that perceptual learning of speech sounds could adapt to a particular speaker's accent (in other words, a familiar talker's accent). Research from Eisner and McQueen (2005) showed that learning did not generalize from ambiguous fricative sounds to a new speaker. Kraljic and Samuel (2006) reported that cross talker generalization was obtained from perceptual training with ambiguous stop phonemes. However, it was noted that all of these studies used nonsense syllables as test sounds, which indicates that phoneme perceptual learning does not transfer across speakers in all speech situations. Results from Norris *et al*. (2003), Eisner and McQueen (2005), and Kraljic and Samuel (2006) also suggest that the auditory system can make use of different levels of sound information, and that a generalization of learning could occur when contextual and speaker-related information is available to be detected. Some other studies on talker variability induced perceptual learning complemented the conclusions above, and indicated that if listeners were exposed to training with similar patterned speech (for example, purely Japanese accented English), then speech identification performance improvement could generalize to previously unheard speakers with the same accent (Japanese accented English but spoken by unfamiliar people) (Weill, 2001; Bradlow & Bent, 2008).

Bradlow and Bent (2008) examined whether highly variable training sessions led to better performance. Native speakers of American English in the test group were trained with five different Chinese accented English speakers. In the meantime, there was another test group of American native listeners was trained with a single Chinese accented English speaker. During the training sessions, they were all required to transcribe the accented English sentences that they were trained with. In order to acquire a baseline measure of the training effect for accented English sentence transcription, participants in the control group were tested with native non-accented English sentences. All of the tests were carried out with multi-talker babble background noise, and the signal to noise ratio (SNR) was +5dB SNR. Results showed that the performance of participants who trained with multiple accented talkers improved to a greater extent than participants in the control group (almost 10% pts performance improvement). Improved performance was also found in the test group, which was trained with single talker speech, and tested with the same talker. However, the results differed when listeners who were trained with a single talker were tested with sentences spoken by a different talker (their improvement was not as good as the control group). These results indicate that training with a single accented talker led to an improved perceptual learning effect with the same talker's speech, but learning was not transferred to other speakers even with similar accents. However, training with multiple accented talkers' speech led to both perceptual learning and generalization effects. But the transfer of learning was limited to speakers with the same type of accent, not to disparate accents, such as a Slovakian accent. Baese-Berk *et al*. (2013) extended the work of Bradlow and Bent (2008). They exposed listeners to five different language backgrounds during training. The results from their study demonstrated that multiple accented speech training generalized to novel talkers and the transfer of learning were caused by systematic variation during training.

2.3.2.4 Lexical information induced perceptual learning

Lexical information (morphemes, words or meaning clues in a text) induced perceptual learning plays an important role in the acquisition of speech information. It can help people to make a slow speech process faster (Strange, 1995). Fu *et al.* (2005) compared both word (1,000 monosyllable words) and sentence-training (HINT sentences) methods to find which training approach led to effective improvement for normal hearing listeners to identify spectrally distorted vowel and consonant sounds. The test results from their studies showed that both these training methods achieved significant improvement in identification of consonants, but the word training method was more effective than the sentence based training method for improving the ability of listeners to discriminate vowel sounds. So word based training might be better than sentence based training in developing speech-perception skills for cochlear implant (CI) users. However, there were two limitations in the above study: first, they did not include a control group (without training) to account for procedural learning effects, which may be caused by exposure to the experimental stimuli (Robinson & Summerfield, 1996); secondly, they did not make use of lexical information cues from sentences in their study (they did not test sentence perception). They purely focused purely on the ability of listeners to identify vowels or consonants in monosyllabic words. Nevertheless, information about the perception of a sentence could provide more systemic information about our daily communication. Therefore, if lexical information cues were used from sentence-based training, the opposite conclusion might be obtained from Fu *et al*. (2005).

Another study of speech perceptual training from Stacey and Summerfield (2007) used the same training method (word- and sentence-based training) as Fu *et al*. (2005). They made some changes, such as extending the duration of each training session (from 15 minutes to 1-2 hours), decreasing the number of test days from 5 to 3 days, and they evaluated the training effect of noise-vocoded speech with normal hearing people. Stacey and Summerfield (2007) found that both word and sentence training led to improvements in the identification of spectrally distorted speech (words in a sentence), and that training with several talkers was more effective than training with a single talker.

Kidd *et al.* (2007) investigated lexically induced perceptual learning with environmental sounds (such as dogs barking, doors slamming and cars starting), and also with speech sounds, and suggested that the auditory perception of both speech and environmental sounds was highly reliant on the recognition of familiar stimuli. However, research from Burkholder (2005) showed that speech identification performance from participants trained with anomalous sentences generalized to new speech materials and environmental sounds. The transfer of learning was found to be larger when training with non-meaningful speech sounds than training with meaningful speech sentences. Loebach and Pisoni (2008) also demonstrated that perceptual learning transferred from training with environmental sounds to both untrained environmental and speech sounds. Following the study of Loebach and Pisoni (2008), Shafiro *et al*. (2012) found that perceptual learning generalized from environmental sounds to speech and novel environmental sounds in patterns of exposure (repeated short test) and training. However, the greatest improvement in performance occurred for patterns of training rather than patterns of exposure alone (repeated short test).

2.3.2.5 Perceptual learning for speech in noise

The flexibility of the auditory system may provide potential cues for training people to better detect speech in noisy environments. In order to improve people's speech perceptual ability via speech-in-speech environments, it is better to train people with speech background noise rather than train them with non-speech masking sounds (Van Engen, 2012). Research on the central auditory system's plasticity has shown that auditory training with short-term speech stimuli changes cortical and subcortical responses, and can enhance speech perceptual learning, especially with speech in noise training (Tremblay & Kraus, 2002; Wong & Perrachione, 2007). Davis *et al*., (2005) demonstrated that even without feedback, a naive listeners' speech intelligibility displayed rapid changes as a consequence of short periods of initial exposure to vocoded stimuli, and that this ability could also generalize to untrained speech frequency bands with some variation or different vocoder carriers (Hervais-Adelman *et al*., 2011). In addition to the generalization of noise-vocoded speech to untrained stimuli, Clarke and Garrett (2004) found that generalization to untrained words was also observed in the identification of accented speech in noise.

In recent studies, some used fixed babble noise (same section of babble noise) as the background noise, whilst others used random babble noise (different section of babble noise) (Wilson, 2003; Killion *et al.*, 2004 Engen & Bradlow, 2007). However, not many researchers have investigated the influence on perceptual learning to compare any differences that arose when the target sound's background noise fluctuated or was held constant across training sessions. A study from Felty *et al*. (2009) demonstrated that listeners obtained better word recognition performance (words in a sentence) as a result of fixed babble noise rather than random babble noise. In their study, they compared listeners' word identification performance with fixed and random babble noise, this occurred within one test session without training.

2.3.3 Types of learning

Based on the review of the literature in section 2.3.2, many studies on auditory perceptual learning demonstrate that auditory training leads to both perceptual learning and a generalization effect. Performance learning improvement is mainly attributed to stimulus learning, task learning, and procedural learning. Details of these three perceptual learning types are summarized below.

2.3.3.1 Stimulus learning

Stimulus learning refers to learning that is associated with specific feature values of the stimulus (Ahissar & Hochstein, 1996; Robinson & Summerfield, 1996; Rubin, Nakayama, & Shapley, 1997), such as a specific tone frequency (e.g., Demany, 1985; Irvine, Martin, Klimkeit, & Smith, 2000; Delhommeau, Micheyl, Jouvent, & Collet, 2002), or particular line orientation (e.g., Vogels & Orban, 1985; Shiu &

Pashler, 1992). Stimulus learning can be observed for the stimulus that was encountered during training (purely learning, not including transfer learning or generalization). It can also be obtained from a stimulus that shares a particular feature with the training stimulus (including transfer learning or generalization) (Ortiz & Wright, 2009). The greater the familiarity listeners have with the stimulus materials, the faster is the rate of the learning. For example, if the aim of research is to train listeners to understand speech in daily life, in order to obtain more stimulus learning performance improvement the selected training stimuli material should include a variety of talkers and phonetic contexts, rather than be limited to few stimuli (few talkers and phonetic contexts) in the training session (Greenspan, *et al.*, 1988, Robinson & Summerfield, 1996). Ortiz and Wright (2009) demonstrated that stimulus learning occurred in both multiple- and single-session training.

2.3.3.2 Task learning

Task learning refers to learning that is associated with the particular perceptual judgment to be made for the purpose of the task (Robinson & Summerfield, 1996; Ortiz & Wright, 2009), such as frequency discrimination (e.g., Demany, 1985; Wright, 2001; Delhommeau *et al.*, 2002; Delhommeau, Micheyl, & Jouvent, 2005), sound duration discrimination (e.g., Wright, Buonomano, Mahncke, & Merzenich, 1997; Wright, 2001), and speech discrimination (e.g. Culter *et al.*, 2006). In contrast to the various cases of stimulus learning following either multiple- or single- session training, task learning follows training experiments with a multiple-session design (Ortiz and Wright, 2009). A greater generalization of auditory learning performance can be obtained if a similar task is used in both the outcome measures and the training sessions (Robinson & Summerfield, 1996). For example, if the purpose is to train listeners to do speech identification tasks, in order to obtain greater transfer learning performances, then the training task should incorporate speech identification tasks rather than speech discrimination tasks.

Perceptual learning research in vision demonstrated that task difficulty (a change of test stimulus/task lines or time intervals between the target and the mask) can affect the perceptual learning progress (Linkenhoker & Knudsen, 2002; Ahissar & Hochstein, 2004). Linkenhoker and Knudsen (2002) found that barn owls adapted to small size shifts in visual experiences, but if the shifts were made in larger increments, no learning was observed in these adult barn owls. In addition, research with humans, by Ahissar and Hochstein (2004), noted that manipulating training task difficulty could cause changes in visual perceptual performance. They required participants to view arrays of oriented lines, and to decide which one included a single, oddly oriented line. The task difficulty was adjusted by changing the time interval between exposure to the target and a subsequent mask. They found that when the task was made easier (with a longer time interval between target and mask), perceptual learning improved more quickly and generalized to novel orientations. In contrast, when the task was made more difficult (with a shorter interval between target and mask), learning was slower, and was specific to the trained orientation and location. Amitay *et al*. (2006) reviewed both animal (Linkenhoker & Knudsen, 2002) and human visual perceptual research studies (Ahissar & Hochstein, 2004) and concluded that task perceptual learning performance improvement was observed from easy to difficult training, however, if the task was too difficult at the beginning, training might not occur.

2.3.3.3 Procedural learning

Procedural learning is the learning of the test components because of familiarity with the test, but excluding the training experience from trained tasks and stimulus (Robinson & Summerfield, 1996; Ortiz & Wright, 2009). The components may include the experimental setting, the test methods, test response requirements and general strategies for doing the test tasks (Robinson & Summerfield, 1996; Delhommeau *et al*., 2002; Demany & Semal, 2002; Hawkey, Amitay, & Moore, 2004). In order to speed up part of this learning process, many researchers require participants to take part in a brief pre-training (practice) session before the real test. The pre-training session can help the listeners to familiarise themselves with the test procedure (Irvine, Martin, Klimkeit & Smith, 2000; Demany & Semal, 2002). Hawkey, Amitay and Moore (2004) suggested that the length of pre-training should be limited, otherwise the participants might lose interest in the test or the early learning effect could happen before the real test. All of these factors will affect final learning and generalization effects. Wright and Fitzgerald (2001) demonstrated that a rapid early phase perceptual learning or generalization effect could be considered as procedural learning.

2.4 Gaps in knowledge

Based on perceptual learning literature on hearing over the last few years, researchers have focused on non-speech perceptual learning, on frequency discrimination, intensity discrimination and spatial hearing. However, a gap exists around perceptual learning and the importance of different modulation depths for SAM stimuli. Wright and Zhang (2009) showed that auditory learning generalizes across frequency, ear, stimulus duration, different presentation style, etc. However, Fitzgerald and Wright (2011) argued that the cross-learning effect could not generalize from SAM detection to SAM-rate-discrimination.

Fitzgerald and Wright (2011) investigated whether training with SAM detection task can generalize to SAM rate discrimination task with same trained rate and carrier band or SAM detection task with untrained rates. Eighteen listeners participated and were randomly placed into two groups: trained group (n=9) and control group (n=9). They trained group completed all the pre-, training and post-test sessions, the control group only attended the pre-and post-test sessions. Both the pre- and post-test sessions including five SAM detections (target sound was a 3-4 kHz carrier modulated at 30, 80, or 150 Hz, or 0.5-1.5 kHz band carrier or 5 kHz low-pass carrier modulated at 80Hz) conditions and one SAM rate discrimination (standard sound was 3-4 kHz carrier modulated at 80 Hz with 100% modulation depths, target sound was 3-4 kHz carrier with a faster modulation rate) condition. Listeners were trained with 720 trails SAM detection task (3-4 kHz band pass carrier modulated at 80 Hz) per day for six to seven days. Results showed that training-induced learning did not generalize to SAM detection untrained carriers at 0.5-1.5 kHz and 5 kHz low-pass and also not generalize to rate discrimination task with the trained rate (80 Hz) and carrier band (3-4 kHz). However, the learning generalized to SAM detection with two untrained rates at 30 and 150 Hz. Fitzgerald and Wright (2011) demonstrated that sensitivity training on detection modulation depth had no advantages to a rate discrimination task with 100% modulation depth, as 100% modulation depth was well above any minimum threshold to get the best performance. Therefore, the first step of this PhD project is to see whether there will be a generalization effect from training on SAM detection test to SAM-rate discrimination test with three different fixed modulation depths. It was hypothesized that a generalization effect may occur from SAM detection to SAM-rate-discrimination – if lower modulation depths are used for the SAM-rate-discrimination tasks.

Regarding auditory perceptual learning studies in speech, most clinical rehabilitation work has attempted to train people to understand speech material better in a quiet environment. Typically, auditory perceptual learning programmes present speech with no background noise or other competing environmental factors. Auditory training in a quiet environment may help to focus listeners' attention on detecting subtle listening cues, but it is not certain whether these auditory cues can be made use of in noisy environments (Fu & Nogaki, 2005).

Even though some speech studies use noise in their research, most speech in noise auditory perceptual training studies are highly specific to changing the tasks by using several different signal sounds (such as speech contrasted sound tasks, accented speech tasks, or different talkers) alongside the same background noise (Burk *et al.,* 2006; Yund & Woods 2010). They also show that learning occurring from training is specific to certain trained speech materials and parameters of the background noise (such as the signal to noise level; and noise type: white noise, speech shaped noise or babble noise; etc.). In addition, although most speech perceptual studies in this area show a generalization effect from trained to untrained stimuli, they are mainly focused on changing the target stimuli using amplitude modulated sounds or speech stimuli, less research has examined the generalisation effect by changing target stimuli from speech sounds to environmental stimuli.

To date, although some studies have used environmental sounds in their experiments, environmental sounds are usually used as the target sound, not as background noise. It is known that the ability to detect speech signals in a noisy environment is critical in people's daily communications. However, not many researchers have explored the auditory learning and generalization effect of using environmental sounds as the background noise.

Perceptual learning studies in the visual domain show that people can improve their detection performance by learning to ignore (visual) noise. One study showed that once participants have learned to ignore fixed (repeated) trials of visual noise, and can successfully detect targets then this skill transfers to new, random visual noise (Schubö, Schlaghecken and Meinecke, 2001). In the visual experiment, a texture segmentation task was used. For this task, a surface texture (a field of short, tilted lines) was presented very briefly (33 ms) and followed immediately by a masking stimulus. The texture is either continuous ("no target") or contains a discontinuity (a small area where lines are tilted in a different direction – the "target"). Participants have to indicate whether or not a target was present. With a homogenous mask (where the mask has a simple, repetitive structure, and the same mask is presented in every trial of the experiment), participants quickly learn to distinguish between target and no-target stimuli. If, in contrast, the mask has a heterogenous structure (no simple repetitive pattern), and a different mask is presented on each trial, no learning occurs. However, if a heterogenous mask is 'frozen in time' (i.e., the mask stimulus is fixed. It means that the same unstructured mask is repeated on every trial of the experiment), small but significant learning can be observed. If the same participants who were trained with frozen (fixed) heterogenous masks are then tested with ‘unfrozen’ (random) heterogenous masks, they show successful performance learning even when the mask is renewed on each trial.

Motivated by the evidence above in visual research, I will explore if it is possible to improve the ability of listeners to process auditory stimuli by training a listener to recognise a stimulus sound in a fixed sample of background noise (fixed noise), or a sample that changes at random over time (random noise). In daily life, people experience more communication conditions with random background noise than with fixed background noise. It will be useful to find which training method (fixed or random babble) is better to obtain auditory learning or generalization to speech sentences against different random noise conditions. The next stage of this PhD project will be focused on training more complex speech sounds (such as vowels, consonants, words, sentences and so on) with different background noises (car noise, fixed babble noise, random babble noise, etc.) as well as exploring the plasticity of auditory sound identification in noisy environments.

The aim of the research reported in this thesis was to learn from studies of perceptual learning in normal hearing people with various background noises, and to use this knowledge to explore effective ways of improving hearing ability in challenging conditions. The learning outcomes from the studies reported in this thesis can provide suggestions for further studies towards creating clinical tools for the training of hearing-impaired persons to improve their hearing ability in everyday noisy environments. Based on this broad project aim, the general question for this project was

Whether changing background noise can benefit auditory training for normal hearing people in challenging conditions?

In order to answer this general question, four objectives were set below:

Research objectives:

1. To determine whether after training people on a non-speech task (SAM detection) learning will generalize to another non-speech task (SAM-discrimination) with lower modulation depths (70%, 40%) (Chapter 4).

2. To investigate perceptual learning effect using nonsense syllables speech sound identification performance with fixed and random babble noise training (Chapter 5).
    - To identify whether nonsense syllables speech sound identification performance is affected by fixed babble noise training compared to random noise training.
    - To identify whether nonsense speech sounds in a fixed sample of babble noise generalizes to the identification of non-sense speech sounds in random samples of babble noise.

3. To investigate single session nonsense speech sounds adaption to fixed babble noise and to random babble noise (Chapter 6).
    - To identify whether nonsense speech sounds adaption to fixed background noise is different to nonsense speech sounds in random babble noise background noise.

  - To identify whether test method differences in Chapter 6 (single session nonsense speech sound recognition in noise) will lead to result differences in Chapter 5 (multiple sessions of nonsense speech sound recognition in noise).

4. To test whether training people on speech tests (such as those involving words in sentences) in babble noise will improve their ability to understand speech in other background noises (such as car and rain noise) (Chapter 7).

Chapter 3 Research methodology

3.1 General plan for the whole project

In order to answer the research questions, the whole PhD study was divided into two general categories, one was about the non-speech test with normal hearing people (NHs), it is considered as the step one in the general plan, and the other one was the speech test. Regarding the speech test category, it was separated into another three steps (step two, three and four), the step two and three was conducted on NHs with nonsense syllables stimuli, and the identification of speech sentence with various environmental background noises was tested in the final step four. The general plan of the whole PhD work can be seen from Fig. 3.1 and the following paragraphs will describe more details about each of the experiments.

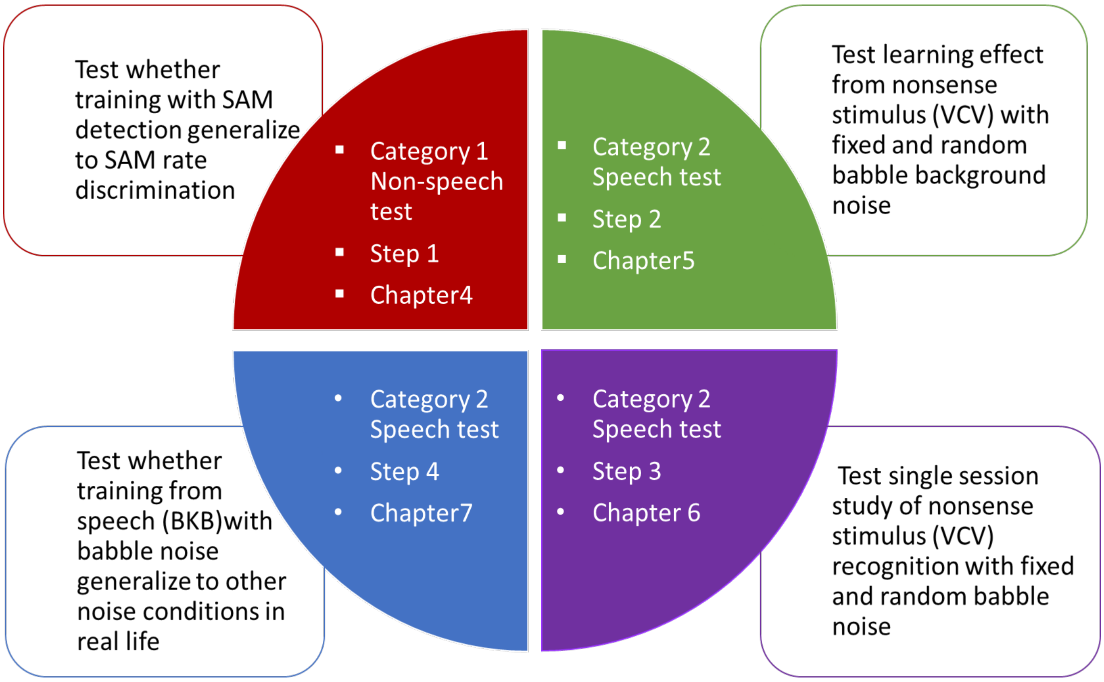


Fig. 3.1 The general plan for the whole PhD project. (SAM: sinusoidal amplitude modulation VCV: vowel-consonant- vowel BKB: Bamford-Kowal-Bench)

The first step involved a psychoacoustic training study (Chapter 4). It explored whether learning through SAM-detection tasks would generalize to SAM-rate-discrimination tasks with different fixed modulation depths. Based on test experiences and results from the preliminary psychoacoustic study in step one, the second speech test step (Chapter 5) was carried out to see what learning effect would be obtained from nonsense speech tasks with fixed or random babble background noise. The identification of nonsense speech sounds in fixed babble and random babble noise were compared, and intended to determine which was more effective to observe perceptual learning. Following the experiments from step two, the third step (Chapter 6) investigated perceptual learning for nonsense speech in fixed and random babble noise during a single session. Results from step three were also compared with the study from Chapter 5 to assess whether differences in sound perception results were due to test design differences (multiple training sessions in chapter 5 versus a single training session in Chapter 6). Step four (Chapter 7) applied the test theory, from step two and three, to train people with random babble noise to see whether it was transferable to other noise conditions in real life.

3.2 Participants

Test listeners were recruited from the student and staff population of The University of Warwick. Participant recruitment was approached through word of mouth and advertisements on university notice boards. The first psychoacoustic training study did not request any specific language requirements from listeners. All other speech-related experiments required the participants to be native English speakers. A consent form was provided before potential participants attended the study to give them enough time to consider/reconsider taking part. Participation in the study was completely voluntary. No pressure was exerted on potential participants to take part. Any refusal to attend did not affect the participant in any way. Participants had the right to withdraw from the study completely, and to decline any further contact from the researcher after withdrawal. A decision to withdraw at any time, or a decision not to take part in this study, did not affect the standard of care they received (such as hearing threshold check and knowledge about how to protect their hearing in daily life). Participants were paid £5 per hour for their time.

Participants with normal hearing were selected for each experiment based on the following inclusion standards:

- Adult subjects aged between 18 years to 40 years, who are willing to participate in this study.
- Normal hearing subjects (pure tone audiometry threshold ≤ 20 dB HL, 250 Hz -8 kHz).
- Have normal middle ear and external ear. Have no current ear problems (e.g. pain, ear infection, medication for ear problems) (information obtained by asking about ear problem history).
- No complaints of suffering from tinnitus or sensitivity to loud sounds.
- No exposure to loud noise(s) in the past 24 hours.
- Not a regular user of known ototoxic drugs (e.g., aspirin, gentamicin, tobramycin, cisplatin and carboplatin).

Each listener was given a consent form (see Appendix 1) and participant information leaflet (see Appendix 2 for an example of the information leaflet about the SAM stimuli experiment) to provide them with ample time to read the information again before they signed anything. Participants were free to ask questions about the nature of the study. The main test was only carried out once participants' questions had been answered and they had completed the consent form. Before the main test was conducted, the participant was trained to be familiar with the test procedures; training in this part included two parts: firstly, to go through the test procedures with the participant; secondly, to let them listen to the test stimuli and have a feel for what the test stimuli sounded like (details can be seen in each chapter method session). We ensured that they understood the purpose of the study.

3.3 Ethics consideration

Considering all of these experiments were conducted with human participants, ethics approval needed to be obtained before conducting any experiments. The ethics approval of the first experiment was agreed by the Biomedical and Scientific Research Ethics Committee (BSREC) of The University of Warwick on 5th March 2013 (See Appendix 3 and 4). Apart from the stimuli used, the procedure and calibration were the same in the rest of the experiments. Therefore, the BSREC gave an extension of their ethical approval to cover all subsequent experiments for the PhD. The ethics number is REGO-2013-065 (See Appendix 4).

All information collected during the research was kept strictly confidential. The data were made anonymous during the data analysis stage. Each participant is distinguished by a unique ID number in each experiment. Therefore, participants' personal information can not be identified in any report/publication. The un-anonymised data was stored only on the Chief and Principal Investigators' personal computers for 10 years. Only primary research data, which cannot identify individuals, will be published in research journals or presented at conferences. The data collected from this experiment will make a contribution to research knowledge.

3.4 Test procedures consideration

Several methodologies are commonly used to investigate human auditory perceptual learning. Each of approach has advantages and disadvantages (Shofner & Nieiec, 2010). Normally the maximum-likelihood procedure, the adaptive staircase procedure, and alternative forced choice methods were used to detect hearing thresholds in perceptual learning studies (i.e., pure tone thresholds; speech in noise thresholds, etc.). The former two test procedures are usually applied in hearing studies with one test sound. They were categorised as adaptive tracking procedures. An alternative forced choice method is typically used when a target sound is presented with another two or more options. Some other test procedures can also be used, such as percentage of correct responses, electroencephalography (EEG), functional magnetic resonance imaging (fMRI), etc. The following sections 3.4.1 to 3.4.3 describe more about the adaptive tracking procedures, the forced choice procedure and other procedures.

3.4.1 Adaptive tracking procedures

There are two adaptive procedures normally used in psychoacoustic tests: the maximum-likelihood procedure, and the adaptive staircase. For the maximum-likelihood procedure, the initial sound level used for a test is an estimation of the supra-threshold, which is from the probable range of the accurate signal threshold. For this test procedure, several sets of stimulus values are used to cover the correct threshold area. A psychometric function is assumed for each stimulus set. Any of the stimulus set can be presented at the initial trial. Each wrong stimulus is considered as the potential threshold until the most accurate one is found. The most probable threshold is calculated based on the observed accuracy of the responses. All previously accumulated information is used to calculate the most probable threshold. The test procedure concludes with an accurate signal threshold (the most probable threshold). The procedure has a variable step size, and provides the correct threshold after several trials (Shelton & Scarrow, 1984). The strategy used in adaptive staircase techniques is different from the maximum-likelihood. For the adaptive staircase procedure, the starting stimulus level is fixed and, usually, nearby the presumed signal threshold. During this test, the signal threshold is determined by increasing (missed sound) or decreasing (correctly detected sound) the test stimuli level in a fixed step size. The most commonly used sequence is to decrease the stimulus level after two right responses, and to increase the signal level following a wrong answer (two-down, one-up) (Shelton & Scarrow, 1984). However, Saberi and Green (1997) showed that when comparing the two-down, one-up algorithm with the three-down, one-up algorithm, the latter one has a steeper slope on the psychometric function. Fig. 3.2 shows a transformed up-down staircase with a three-down, one-up.

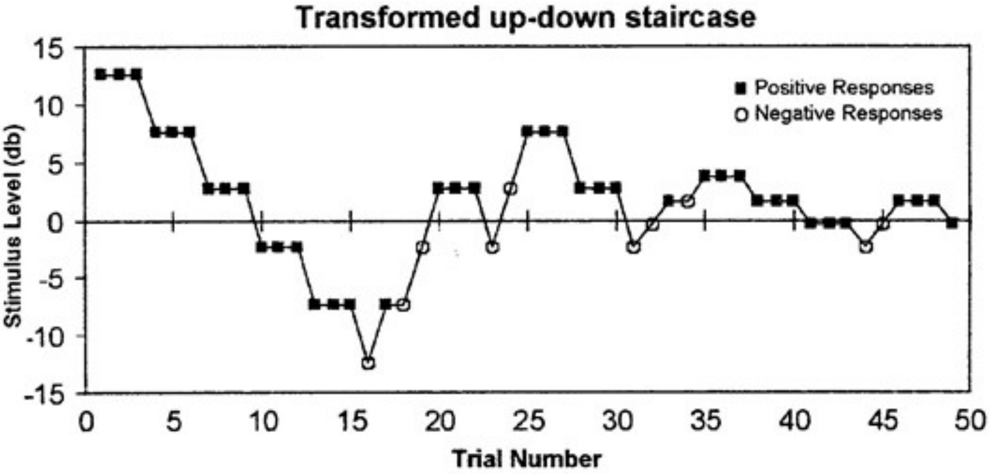


Fig. 3.2 A transformed up-down staircase with a three-down, one-up algorithm (adapted from Leek, 2001).

3.4.2 Alternative force choice (AFC) method

The Alternative Forced Choice (AFC) method is commonly used in both psychophysical and speech threshold in noise experiments. It aims to obtain an estimate of a characteristic threshold value (i.e., the threshold for the audible sound pressure of a listener's response to temporally modulated sound waves). This forced choice detection method works by alternatively presenting the target sound signal (e.g. modulated sound) and other options (e.g. non-modulated signals) over pre-defined intervals to listeners. The listeners are then required to decide which interval possesses a target signal. Noticeably, the target signal appears to be randomly presented during testing procedures. In addition, depending upon the purpose of the experiments, the presence of test (reference) signals and target sounds can be simultaneous or sequential (Jennings, 2005). For example, if the purpose of the experiment is to identify signal (target sound) from background noise (reference sound), the test and target sounds will be simultaneous. If the purpose of the experiment is to identify two similar sound level differences, these two signals will be sequential. The listeners' responses should be based on their sensations of the different stimuli, or matched impressions in their minds. The minimum number of forced choice intervals is two (2-AFC), whereas, in practical experiments, a range of 2 and 8 choices could be possibly used (m-AFC). A disadvantage of this method is that threshold values can not be obtained directly and have to be derived from the psychometric function (Leek, 2001).

Schlauch and Rose (1990) identified that less variability and bias in threshold measurements occurred as the number of intervals increased − especially for comparisons between 2AFC to 3AFC, and 3AFC to 4AFC. They also suggested that this bias was a result of behaviour near chance performance, and the effects of guessing. By fitting trial-by-trial data, using a pre-fit method, the thresholds recovered some of the bias that was associated with all these adaptive procedures. In order to improve efficiency and reduce bias, it is recommended to use an adaptive track with small step sizes (e.g. to use 1dB as step size rather than 5dB) to enable the estimation of a threshold in the tracking procedure (Leek, 2001).

Klein (2001) pointed out that 2-AFC discrimination data had the advantage of providing a straightforward view of psychometric functions, and a simple calculation of the interval bias. However, it was also claimed that the index of perceptual detectability (the probability of correct detection), not just for 2-AFC but for m-AFC, suffered from a bias of interval choice which, in turn, produced an under-estimate of the detection threshold. Katkov, Tsodyks and Sagi (2006) revealed that 2-AFC was not always a suitable option for acquiring reliable estimates of mean internal response and noise amplitude from psychophysical sensory discrimination data, which was mainly subject to high sensitivity of sampling errors.

Moore *et al*., (2005) pointed out that although 2-AFC seemed to be more robust compared to simulated variations for inattentive observers, 3-AFC appeared to be more accurate to track the mean threshold of attentive subjects. In terms of between participant variability, 3-AFC can reduce between-subject variability, which for 2-AFC, is elevated. Therefore, 3-AFC is superior to 2-AFC in this aspect (Shelton & Scarrow, 1984). Grose and Hall (1993) indicated that 3-AFC could minimize stimulus uncertainty, and decrease the guessing probability for a correct response. It could also allow the listeners to choose the different interval from the test intervals without being familiar with the signals' characteristics. Both 2-AFC and 3-AFC could use the adaptive procedure to obtain a threshold, but considering some other factors, 3-AFC is more suitable to obtain an accurate signal threshold in psychoacoustic tests. Table 3.1 shows the main differences between 3AFC and 2AFC.

Table 3.1 The main differences between 3AFC and 2AFC

Names Differences	***3.4.2.1 3AFC***	2AFC
Characteristics	3 Zero- order	2 Zero- order
	9 First- order	4 First- order
Advantages	Attentive observers	Inattentive observers
	Minimize uncertainty	Less time consuming
	Reduce guess probability	Less stimulus variance

3.4.3 Other procedures

Apart from the test procedures listed above, a percentage of correct responses is frequently used in hearing studies to evaluate people's performance in terms of accuracy and improvement. Results from this test calculate a percentage of correct responses based on the total stimuli presented, thus the results range is between 0 to 100%. A percentage of correct response measurement procedures was used for all of the speech test studies in this project. A percentage of VCV stimuli or word corrections in background noise was calculated to measure listener's performance. There are also some other procedures, such as electroencephalography (EEG), functional magnetic resonance imaging (fMRI) etc., used for auditory perceptual learning studies. However, as these methods investigate human behaviour with neuroscience aspects they were not included in this project.

In summary, apart from the non-speech study which used a three interval, three alternative forced choice (3I-3AFC) adaptive procedure, the rest of the studies, in this project, used a percentage of correct responses to measure listeners' performance.

3.5 Considerations of feedback

There was a debate about whether to use feedback for perceptual learning. Some research have demonstrated that feedback is necessary for perceptual learning (Herzog & Fahle, 1997; Seitz, Nanez, Holloway, Tsushima, & Watanabe, 2006), However, others insisted that feedback is not necessary for perceptual learning (Fahle, Edelman, & Poggio, 1995; Karni & Sagi, 1991). Mollon and Danilova (1996) demonstrated that if feedback was not provided for people who did experiment using above threshold stimuli, participants could have various degrees of confidence about their performance. If feedback was provided for these people, it could be used as a de facto feedback for them to confirm their responses were correct. As feedback can give trial to trial basis and affect listeners' perceptual learning process, some studies prefer to provide feedback when a series of tasks had been completed.

Davis *et al.* (2005) compared auditory perceptual learning performance with / without feedback. In their experiment, there were three groups (two feedback groups and one no-feedback group). All the groups listened to an artificially distorted speech (marked as D) and then listeners were required to write down what they could hear. Following the artificially distorted speech, a clear version of the speech (marked as C) or see a written presentation of the speech (marked as W) were presented to each feedback group and then the same distorted speech were repeated for both the feedback groups. The order of the speech sound was D-C-D or D-W-D. For the no-feedback group, there was no C or W session and this process was marked as D-D. The results showed that the D-C-D feedback group had the best performance among these three groups, but the results were easily to reach listeners' asymptotic performance. However, the results for the no-feedback group, although the sentence correction is lower but it kept increasing during the training session. All feedback and no-feedback groups demonstrated perceptual learning during the training session.

For the studies in this thesis, in order to reduce the risk of reaching listeners' asymptotic performance too early and get more information of the perceptual learning process according to the data from auditory perpetual learning studies, feedback was not provided for the participant to 'slow down' their learning process in purpose.

3.6 Experiment preparation

All tests were carried out in a sound-attenuating room, which is based in WMG (Warwick Manufacture Group). The sound-attenuating room is separated by a wall (including a window) to be two rooms. One is the test room, the other one is the monitor room. As seen in Fig. 3.3, during the test, the participants sit on a comfortable sofa in the sound-attenuating test room (room 2) and using computer two to do the hearing experiment. The experimenter (tester) is located at the monitor room (room 1), which is outside test room 2. Computer one was used to monitor and record the listeners' responses (see Fig.3.3, room 2).

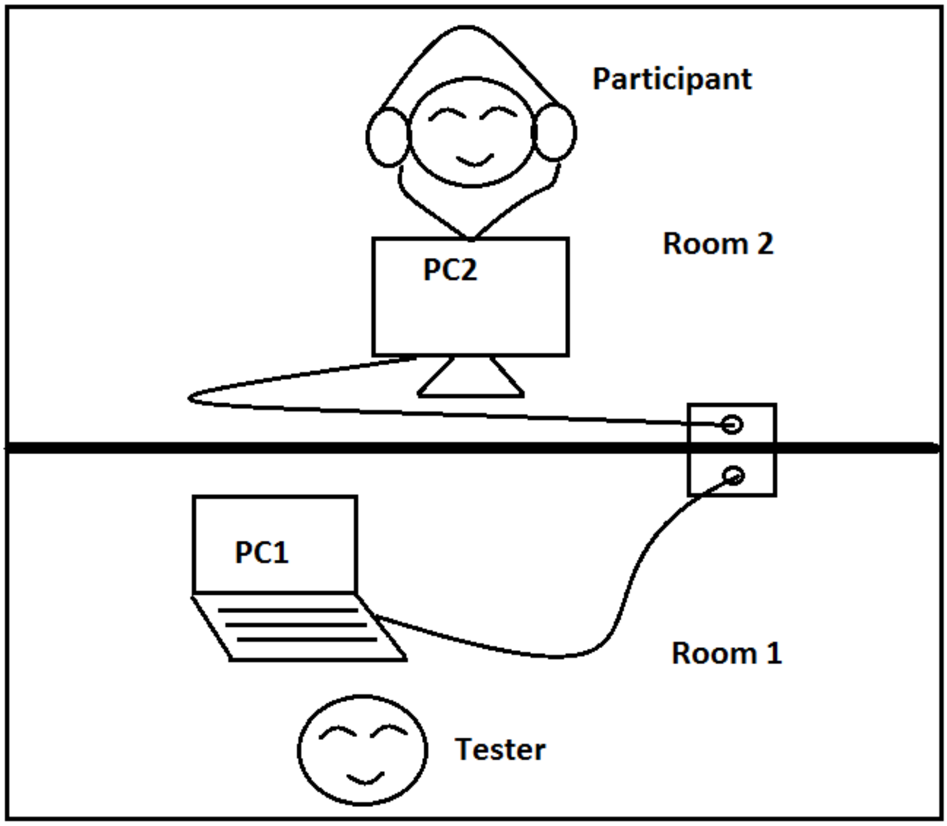


Fig. 3.3 Test position for tester and participant

A pure tone audiogram test was carried out to make sure the participant was qualified to take part (pure tone audiometry threshold ≤20 dB HL) in the study. After that, the instructions for the experimental tests were given to the participant to read to ensure they understand the experiment. After this, participants were allocated to either a control or a test group. All participants were informed that they could withdraw from the study at any time. Stimuli calibration was carried out before the main test took place. All equipment used in this experiment was checked to ensure it met the safety and calibration standards. The audiometer was checked within the specified calibration period. The headphones used with the audiometer were checked by subjective listening.

Noise over-exposure is the predominant risk in this study, and so sound levels were controlled via software and hardware. Damage to participants' hearing is extremely unlikely because the volume of the presented sound was kept below participant's maximum uncomfortable level. The risk of damage to hearing was also minimised by making sure the maximum sound pressure value guided by the Control of Noise at Work Regulations 2005 (http://www.hse.gov.uk/noise/regulations.htm). The aim of the Noise regulations is to ensure that workers' hearing is protected from excessive noise at their place of work and make sure their hearing is not damaged either by loss of sensitivity or tinnitus. In this research, we ensured that each experiment session (only one per day maximum) is below the lower exposure action value as stipulated by the Noise regulations – i.e. limit daily personal noise exposure to below 80 dB (A-weighted) and ensure that no peak sound pressure should be above 135 dB (C-weighted). The Noise regulations actually allow exposure up to 87 dB (A-weighted) and a peak sound pressure of 140 dB (C-weighted). Therefore, by ensuring the experiments are below the much lower (recall that dB is a logarithmic scale) exposure action value ensures there is no chance of hearing damage. The stimulus levels were calibrated by using the industrial standard IEC 711 acoustic coupler and a precision microphone. Then the maximum sound pressure levels from the PC were controlled to make sure the output from the software (MATLAB) was within exposure action value (65 dB SPL) prior to presentation to the participants. Table 3.2 provides details of the equipment used during the experiments. Pictures of the listed equipment are shown in Fig 3.4.

Table 3.2 Apparatus used in the experiments

Item	Quantity	Use	Photo
Sound Amplifier	2	One is used to make the stimuli audible to listeners during the test and the other one is used for calibration	A
Sound Calibrator	1	Used as a reference sound to calibrate the sound level from headphones	B
MAICO30 Clinical Audiometer	1	To make sure the listeners' hearing thresholds were within normal limits.	C
IEC 711 coupler	1	To calibrate the test stimuli and make sure the sounds are below 65 dB SPL	D
Circumaural Headphone Sennheiser HD 558	1	To make sure the participant can hear the test stimuli without any other background noise	E
PC	2	One is used for the presentation of the test sounds and the other one is used to monitor the experiment	N/A

Fig. 3.4 Pictures of the apparatus used in the experiments: A, Sound Amplifier; B, Sound Calibrator; C, MAICO30 Clinical Audiometer; D, IEC 711 coupler; E, Headphones

Chapter 4 No generalization from training on a SAM detection task to a SAM-rate discrimination task with different depths

4.1 Introduction

Information is carried through speech and sounds both in subtle amplitude and frequency variations over time. The listener's ability to detect fluctuation or modulation in sound contributes to their perceptual accuracy of many sounds, such as nonsense syllables, speech, etc. (Plomp, 1983; Rosen, 1992; Shannon *et al.,* 1995). Therefore, amplitude and frequency fluctuations or modulations in sounds are important information carriers for understanding speech. People with normal hearing can make use of these cues in sound to understand another speaker, but people with hearing-impairment have a reduced ability to detect these cues, particularly in challenging auditory environments. Any improvements in low-level perceptual tasks might help to alleviate some hearing difficulties in speech perception for people with hearing impairment (Lorenzi *et al*., 2000; Fu, 2002; Rocheron *et al*., 2002; Witton *et al.,* 2002). Although speech recognition using hearing prosthetic devices has improved significantly over the past few years, hearing aid and cochlear implant users still face major difficulties in noisy environments (Dorman & Wilson, 2004; Ricketts & Hornsby, 2005). Moore and Shannon (2009) suggested that rehabilitation and auditory training have the potential to optimise the performance of hearing-impaired users to help them get more benefit from their prosthetic devices.

In literature, auditory learning generalization occurs within different stimulus/tasks properties, such as frequency (Demany & Semal, 2002; Amitay *et al.* 2005), ear (Roth *et al.* 2003; Micheyl *et al.* 2006) and stimulus duration (Delhommeau *et al.* 2002). Previous studies showed that sufficient auditory training could help to improve humans' perceptual skills to detect and discriminate sounds, and lead to better performance to detect the changes in amplitude-modulated stimuli, especially for people with problems in detecting amplitude-modulated sounds (Hall and Grose, 1994; Irvine *et al*., 2000; Hawkey *et al.,* 2004). In a previous study, Fitzgerald & Wright (2011) demonstrated that sinusoidal amplitude modulation (SAM) detection and rate discrimination tasks involve different perceptual cues that the auditory system uses during decision-making. The SAM detection test focuses on the differences of amplitude-modulated depths between the target and the standard stimulus (the reference sound). The critical cue for the SAM-rate discrimination condition is the modulation rate differences between the target stimulus and the standard one. The SAM detection threshold is the minimum difference between the SAM depth of the target sound and standard SAM sound (reference sound in the test) and is usually measured by a logarithmic scale (in decibel: dB). The SAM-rate discrimination threshold is the minimum difference in the SAM rate required to discriminate between a higher SAM rate (target sound in test) and the standard slower SAM rate (reference sound in test). It is measured as a function of modulation rate, and the unit of measurement is Hertz (Hz).

The study from Fitzgerald and Wright (2011) demonstrated that AM-detection learning generalizes from trained rates (80 Hz) to untrained rates (30 and 150 Hz), but not to a new task (AM rate discrimination task). However, a 100% modulation depth for the SAM-rate discrimination task was used by Fitzgerald and Wright (2011). The modulation depth is how much the modulated signal varies around the original level of unmodulated signal, and it is one of the characteristics of the fluctuation of the signal. An optimal SAM modulation depth could help normal hearing people to achieve maximum scores during SAM detection and discrimination tests. Fitzgerald and Wright (2011) concluded that training on detection modulation depth would have no advantages to a rate-discrimination task with 100% modulation depth, as 100% modulation depth was already well above any minimum threshold to get the best performance for discrimination. Training on modulation detection would primarily require the presentation of stimuli with modulation depths substantially below 100%. For an amplitude modulated rate discrimination task, training with the full 100% modulation depth can determine the minimum discriminable rate change, and make the asymptotic performance for the sound easier to achieve (Grant *et al.* 1998). As authors state, this may explain why training with SAM detection did not generalize to SAM-rate discrimination in the experiment from Fitzgerald and Wright (2011). Based on the conclusion, this current study hypothesises that if significantly lower modulation depths are used for SAM-rate discrimination tasks, a generalization effect may occur from training with SAM-detection to SAM-rate discrimination.

In order to explore whether listeners' auditory perceptual learning abilities can be improved via the training of AM sounds, the present perceptual training study investigates the influence of multiple-session training on AM detection tasks in normal hearing people. It intended to investigate whether the lack of generalization found in Fitzgerald and Wright (2011) was due to the use of 100% AM depth in the rate-discrimination task. The modulation depths in this study (full modulation depth 100%, mid-depth 70% and low-depth 40%) were based on the same paradigm as the one used in the modulation rate detection and discrimination study by Grant *et al.* (1998). In their study, they used three different modulation rates (80Hz, 160Hz and 320 Hz, which are fundamental modulation frequencies covering male, female and children’s voices) and three different ranges of modulation depth (full 100%; mid 70%‑80%, and low: 40%‑60%) to investigate the modulation detection and rate discrimination for both normal hearing and hearing impaired people. The modulation rate of 80 Hz was the same as the modulation rate in the study of Fitzgerald and Wright (2011). The modulation depth 100%, 70% and 40% was selected as the full, mid and low modulation depth for the modulation rate discrimination task with 80 Hz. Their results showed that both modulation detection and rate discrimination threshold increased with modulation rate, this was true for both hearing impaired and normal hearing people. For the modulation rate discrimination task, the threshold decreased with increased modulation rate. Most hearing-impaired people could not discriminate any change in rate at the fastest/highest rate (320Hz) with less than 100% modulation depth.

In my experiment, I used the same modulation depths (100%, 70%, 40%) as Grant *et al.* (1998), the modulation rate (80 Hz) and carrier frequency range (3-4kHz) was the same as Fitzgerald and Wright (2011). Both SAM detection and SAM-rate discrimination tasks were used in this present experimental design. Listeners’ performances on picking up target SAM-detection and SAM–rate discrimination stimuli across test sessions were recorded. The main purpose of the present study is to investigate whether training with AM detection tasks generalized to AM rate discrimination tasks by using lower AM depths, such as 70% and 40%. Apart from this main purpose, there were two additional aims for this study. Details of all the three experiment objectives are listed below.

- **Objective 1:** Compare SAM detection thresholds between the pre-training and post-training tests of SAM detection training to examine whether there is an improvement after the training session.
- **Objective 2:** Compare SAM-rate discrimination thresholds between the pre-training and post-training SAM-rate discrimination tests to see whether there is a generalization effect from SAM detection training to SAM-rate discrimination with different fixed modulation depths, such as 100%, 70%, and 40%.
- **Objective 3:** Compare SAM-rate discrimination thresholds from three different modulation depths (100%, 70%, and 40%) to investigate which SAM-rate discrimination modulation depth attains the largest improvement after SAM detection training.

4.2 Test method

4.2.1 Participants

Twenty volunteers with normal hearing (13 males and 7 females) participated in this experiment. All of the participants had no prior experience participating in psychoacoustics experiments, and their pure tone thresholds were less than 20 dB HL. The age range was from 18 to 36 years old (with a mean age of 27 years). The participants were volunteers recruited from the student and staff population of the

University of Warwick (for further details of participant requirements, see Chapter 3, section 3.2).

4.2.2 Design

Twenty participants were randomly divided into a training group (n=10) and control group (n=10). For the main test, both groups were required to attend a pre-test and post-test session. These two sessions lasted approximately 2 hours (one hour for each session). Each participant was presented with a series of band-limited noises at a spectrum level of 40dB sound pressure level. The pre- and post-test session included one SAM-detection condition and three SAM-rate-discrimination conditions. As implemented by Fitzgerald and Wright (2011), the order of the four conditions was randomised across test participants but was the same order in the pre- and post-tests for each individual participant. A three interval, three alternative forced choice (3I-3AFC) adaptive procedure was used to determine the thresholds for SAM detection and SAM-rate discrimination. The modulation depth and modulation rate were varied for SAM detection and SAM-rate-discrimination tasks respectively, targeting a 79.4% correct performance on the psychometric curve (Levitt, 1971).

Five thresholds were obtained for each condition. Participants in the training group were required to attend 7 consecutive daily (except weekends) training sessions on SAM detection tasks between the pre- and post-session. Twelve SAM detection thresholds were obtained in each training session (training involved exactly the same task as testing). All experimental sessions were carried out within a single-walled soundproofed room. Sound levels for SAM detection and SAM-rate discrimination stimuli were calibrated using an IEC 711 acoustic coupler to 65dB SPL (or at a spectrum level of 40dB sound pressure level). The experiment was approved by the biomedical and scientific research ethics committee (BSREC) of the University of Warwick (see Appendix 3). Below are details of the test procedures. Fig. 4.1 shows the flowchart for SAM detection and SAM-rate discrimination tests.

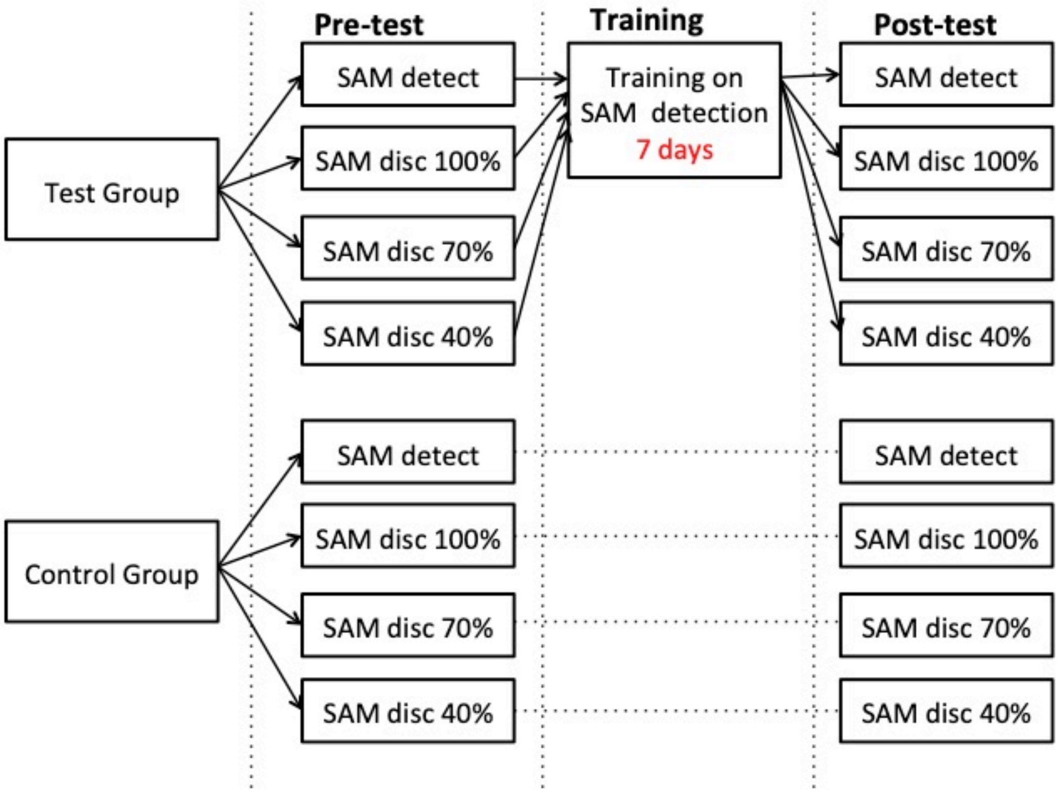


Fig. 4.1 Flowchart for Sinusoid Amplitude Modulated (SAM) detection (labelled 'SAM detect') and SAM-rate –discrimination (labelled 'SAM disc') tests

4.2.3 Test stimuli

For the SAM detection test, the target sound was a 3-4 kHz band-pass noise carrier modulated at 80 Hz (as used by Fitzgerald and Wright, 2011; See Fig. 4.2), while the reference sound was un-modulated. Stimulus duration was 400 ms and inter stimulus interval was 600 ms. Under these test conditions, the modulation detection threshold was determined with an adaptive tracking procedure. There were three intervals, which included two reference signals and one target, randomly presented (the order of the three intervals presented randomly). The target signal was presented randomly during the test procedure. The listener was instructed to decide which interval contained the target amplitude modulated stimuli. The starting modulation depth (m) was 100% modulation and the modulation index in decibels was 20Log10 (m). The initial step size was 4dB and then reduced to 2dB after three test reversals. The SAM detection threshold was defined as the mean of the last 10 reversals in the adaptive tracking procedure (60 trials were displayed for each threshold, only the mean of the last 10 reversals was used to calculate the threshold).

For the SAM-rate-discrimination test, a 3-4 kHz band-pass carrier modulated at 80 Hz with three depths (high: 100%, mid: 70% and low: 40%) was used as the reference sound, and the target sound was the same carrier with a higher modulation rate. During this test, the modulation rate of target sound was measured to determine the modulation detection threshold by using a 3I-3AFC adaptive tracking procedure. Participants were required to give a response about which interval was different from the other two. The initial step size between standard and target stimuli was 15 Hz, then decreased to 3 Hz after the third interval, and 1 Hz thereafter, until the threshold was reached. Fig. 4.2 shows sound waveforms about 3-4 kHz bandpass noise, and 3-4 kHz bandpass noise with 80 Hz sinusoidal amplitude.

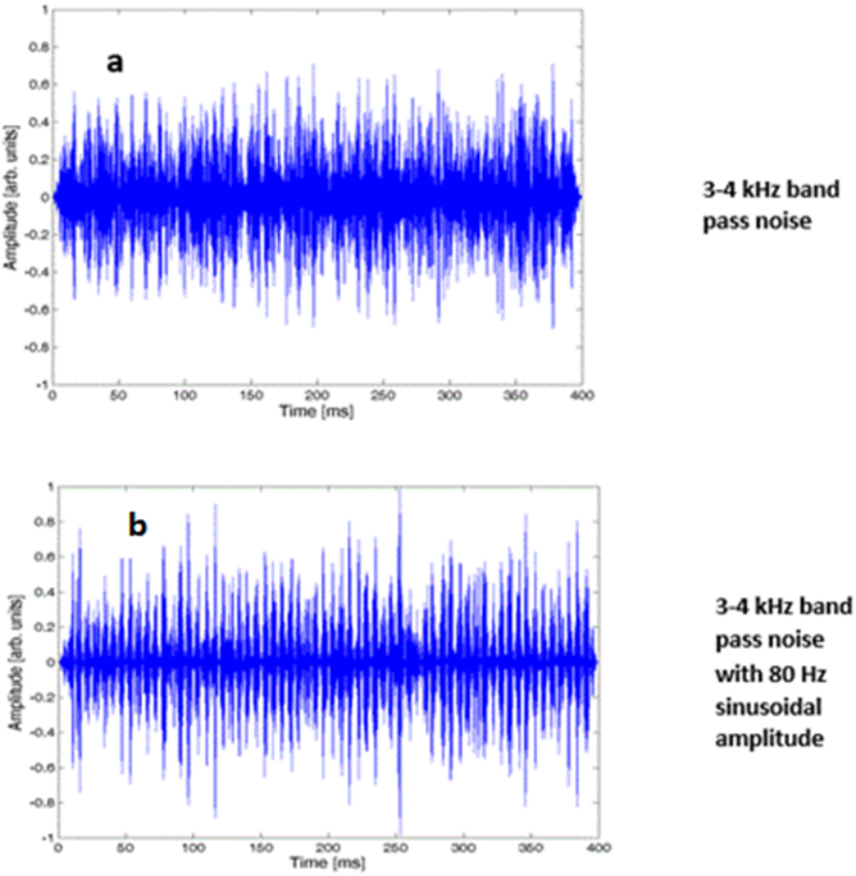


Fig. 4.2 Sound waveforms for a) 3-4 kHz bandpass noise, b) 3-4 kHz bandpass noise with 80 Hz sinusoidal amplitude.

4.2.4 Data analysis

Data from those participants whose pre-test thresholds were above two standard deviations from the mean of all the pre-test thresholds in that condition were removed. However, all participants produced pre-test threshold values within two standard deviations of the mean in this experiment, so no data was removed (i.e. identified as outliers) from prior to the data analysis. An analysis of covariance (ANCOVA) with pre-test thresholds as the covariate was used to compare the test results between the trained and control group. Two groups (trained vs control) × two time (pre vs post) or two groups (trained vs control) × two time (pre vs post) × three depth (100% vs 70% vs 40%) mixed analysis of variance (ANOVAs) with repeated measures on time factors, and t-tests were also used to analyse pre- and post-test results. Day to day individual SAM Detection training performance was analysed using a one-way repeated measures ANOVA.

4.3 Test results

4.3.1 SAM detection tasks

As can be seen in Fig. 4.3a, although the mean threshold for trained listeners in the pre-test SAM detection condition (M= -6.84 dB, SD= 0.59 dB) was higher than that of the control listeners (M= -8.25 dB, SD=0.59 dB), the mean threshold for the trained listeners in the post-test of SAM detection condition (M= -10.01dB, SD= 0.74dB) was lower than the mean threshold for the control participants in their post-test SAM-detection (M= -9.42 dB, SD= 0.63dB). The main effect of comparing the performance of two groups showed that there was no overall performance difference between training and control groups (ANOVA: group: $F_{1,18} = 0.22$; $p > 0.05$). Both the two-way ANOVA and the ANCOVA test indicated that there was an overall learning difference between pre- and post-test results for the trained and control group (ANOVA: time: $F_{1,18} = 100.73$; $p < 0.005$; group × time interaction, $F_{1,18} = 21.33$; $p < 0.05$; ANCOVA: $F_{1,17} = 18.51$; $p < 0.05$).

Paired t-tests were conducted with threshold values from both the SAM detection trained and SAM detection control groups. For the control group, there was a statistically significant decrease in thresholds from the pre-test SAM detection thresholds (M = - 8.25 dB, SD = 1.87) to post-test SAM detection thresholds [M = -

9.42 dB, SD = 1.99), t (9) = 4.34, *p* = 0.002]. For the trained group, there was also a statistically significant decrease in thresholds from the pre-test SAM detection thresholds (M = - 6.84 dB, SD = 1.88) to post-test SAM detection thresholds [M= - 10.01 dB, SD = 2.34, t (9) = 9.38, *p* < 0.0005]. An independent samples t-test was carried out comparing the thresholds improvement between control and trained group. There was a statistically significant difference in improvement between the control group (M = 1.17 dB, SD = 0.85) and the trained group [M = 3.17 dB, SD = 1.07, t (18) = - 4.62, *p* < 0.0005).

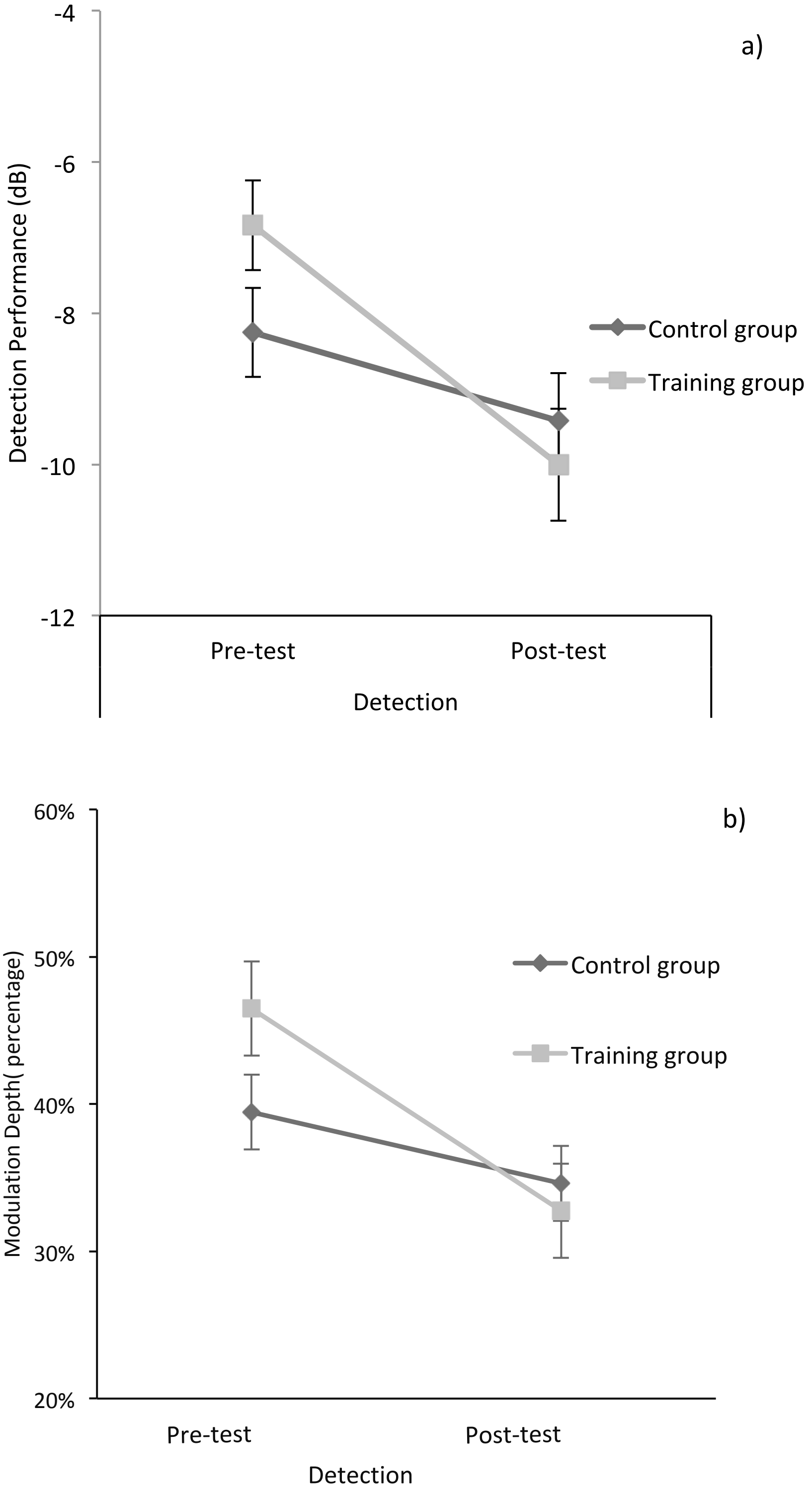


Fig. 4.3 Mean pre-test and post-test SAM detection thresholds for the training (n=10), and control group (n=10). Error bars indicate ± one standard error of the mean; a) SAM detection threshold unit in dB; b) SAM detection threshold unit in percentage (%).

4.3.2 Individual day to day SAM detection training performance

Fig. 4.4 shows individual trained participants' SAM detection performance across each pre-test, post-test and training session. A one-way repeated ANOVA showed that there was an overall performance difference across the whole test [$F_{8,\ 72} = 17.76$, $p < 0.001$]. However, as shown in Fig. 4.4, there was considerable variability in improvement during the test sessions regarding participants' SAM detection performance. Subsequent pairwise comparisons between consecutive days (Pre vs day1, day1 vs day2, and so on), demonstrated that, across participants, a significant change in threshold occurred only between pre and day1 (Mean differences = 1.89, $p = 0.002$). However, it is clear from visually analysing individual results (Fig 4.4) that while some participants did not improve and showed fluctuated performance (e.g. T1,T9) or plateaued after the first session (e.g. T2, T3, T4, T7, T8), others continued to improve (e.g. T6, T10), albeit at a slower rate.

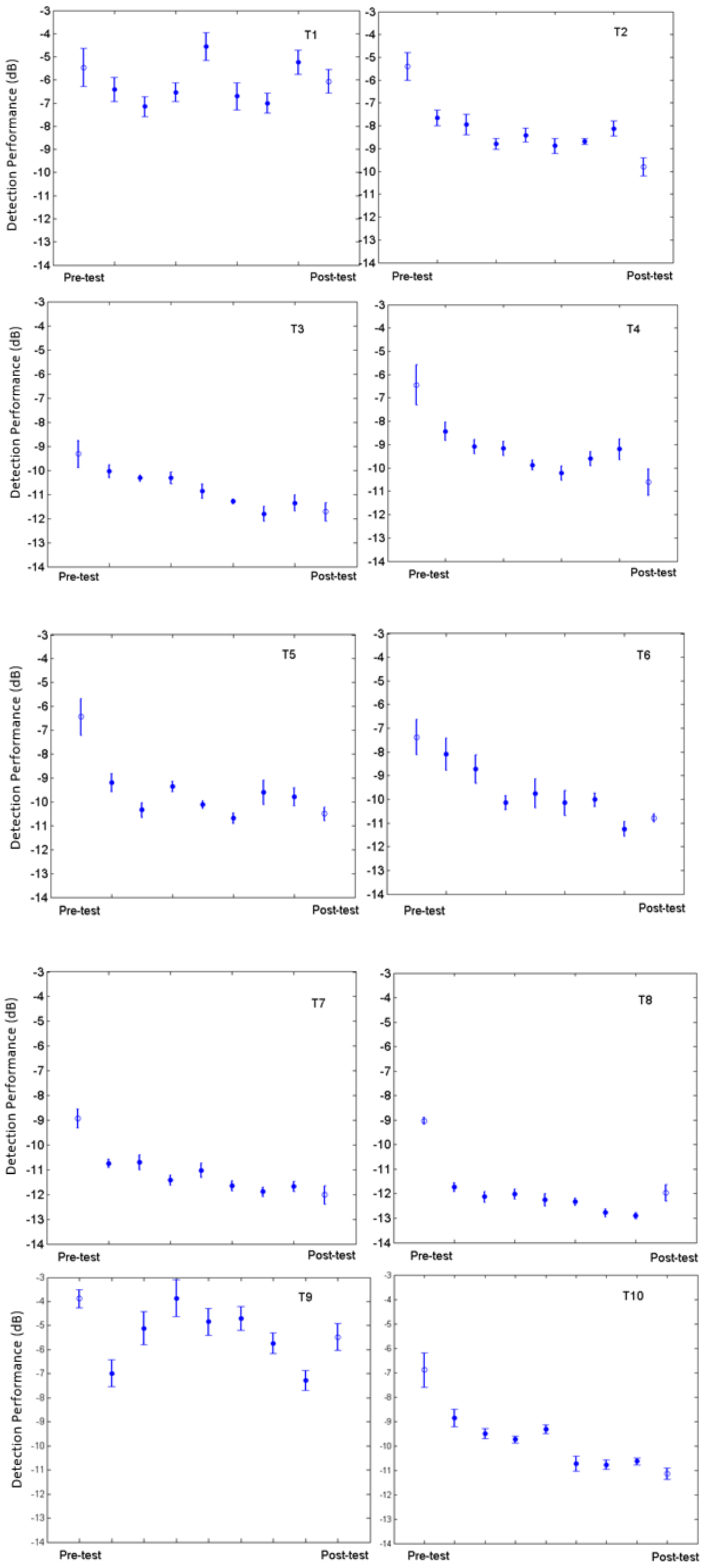


Fig. 4.4 Mean pre-test and post-test SAM detection thresholds on the trained condition (open circles) and during the training phase (open squares) are shown for all ten trained participants. Error bars indicate ± one standard error of the mean within a given listener.

4.3.3 SAM-rate Discrimination tasks

As can be seen in Fig. 4.5, both ANCOVA and mixed ANOVA showed the main effect comparing the pre- and post-test performance of the two groups among three modulation depths was not significant (group: ANOVA: $F_{1,\ 18} = 0.23$; $p > 0.05$; ANCOVA: $F_{1,\ 15} = 0.68$; $p > 0.05$). However, there was a significant difference between SAM-rate discrimination pre- and post-training sessions (time, $F_{1,\ 18} = 49.00$; $p < 0.0005$), but this effect did not differ between the two groups (group × time interaction, $F_{1,\ 18} < 1$). Discrimination performance differed for the three modulation depths (depth, $F_{2,\ 36} = 53.37$; $p < 0.05$), but this effect did not differ between the trained and control groups (group × depth, $F_{2,\ 36} = 0.79$; $p > 0.05$; group × time × depth interaction, $F_{2,\ 36} = 0.17$, $p > 0.05$). The mean SAM-rate discrimination thresholds were 20.14, 22.47, and 30.68 Hz for modulation depths of 100 %, 70 % and 40 %, respectively. The following pairwise test demonstrated that the three values were all significantly different between each other (100 % vs 70 %, 100 % vs 40 %, 70 % vs 40 %, all $p < 0.05$). It suggested that among all the three different modulation depths (100%, 70%, and 40%) for SAM-rate discrimination conditions, participants had the highest SAM-rate discrimination threshold with modulation depth of 40%.

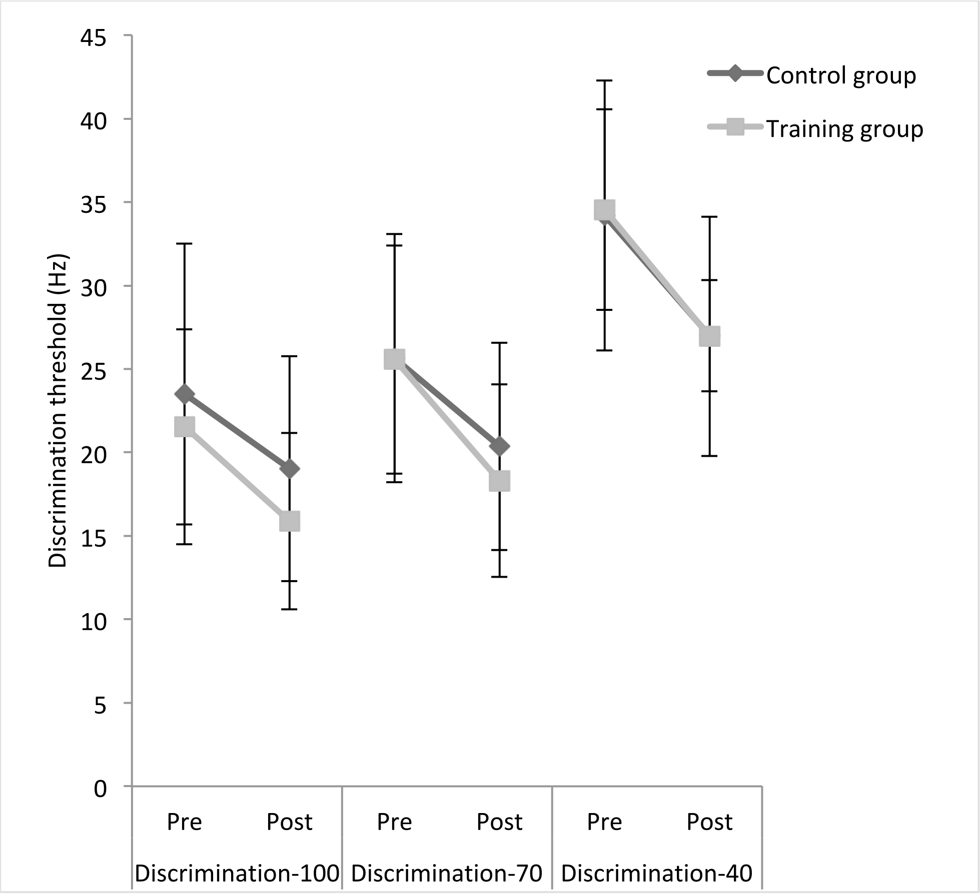


Fig. 4.5 Mean pre-test and post-test SAM-rate discrimination thresholds for the trained (n=10) and control group (n=10) under three conditions (100%, 70% and 40%). Error bars indicate ± one standard error of the mean.

A mixed ANOVA was carried out to test the SAM-rate discrimination threshold improvement from pre- to post-test sessions (within factor: two times) between the trained and control group (between factor: two groups) among three modulation depths (within factor: three depth). The results showed that the main effect of comparing improvements between these two types of intervention groups was not significant (group: $F_{1, 18} = 0.44$, $p > 0.05$). These results suggested that the trained group had similar improvements to the control one. There was no significant difference among SAM-rate discrimination thresholds improvement with three modulation depths (depth: $F_{2, 36} = 1.48$, $p > 0.05$); the same was true for the interaction between group and depths (group × depth: $F_{2, 36} = 0.17$, $p > 0.05$).

4.4 Discussion

Comparing the results from pre-and post- SAM detection thresholds and SAM-rate discrimination thresholds revealed that both trained and control groups demonstrated significant improvements. Thus, learning effects were observed for the SAM-detection and SAM-rate discrimination tests. Even though the initial pre-test session SAM-detection performance from the trained group was poorer than the control group there was significantly more improvement for the trained than the control group on SAM detection. These findings of the SAM-detection results were the same as those reported by Fitzgerald and Wright (2011); that is, training improves abilities in a SAM detection task (listeners' performance improved with SAM detection training). While Fitzgerald and Wright (2011) did not provide specific values of the SAM detection data, according to the SAM detection performance from Figure 1 in their paper, visual inspection suggests their trained group improved roughly 3dB (Pre: M= -9dB; Post: M= -12dB) and control group improved around 1dB (Pre: M=-9.5dB; Post: M=-10.5dB). These are relatively similar improvements to the SAM detection results in current study (trained group: Pre: M= -6.84 dB, Post: M= -10.01dB, Improved: M=3.17 dB, Control group: Pre: M= -8.25 dB, Post: M= -9.42 dB, Improved: M = 1.17 dB). However, when comparing the mean thresholds of the SAM-rate discrimination tasks in current study, no significant difference was found between trained and control groups. Therefore, the study does not demonstrate a generalization effect from training on a SAM-detection task to a SAM-rate discrimination task with any of the three modulation depths (100%, 70%, and 40%). The mean threshold of the SAM rate discrimination with 100% depth from Fitzgerald and Wright (2011) also demonstrated no significant differences between their trained and control groups. Comparing the thresholds improvement of SAM rate discrimination with 100% depth between Fitzgerald and Wright (2011) and this study, visual inspection from their Figure 1 indicated their trained group improved approximately 5.5Hz (Pre: M= 19.5Hz; Post: M= 14Hz) and control group improved around 3.5Hz (Pre: M=19.5Hz; Post: M=16Hz). The current study shows similar improvements in the SAM rate discrimination (100% depth) task (trained group: Pre: M= 21.5Hz, Post: M= 15.9Hz, Improved: M = 5.6Hz, Control group: Pre: M= 23.5Hz, Post: M= 19Hz, Improved: M = 4.5 dB). In this study, the most difficult condition (40% modulation depth) had the highest SAM-rate discrimination threshold among the SAM-rate discrimination thresholds of the other two conditions (100% and 70% modulation depths), regardless of training. As for individual learning ranges, participants' performances varied across the training session, T06 demonstrated the largest improvement (3.15 dB) from day 1 to day 7 among all of the trained participants. Details of the test results will be summarized and explained in the following paragraphs.

4.4.1 No generalization from SAM detection to SAM-rate discrimination

Fitzgerald and Wright (2011) demonstrated that sensitivity training on SAM detection task had no advantages to SAM rate discrimination task with the 100 % modulation depth. In the present study it was hypothesised that training on SAM modulation detection task may result in improvements in lower modulation depths (40% or 70%) rather than with the high (100%) modulation depth. However, results from this study showed that regardless of training, 40% modulation depths had the highest SAM-rate discrimination thresholds (poorest performance) compared to both modulation depths of 100% and 70%. This agreed with the rate discrimination performance in Grant *et al.* (1998), according to their results (Figure 4 in their paper) by visual inspection, it showed that the mean SAM rate discrimination threshold at 80 Hz for the low modulation depth (40%) was higher than for both the middle (70%) and the high (100%) modulation depths. The mean SAM rate discrimination thresholds at 80 Hz for the middle (70%) and high (100%) modulation depths were similar. Unfortunately, Grant *et al.* (1998) did not provide any further pairwise statistics analyses for SAM rate-discrimination performance at 80Hz across three different modulation depths (40%, 70% and 100%) and hence it is only possible to report the rate discrimination performance visually here.

Results from this study also showed that detection training made no difference in the SAM-rate discrimination performance improvement with lower modulation depths (40% and 70%). It did not agree with the original hypothesis. Performance improvement (from pre- to post-test) of a SAM-rate discrimination task was almost the same for the three modulation depths (the mean of SAM-rate threshold improvement for each modulation depth was: 100%: 5.66Hz, 70%: 7.27 Hz; 40%: 7.58 Hz). Despite adding a SAM detection training session for the trained group, this group demonstrated similar SAM-rate performance improvement as the control group. Therefore, no transfer learning was found from SAM detection task to SAM-rate discrimination task, regardless of modulation depth. These results suggested that rate discrimination tasks with different modulation depths had similar rate discrimination threshold changes after SAM detection training. In spite of this, it was supposed that a certain amount of AM stimuli training could lead to significant threshold differences for SAM-rate discrimination tasks with different modulation depths, but as the results showed, in this study, AM sound detection auditory training did not lead to improvement (learned values from pre-test to post-test) differences for SAM-rate discrimination tasks with different modulation depths.

A potential reason why training with SAM detection did not generalize to SAM-rate discrimination was that the auditory system processes these two tasks differently. Millward *et al*. (2011) presented evidence to suggest that the generalisation effect from trained auditory tasks to other tasks is more likely if both share a common stimulus dimension, i.e. the same masking noise or target stimulus. Otherwise, they demonstrated an opposite effect to the desired synergistic generalisation effect, where training in one task actually suppressed or reduced performance in another. This was more likely to occur if the two tasks did not share a common stimulus dimension. In the present study, although target sound in the SAM rate-discrimination test used the same carrier as that used in the SAM-detection tasks, training with SAM detection did not generalize to the SAM-rate discrimination task. The SAM detection task neither improved nor suppressed the performance of SAM-rate discrimination. It was probably that sharing part of the stimuli parameter (carrier frequency) between the SAM detection and SAM-rate discrimination tasks had no influence on the auditory system to process these two tasks.

As described in section 4.1, the stimuli feature of interest for the SAM detection task and SAM-rate discrimination task, namely modulation depth versus modulation frequency, which were different features between these two tasks. It could be argued that the lack of generalization from SAM detection to SAM-rate discrimination training might arise as a result of the auditory system processing these two tasks separately. Different cues could be used to give the best performance on these two tasks, and thus training could affect the two tasks differently. Therefore, further training with SAM detection did not add any benefit for the post-test performance of the SAM-rate discrimination task.

4.4.2 Perceptual learning for SAM detection and SAM-rate discrimination

Both the pre- and post-test sessions in this present study included mixed SAM detection and SAM-rate discrimination tasks. Demany (1985) demonstrated that listeners required practice to achieve their asymptotic detection or discrimination threshold. Research shows that perceptual learning improvements may be obtained in the first few trials (Gilbert, 1994; Atienza, 2002; Moore, *et al*., 2003; Ben-David, *et al*., 2010). The preliminary performance in the pre-test, for both the SAM detection and SAM-rate discrimination task, could help listeners to improve their performance of these two tasks in the post-test session. In this experiment, fast perceptual learning had already occurred in the pre-test session and was enhanced by the post-test tasks. As a consequence, the trained and control group revealed significant improvement from their pre- to post- SAM detection thresholds and SAM-rate discrimination thresholds.

However, Wright *et al.* (2010) and Szpiro *et al.* (2014) demonstrated that mixed perceptual learning tasks led to more fine-tuning based on different task cues, such as modulation rate and depth. Fitzgerald and Wright (2011) found that no generalization occurred from training with a SAM detection task to a SAM rate discrimination task. In order to explain these findings they suggested that SAM detection and rate discrimination tasks use different task cues during auditory decision-making. Therefore, listeners probably needed more fine-tuning between SAM detection and rate discrimination tasks (the pre-test and post-test sessions in this study). The fine-tuning process might affect perceptual learning between SAM detection and rate discrimination tasks, but we could not determine the amount of influence induced by this experimental design. To confirm this point, further studies using SAM detection and SAM-rate discrimination tests separately across sessions need to be conducted.

4.4.3 Overtraining

Based on the tests results, overtraining occurred. According to individual day-to-day SAM detection training performance, the largest learning gradient occurred between Pre-test and Day 1 (due to the significant effect found from pairwise comparisons). However, following Day 1, individual learning performance varied, with some participants continuing to show substantial learning (e.g. T6, T10), while others showed fluctuated learning performance (e.g. T1, T9) or an asymptotic learning effect (e.g. T2, T3, T4, T7, T8). The majority of listeners had already achieved their asymptotic performance for the SAM detection tasks after day one's training sessions (see results section 4.3.3).

From perceptual learning in visual domain, there is an effect named perceptual deterioration, which caused by overtraining during perceptual learning. Overtraining in perceptual task can generate an improvement in performance at the beginning and follow gradual decline afterwards (Censor, Karni &Sigi, 2006; Mednick, Nakayama & Cantero, 2002; Ofen, Moran & Sagi, 2007). Perceptual deterioration occurs due to the limited capacity of early visual area, when the visual area become saturated with information during overtraining, it will be hard to consolidate newly acquired changes (Mednick, Nakyama & Stickgold, 2003; Ashley &Pearson, 2012).

Early perceptual training trials were crucial for perceptual training studies, especially for long-term training ones. Once participants' perceptual training performance achieved a certain level, additional training added no significant improvement benefits to listeners' performances, they may affect some other aspects of perceptual learning, for example generalization or retention (Wright & Sabin, 2007). In this study, no significant SAM detection performance differences were found after the training session on day one. This result suggested that the majority of participants performance was improving, but only at a relatively slow rate, so there was limited improvement after the following days training tests. In this case, overtraining was found and as a consequence, there was a slower perceptual learning gradient (day to day improvement rate) after they reached their asymptotic performance. Therefore, the influences of daily learning limits should also be taken into consideration in auditory training design for future study design. Wright and Sabin (2007) also demonstrated that if learning on two tasks had modified different circuitries at physiological level, training on one of the task would inherent some features and then made that circuit less amenable to change. Therefore, overtraining would inhibit the learning to be transferred to another task. Therefore, it is assumed that no generalization from training with SAM detection task to SAM rate discrimination task in this study was due to SAM detection and SAM rate discrimination were two different tasks and modified at different levels.

4.4.4 Further thoughts

During seven days of training session with SAM detection tasks, listeners might get used to detecting the modulation depths differences for the SAM detection task. Thus, it can be easier for them to continue to do the same task during the post-test session, but difficult to then shift their focus to a different task (SAM-rate discrimination) in the post-test session. For this study, the training session might have enhanced listeners' perceptual learning on the SAM detection task, but affected their performance on the SAM-rate discrimination task in the first few trials of the post-test session.

Training on a range of different auditory stimuli leads to a greater transfer learning effect (Halliday *et al*., 2012). This transfer learning effect was possibly due to improved attention and/or working memory for different stimuli tasks during the test sessions. Many researchers have emphasized the importance of attention in perceptual learning (Van Wassenhove & Nagarajan, 2007; Yotsumoto & Watanabe, 2008; Paffen, *et al*. 2008; Ahissar *et al*., 2001, 2009). For auditory sound discrimination or detection tasks, attention played an important part in processing and distinguishing complex acoustic stimuli (Näätanen, 1990; Kiehl, *et al*., 2001; Petkov, *et al*., 2004). However, the training task in this study was specific to the SAM detection task. Further research can be carried out to explore whether better generalization occurs when people are trained on more than one type of complex auditory stimuli, such as non-speech and speech sound mixed training.

4.5 Conclusion

In summary, it is interesting to note that although SAM detection and SAM rate discrimination tasks shared similar stimuli features, SAM detection training did not transfer to SAM rate discrimination, regardless of modulation depth. The results indicated that stimulus learning is not sufficient to improve perceptual learning between different SAM tasks. The SAM detection and SAM rate discrimination were two different tasks and modified at different levels. This result may be due to overtraining in the experiment design, a lack of mixed stimuli training, listeners' working memory, and/ or attention. These factors should be kept in mind for any subsequent studies. At this stage, it is suggested that further research should be carried out to explore whether better generalization occurs when people are trained on more complex, ecologically valid stimuli, such as speech sounds, non-speech and speech sound together.

4.6 Summary

This study confirmed that training improves abilities in a SAM detection task, corroborating the results found by Fitzgerald and Wright (2011). However, the results also extended the work of Fitzgerald and Wright (2011), as there was no evidence that a generalization effect occurred from training SAM detection to SAM-rate discrimination, this was true for all of the three modulation depths tested. Moreover, listeners in the trained or control group demonstrated a similar performance improvement in the SAM-rate discrimination task with the three modulation depths listeners from their pre- to post-test sessions. Based on the results of this study, it is suggested that further work could be carried out to explore whether generalization occurs with more complex stimuli, such as speech sounds or speech and non- speech sounds together.

Chapter 5 Auditory training of nonsense stimuli recognition with fixed and random babble noise

5.1 Introduction

Speech is an essential tool for communication, even when it is degraded or masked by other competing sounds in daily life. As evidenced in a study of performance in speech intelligibility tasks, listeners can obtain near perfect speech recognition performance (> 90%) with degraded speech tasks (consonants, vowels, and words in sentences with speech spectrum information reduced) with 8 to 10 hours training over two to three training sessions (Shannon *et al.,* 1995). In clinical studies most researchers have attempted to train people to understand speech material better in a quiet environment. Typically, auditory perceptual learning rehabilitation programmes present speech with no background noise or other competing environmental factors. Auditory training within a quiet environment may help listeners' focus attention on the detection of subtle listening cues, such as pitch, stress, intonation and so on, but it is not certain whether these auditory cues can be made use of in noisy environments (Fu & Nogaki, 2005). Any additional outcomes from speech in noise auditory training research may contribute to devising better training methods for people with hearing impairments.

In the literature, few studies incorporate training with background noise. Most speech in noise perceptual training studies focus on changing the tasks which involves using several different sounds with the same background noise (Burk *et al*, 2006; Yund & Woods, 2010). These studies demonstrate that any learning that occurs from training is specific to certain trained speech materials, and parameters of background noise (such as the signal to noise level; noise type: white noise or speech shaped noise or babble noise). A novel element of the experiments laid out in this chapter is that they focus on training subjects to listen to multi-talker babble noise masking with fixed and random background noise. For the fixed babble noise background condition, the exact same section of babble noise was selected as the test background noise on every single trial (but different sections of babble noise were selected for different participants); while for the random babble noise condition, a different section of babble noise was used as the test background noise on every single trial.

Langhans and Kohlrausch (1992) reported that the detection thresholds for signals presented randomly in noise (in different temporal positions from the mask stimuli) is significantly higher than for signals displayed in fixed in noise (in the same temporal stimuli position from the mask stimuli). In their psychoacoustic perceptual training study a flat power spectrum mask stimulus was used. The spectrum of the stimuli was between 20 Hz and 5 kHz presenting with a duration of 300 ms. Cutler *et al.* (2004) reported that multi-talker babble noise is a form of noise that can be used in speech perception and recognition studies due to its high level of ecological validity. Some hearing studies used fixed babble noise as the background noise, while the others used random babble noise (Wilson, 2003; Killion *et al.*, 2004; Engen & Bradlow, 2007). Felty *et al.* (2009) used one short session test to compare word performance with fixed and random babble background noises. They demonstrated that the listeners obtained better word recognition performance with fixed babble noise than with random babble noise. However, no research has investigated the effect on perceptual learning ability with changing (random patterns) or what effect keeping the target sound's background noise constant (fixed) has across training sessions. From perceptual learning in the visual domain, people can improve their detection performance by learning to ignore (visual) "noise". Once participants have learnt to ignore fixed visual noise and can successfully detect targets, this skill transfers to new random visual noise (Schubö *et al.,* 2001). Motivated by the evidence of visual research, the studies outlined in this chapter will test whether the same pattern occurs in the hearing domain.

In daily life, we normally experience more communication conditions with random background noise than with fixed ones. If we can discover which training method (fixed or random babble) is better to help people to be more sensitive to speech information in random background conditions, then we can apply that noise training method in clinical use, and support hearing-impaired people to improve their speech comprehensions in noisy environments. The studies in this chapter have two aims. One aim is to explore if it is possible to improve listening ability by training listeners to recognise the stimulus sound from fixed or random background noise across time (a learning effect, investigated in VCV study one and two). The second aim is to investigate if this learning effect, from training listeners to detect target sounds against fixed babble background noise, will generalize to random background environments (generalization, investigated in VCV study two). Two auditory training experiments in nonsense stimuli recognition with fixed and random babble noise were conducted in this chapter to fulfil these two aims.

5.2 VCV study one (SNR-24dB)

As described in the introduction of this chapter, this auditory learning study was motived by evidence from visual perceptual learning study (Schubö *et al.,* 2001). The present study evaluates whether perceptual learning findings from vision can also apply in the hearing domain. It is hypothesised that the ability to detect VCV stimuli may be improved more by training listeners in VCV stimuli against a fixed babble noise rather than training them in VCV stimuli against a random babble noise. There were two objectives for VCV study one (listed below).

Objectives:

- Compare pre-training and post-training test results with random babble noise to explore a training effect.
- Investigate whether learning with fixed babble noise produces better identification performance against a random-noise background than learning with random noise.

5.2.1 Test methods

5.2.1.1 Participants

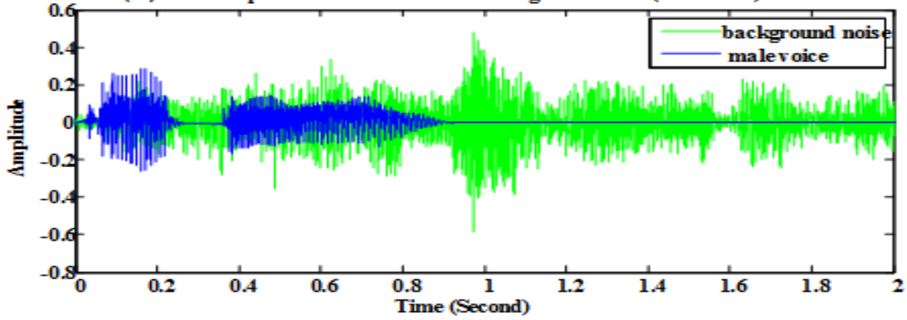
Twenty normal-hearing native English speakers (aged from 18 to 40) participated in this experiment. They were randomly assigned to a fixed or random babble noise training group. All of the participants had no prior experience participating in psychoacoustic experiments, and their pure-tone thresholds were less than 20 dB HL (see details of the participants' requirements in chapter 3). The participants were all volunteers recruited from the student and staff population of The University of Warwick.

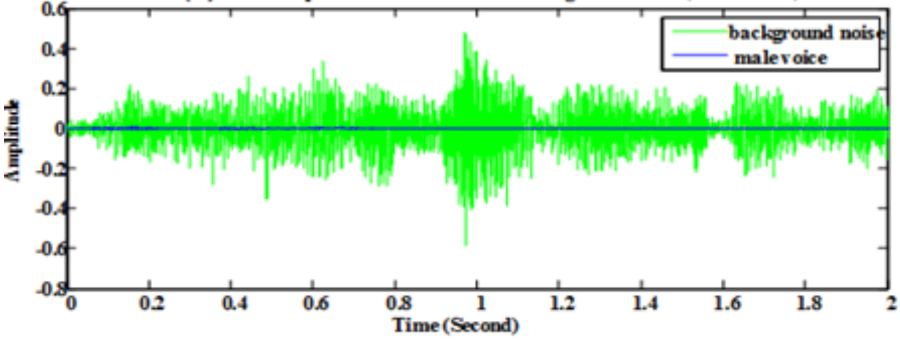
5.2.1.2 Stimuli

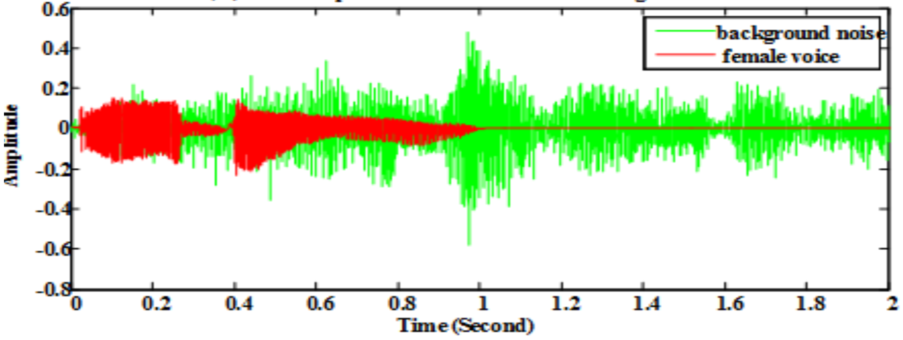
In order to find the required test parameters (SNR level, carrier vowels, and target consonants) for the VCV experiment design, two pilot studies were carried out (see Appendix 5). The vowel /I/ with eight consonants /b/, /d/, /f/, /k/, /m/, /n/, /t/, /z/ were initially used in pilot study 1, at a SNR level of -24dB. The results indicated that the consonants /t/ and /z/ started out around the level of asymptotic performance (/t/: 77.92%; /z/: 63.75%) and remained there (/t/: 66.67%; /z/: 72.50%), and all others started out at (or actually below) chance (around 12.5%) and never improved. In pilot study 2, /t/ and /z/ were discarded and replaced by /g/ and /p/ to avoid the asymptotic performance of those consonants, while the vowel /I/ was replaced by /a/ and the SNR was increased to -18dB to avoid the risk of a floor effect for the consonants. The results of pilot study 2 indicated that all consonants were at a high risk of reaching asymptotic performance, and therefore the SNR level was further reduced for the main test.

Following the results from the two pilot studies, the main test for VCV perceptual learning study one was carried out with SNR -24dB and used the vowel /a/, and eight consonants /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/ with both male and female voices. Comparing babble noise and speech shaped noise, babble noise was more lifelike than speech shaped noise, so the babble background noise was chosen as the background noise in this present test.

All sounds were presented through Sennheiser HD 580 headphones. The SNR for each token was determined by comparing the root mean square average amplitude of each signal file with the babble background noise file. As described before, the SNR was fixed at -24dB across test sessions for the two test groups. Calibration was carried out before the main test took place. The IEC 711 acoustic coupler and a precision microphone were used to calibrate the output of VCV test stimulus. Then the maximum sound pressure levels from PC were controlled to make sure that the output from the software (MATLAB) was within exposure action value (65 dB SPL). The sampling rate for all signals was 44.1 kHz. Fig. 5.1 shows the time domain waveforms of example VCV syllable "/ABA/" under 0 dB and -24 dB input SNR in the background of babble noise.. a) Male voice /ABA/ in babble noise (0 dB input SNR); (b) Male /ABA/ in babble noise (-24 dB input SNR); (c) Female voice /ABA/ in babble noise (0 dB input SNR); (d) Female /ABA/ in babble noise (-24 dB input SNR).


(a) Time responses of male voice and background noise(SNR: 0dB)


(b) Time responses of male voice and background noise (SNR: -24dB)


(c) Time responses of female voice and background noise

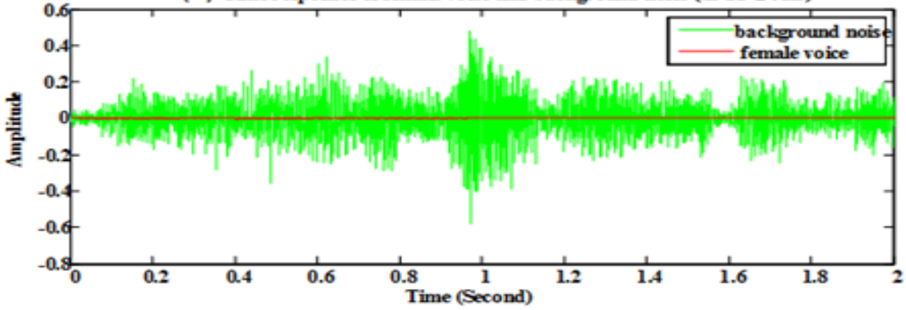

(d) Time responses of female voice and background noise (SNR:-24dB)

Fig. 5.1 Examples of target sound ("aba") in babble background noise with SNR0dB and SNR-24dB. The waveforms are shown for (a) Male voice /ABA/ in babble noise (0dB input SNR); (b) Male /ABA/ in babble noise (-24 dB input SNR); (c) Female voice /ABA/ in babble noise (0dB input SNR); (b) Female /ABA/ in babble noise (-24 dB input SNR).

5.2.1.3 Experimental procedure

All tests were carried out in a sound attenuating room, which is based at The University of Warwick. Before the test, a pure tone audiogram test was carried out to make sure that each volunteer qualified to participate in the study. After that, the instructions for the experimental tests were given to the participants to read and ensure that they understood the experiment. In order for the participants be familiar with the experiment, they were required to do a practice test (see Fig. 5.2), which included eight consonants /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/ with male and female voices, but no background noise. Then participants were randomly assigned to one of two groups: fixed training or random training group. All subjects were informed that they could withdraw from the study at any time.

There were three test sessions in the main test: pre-test, training, and post-test session (see Fig. 5.2). These three test sessions were carried out over three consecutive days, excluding weekends. All of the participants were required to attend a pre-test, training, and post-test session. The pre- and post-test sessions included two blocks of VCV trials (one male voice block and one female voice block) with random babble noise as the background noise. The training sessions took place over three days and the background noise for these training sessions differed for the two test groups. Ten blocks of VCV trials (5 male voice blocks and 5 female voice blocks) with random or fixed babble noise were presented in each training session for three days. Each block of the VCV task contained 64 trials (including eight repeats of each of the eight consonants /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/, in a random order). Consonants were presented in an /a/-consonant-/a/ format with both male or female voices against fixed or random babble noise, depending on each test session and group. There were 12 talkers presented in the babble noise (6 males, 6 females). The length of the target VCV stimulus was 1 second, and 2 seconds for the background noise (selected from a 60 seconds length babble noise). They were displayed simultaneously from the start point.

The fixed group was trained with fixed babble noise, and the other one was trained with random babble noise. As described in the introduction of this chapter, for the fixed babble noise background condition the exact same section of babble noise was selected as the test background noise on every single trial (but different sections were used for different participants). As the babble noise sample rate was 44100 Hz, the sound samples for the 60 seconds babble noise were 44100*60= 2646000.

A 2 second stimulus required 88200 sound sample points to be taken from 60 seconds of babble noise. The same start point (2 seconds) was selected from 60 seconds of babble noise for each participant in the fixed trained group (but from different start points for different participants). The following start indexes were selected for the fixed babble noise conditions: S1-824819; S2-243675; S3-1146848; S4-1915039; S5-1624709; S6-465129; S7-1306451; S8-465129; S9-2342922; S10-275652. For the random babble noise condition, different sections of babble noise were used as the test background noise on each trial. The flowchart of the experiment design can be seen in Fig. 5.2. Details of the test materials can be seen in Table 5.1.

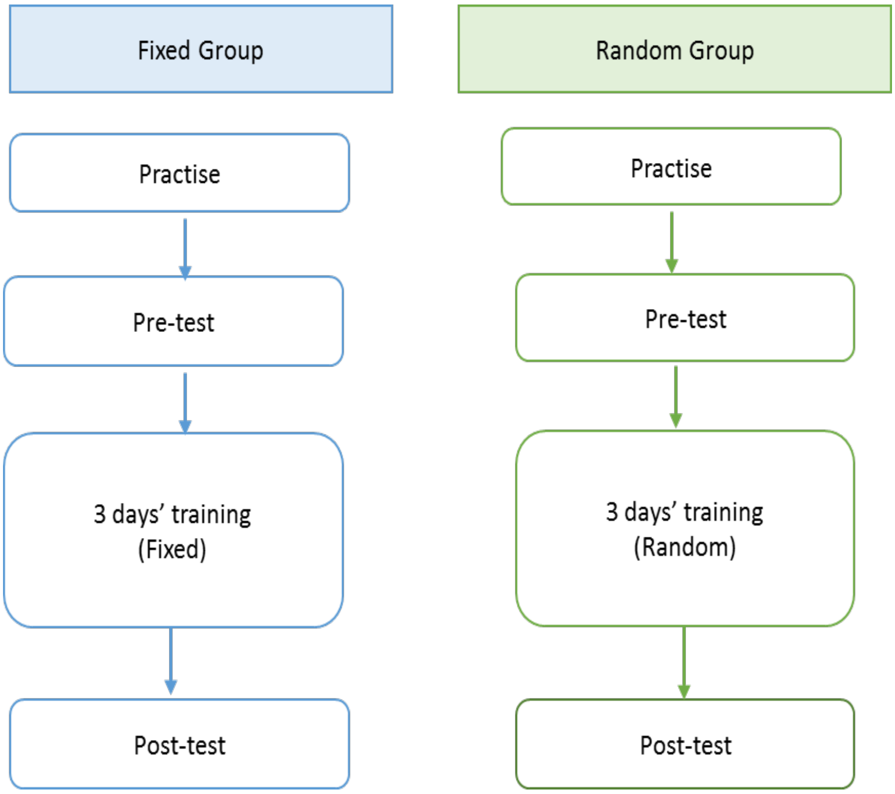


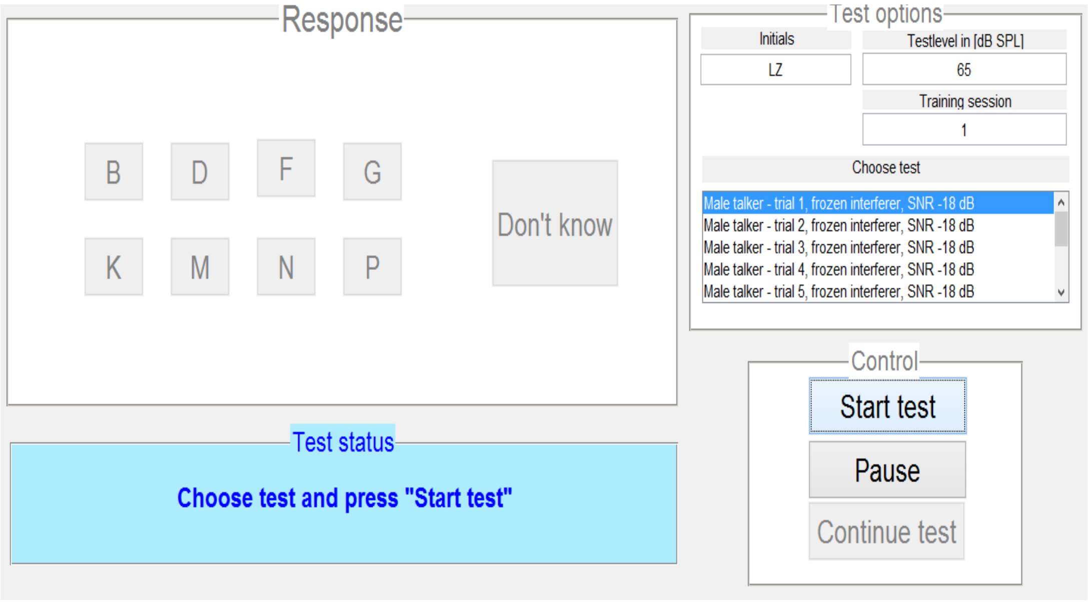
Fig. 5.2 The flowchart for a VCV training test with SNR-24dB

Table 5.1 Experimental procedure and test materials for the VCV training test with SNR-24dB

Session	Target Sound	Background noise	Block	Trials
Pre-test	**Vowel: /a:/** **Consonants : /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/**	**Random babble noise**	**2 blocks :** **1Male Voice** **1Female Voice**	**128**
Training 3 days	**Vowel: /a:/** **Consonants : /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/**	**Fixed or Random babble noise**	**10 blocks**	**640 per day**
Post-test	**Vowel: /a:/** **Consonants : /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/**	**Random babble noise**	**2 blocks :** **1Male Voice** **1Female Voice**	**128**

Fig. 5.3 shows the test interface for VCV study one. As shown in Fig. 5.4, there were nine choices in the "response" panel area of the test interface. Participants were

instructed to click one of the eight consonants that they heard during the test. For example, if they heard /ABA/, then they needed to click /B/ on the screen. If they were struggling to detect the target stimuli from the background noise, then they were instructed to click “Don’t know”. The “test status” showed how many trials were left for each test block. The “test option” panel was mainly used to enter the test parameters, listener’s ID information and for choosing test blocks (male and female voice blocks as displayed). The buttons on the control session indicated by “Pause” or “Continue Test” were used for the listeners to take a rest if they were tired during the test. Across the test sessions (pre-training, training, post-training), no feedback was provided, and participants were encouraged to guess if they were not sure. Both the proportion of correct responses and “Don’t know” responses in each test were calculated as a measure of the participants’ performances (for details of this results calculation from Matlab, see Appendix 5).



Initials: LZ
Testlevel in [dB SPL]: 65
Training session: 1
Choose test:
Male talker - trial 1, frozen interferer, SNR -18 dB
Male talker - trial 2, frozen interferer, SNR -18 dB
Male talker - trial 3, frozen interferer, SNR -18 dB
Male talker - trial 4, frozen interferer, SNR -18 dB
Male talker - trial 5, frozen interferer, SNR -18 dB

Test status
Choose test and press "Start test"

Fig. 5.3 Test interface for the VCV training study with SNR-24dB

5.2.1.4 Data analysis

All participants produced pre-test performance values (of correct and “Don’t know” responses) within three standard deviations of the mean. Hence, no datasets were identified as outliers, and all datasets were used for assessing the influence of training on participants’ performance on the VCV tasks. For both correct and “Don’t know” responses − in order to ensure that participants’ post-test values were not influenced by their pre-test scores, and to reduce the error variance for the analysis of the general results − analysis of covariance (ANCOVA) with pre-test performance as the covariate was used to compare the test results for the two training groups. However, the homogeneity-of-regression requirement for ANCOVA was violated for the correct responses in this experiment, so a two group (fixed vs random) × two time (pre vs post) analysis of variance (ANOVA) was conducted with repeated measures on the time factor (For “Don’t know” responses, both ANCOVA and ANOVA were conducted). In this present study, a significant group × time interaction indicated that the different day-to-day training method affected listeners’ performance. Paired t-tests were conducted as post hoc analyses for each training group to compare their pre-and post-test performance for both correct and “Don’t know” responses. Regarding the proportion of performance improvement for correct responses (learned values from the pre- to post-test), or performance decreases for “Don’t know” responses (decreased values from pre- to post-test), for these two test groups, an independent t-test was used to compare listeners’ improvement, or decreased performance, between the fixed and random babble noise training methods.

For correct responses, a mixed [two time (pre vs post) × two group (fixed vs random) × eight consonant (/b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/)] ANOVA was performed to analyse the proportion of the correct identification of eight consonants from the pre- to post-test (for both of the groups). T-tests were conducted to analyse individual correct responses across the test sessions (pre-test, training period and post-test session). The t-test analysis was an indicator of listeners' learning progress for both of the fixed and random training groups

For the "Don't know" responses, in order to answer the question about whether any improvements in the percentage of correct answers reflected changes in response criterion or perceptual processing, two group (fixed vs random) × two choice (guess rate vs correct improved amount) analysis of variance (ANOVA) with repeated measures on the choice factor was conducted to compare the guess rate (the decreased amount of "Don't know" responses divided by 8) with the increased amount of correct responses (pre to post correct responses improvement).

In this VCV study, there were 9 response choices (/b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/, /"Don't know"/) for listeners to choose during the test. If the improvement of correct responses from pre- to post- test was due to the reduction of "Don't know" choice, the value of improvement should not be significantly higher than the decreased amount of "Don't know" performance/8 (according to a dice throw with guess rate 1/6 example, if a participant did not choose any "Don't know" responses, their guess rate for correct responses was 1/8). So if the decrease (from pre-test to post-test) in "Don't know" responses divided by 8 was less than the increase in correct responses (pre- to post- correct response improvement), we can confirm that listeners' performance improvement was not because of changes in response criterion, it was instead due to listener's perceptual learning. Please see Appendix 6, Table 1 for the proportions of correct responses, incorrect responses, and don't know responses for each condition at each time point for SNR -24 dB.

5.2.2 Correct responses results

5.2.2.1 Pre- and post-test performance

As shown in Fig. 5.4, across the interventions of the test time period (pre- and post-test), the ANOVA showed that the main effect comparing the two groups' identification performance was not significantly different between each group [group, $F_{1,\ 18}$=1.28, $p > 0.05$; ANCOVA was excluded due to a significant heterogeneity of regression-line slope: $F_{1,\ 18}$= 11.60, p= 0.003; It was noticed the VCV identification performance at the pre-test for the fixed group was significantly higher than for the random group (t =2.31, p <0.05)]. However, VCV recognition performance significantly improved between the pre-test and post-test session for both of the fixed and random babble noise trained groups (time, $F_{1,\ 18}$ = 68.15, $p < 0.001$). The following paired t-tests were conducted for each test group to compare each one's pre-and post-test performance. Both of the trained groups showed a significant difference from their pre-test performance to their post-test results [fixed: t (9) = 6.54, $p < 0.001$; random: t (9) = 5.89, p <0.001].

An ANOVA also showed that there was a significant interaction between the group and the time period used (group × time interaction, $F_{1,\ 18}$ = 4.79, $p < 0.05$). An independent t-test showed that performance improvement from the random babble noise trained group was significantly higher than the performance improvement from the fixed babble noise trained group [t (18) =2.19, *p* = 0.042]. Therefore, the random babble noise trained group demonstrated a larger improvement (Pre-test: M = 31.80%, SD = 8.53%, Post-test: M = 50.63, SD = 11.08%; Improvement: M = 18.83%, SD = 10.10%) than the fixed babble noise trained group (Pre-test: M = 40.31%, SD = 7.97%, Post-test: M = 51.25, SD = 11.41%; Improvement: M = 10.94%, SD = 5.29%).

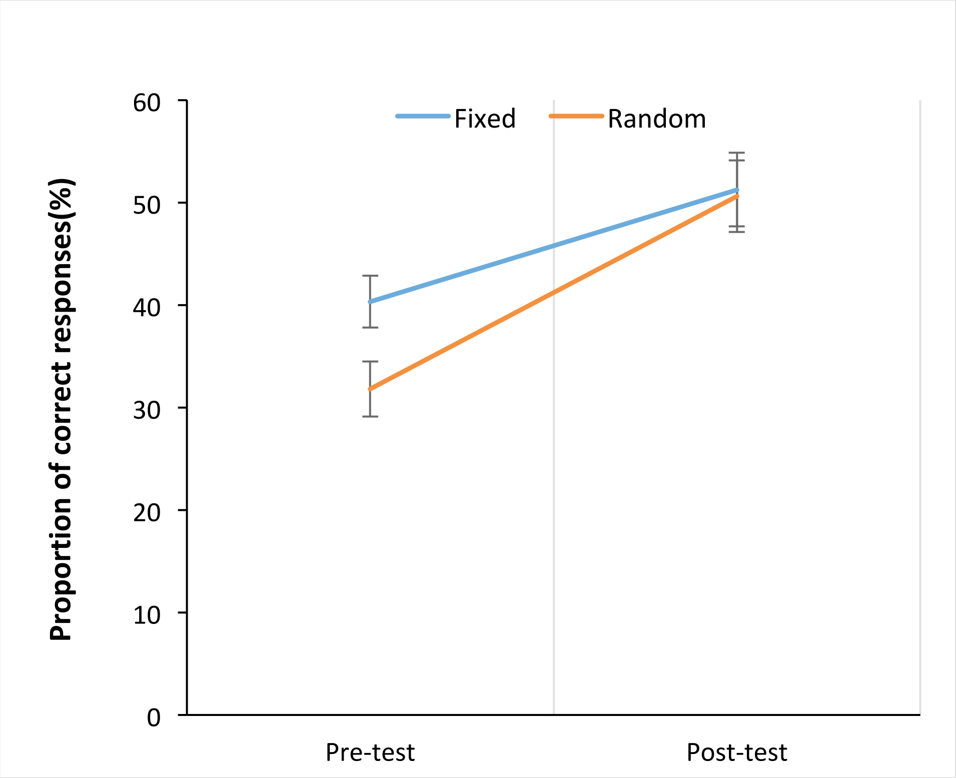
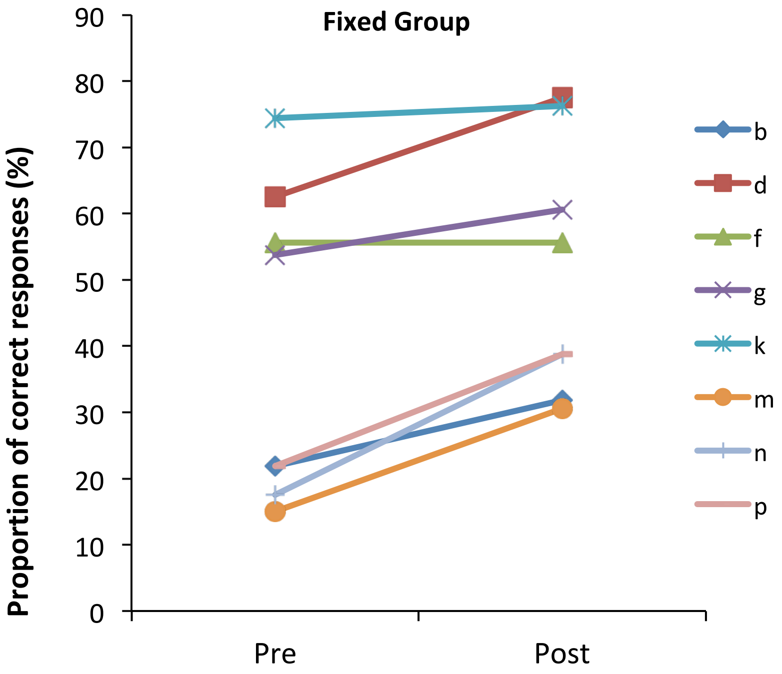


Fig. 5.4 Proportion of correct responses as a function of babble noise training (average across all eight consonants/d,f,g,k,m,n,b,p/), plotted separately for the fixed (n=10) and random babble noise training groups (n=10). Error bars reflect ± one standard error of the mean.

5.2.2.2 Proportion of correct responses for individual consonants

A mixed ANOVA showed that there were significant differences among the main effect of the eight consonants (consonant, $F_{7,\,126} = 53.62$, $p < 0.05$). As shown in Fig. 5.5, the former four consonants (/b/, /m/, /n/, /p/) generally yield low identification scores. In contrast, the latter four consonants (/d/, /f/, /g/, /k/) showed higher performance. The interaction between consonants and time was on the boundary of significance (consonant × time interaction, $F_{7,\,126} = 2.04$, $p = 0.055$). However, there was no significant interaction between time, consonant and group (time × consonant × group: $F_{1,\,126} = 2.66$, $p > 0.05$) across the interventions (pre- and post-test), and all eight consonants, ANOVA showed that the main effect comparing the two groups' identification performance did not result in a significant difference between each group (ANOVA: group, $F_{1,\,18}=1.29$, $p > 0.05$).



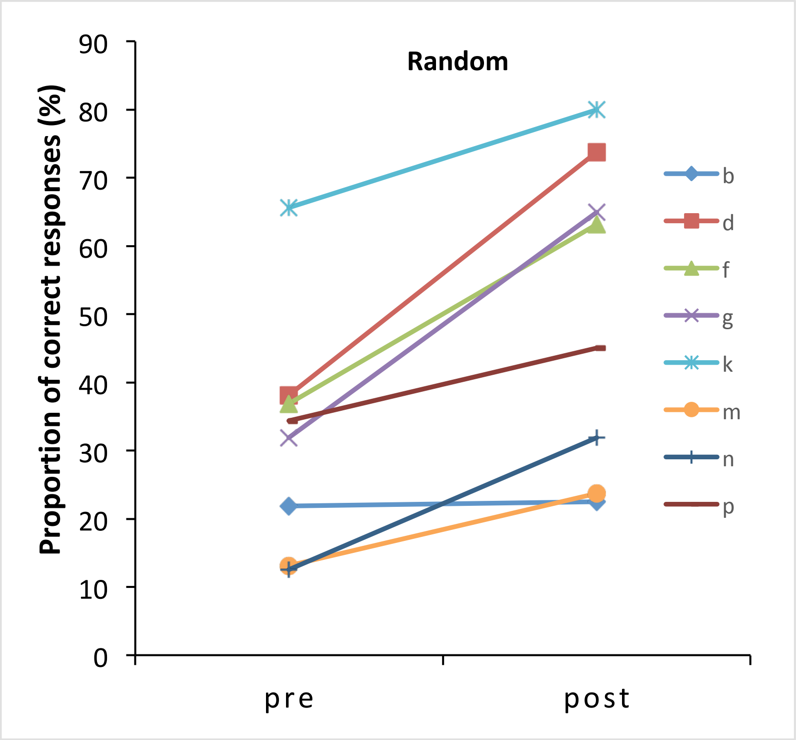
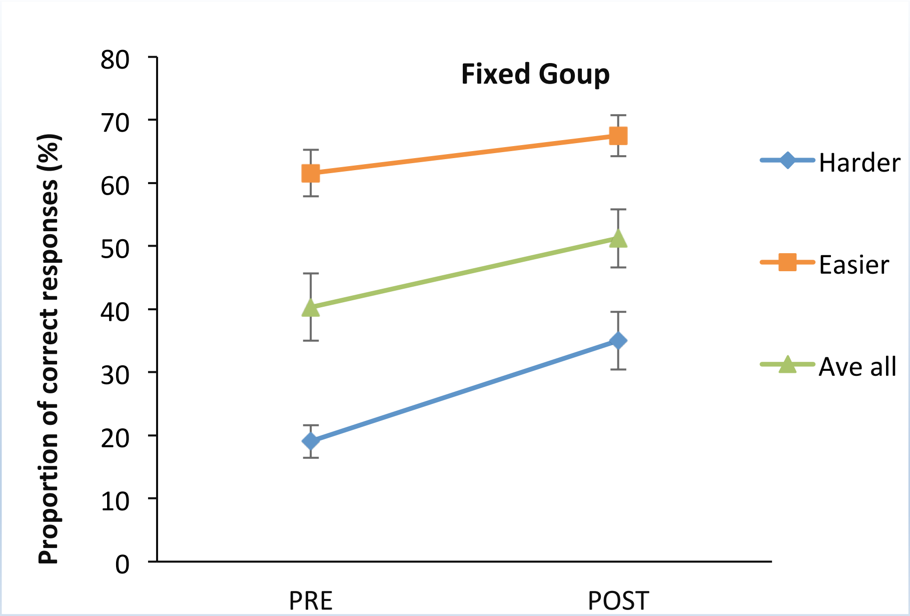


Fig. 5.5 Proportion of correct responses as a function of training with fixed (n=10) and random (n=10) babble noise (Pre- and Post-test with random babble noise), averaged across stimulus types (fixed and random) and plotted separately for stimuli produced by eight consonants (averaged across the male and female speakers).

Based on the results from the above section, the eight consonants were divided into two categories (see Fig. 5.6): easier /d,f,g,k / and harder /m,n,b,p/. A mixed ANOVA showed that the VCV identification performance for the easier consonants group was much higher than the VCV identification performance for the harder consonants, regardless of fixed or random group (consonant difficulty, $F_{1, 18} = 170.90$, $p < 0.001$). There was a significant interaction between consonant difficulty and group (consonant difficulty × group, $F_{1, 18} = 4.79$, $p < 0.05$). The interaction between consonant difficulty, time, and group was significant (consonant difficulty × time × group, $F_{1, 18} = 9.91$, $p < 0.05$). Repeated t-tests were conducted to compare the performance improvement (from pre- to post-test) of easier and harder consonants for both fixed and random babble noise training groups. VCV identification improvement for the fixed training group showed that the improvement of the harder consonants was larger than it was for the easier consonants [t (9) = 2.70, $p < 0.05$]. However, VCV identification improvement for the random trained group indicated that there was a tendency towards statistical significance between the harder and the easier consonants [t (9) = 2.20, $p = 0.055$].



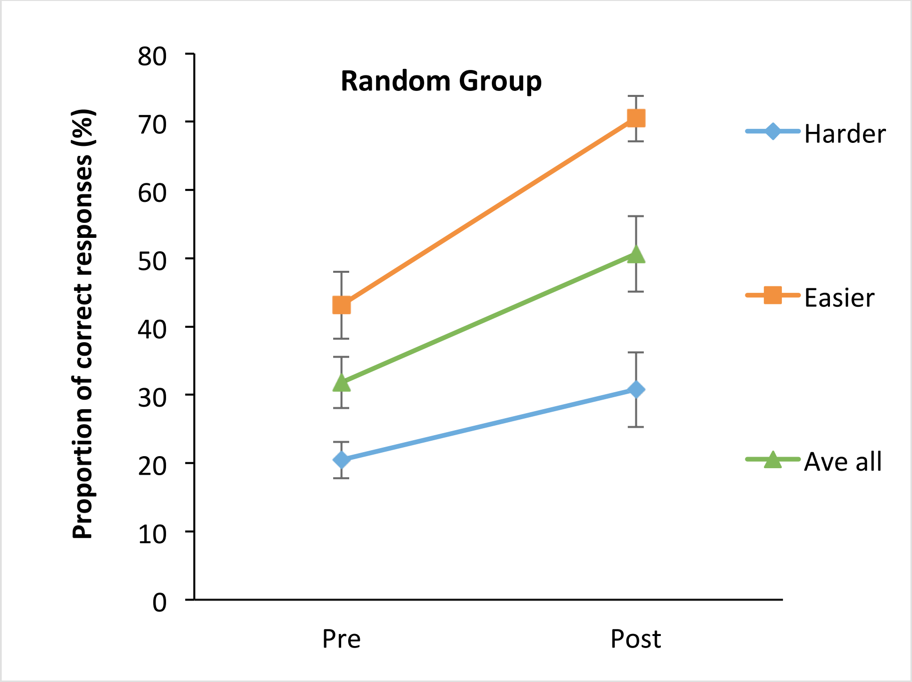


Fig. 5.6 Proportion of correct responses from the fixed and random babble noise training groups (Easier consonants /d,f,g,k/, Harder consonants /m,n,b,p/. Random babble noise was used for the pre and post-tests.), plotted separately for each of these consonants (averaged across the male and female speakers). Error bars reflect ± one standard error of the mean.

As described in section 5.2.2.1, across the interventions of test period (pre- and post-test) and consonant difficulty, an ANOVA showed that identification performance was not significantly different between each group (group, $F_{1, 18}$ =1.29, $p > 0.05$). There was significant difference for the effect of test time from pre- to post-test session (time, $F_{1, 18} = 170.9$; $p < 0.001$). However, there was no significant interaction between consonant difficulty and time (consonant difficulty × time, $F_{1, 18} = 0.76$, $p > 0.05$), and between time and group (time × group, $F_{1, 18} = 1.48$, $p > 0.05$). The interaction between consonant difficulty, time, and group was significantly different (consonant difficulty × time × group, $F_{1, 18} = 9.91$, $p < 0.05$). Independent t tests (comparing between groups separately for each combination of consonant type and test time) were conducted to compare the results of harder consonants /'m', 'n', 'b', 'p'/ and easier consonants /'f', 'g', 'k',’d’/for fixed and random training groups. The results of the harder consonants /'m', 'n', 'b', 'p'/ showed that there was no significant difference in either pre-test or post-test performance for the fixed and random groups [Pre: t (18) = 0.38, Post: t (18) = 0.59; all $p > 0.05$]. However, the results of the easier consonants /'f', 'g', 'k', 'd'/ showed that the fixed group substantially outperformed the random group in their pre-test results [t (18)= 3.02, $p < 0.05$], but there was no significant difference for their post-test results [t (18)= 0.64, $p > 0.05$].

Table 5.2 Proportion of correct responses as a function of fixed or random babble noise training (Easier consonants /d,f,g,k/, Harder consonants /m,n,b,p/) . All the averages are across the four consonants, within either the easier or harder consonants category.

	Fixed			**Random**		
	Pre	Post	Improvement	Pre	Post	Improvement
Harder	19.06	35.00	15.94	20.47	30.78	10.31
Easier	61.56	67.50	5.94	43.13	70.47	27.34
Ave ALL	40.31	51.25	10.94	31.80	50.63	18.83

5.2.2.3 Day to day training performance

Fig. 5.7 shows individual participants' VCV recognition performance for each test session during the fixed babble noise training experiment. There was considerable variability in the amount of identification performance for each listener. A one way ANOVA showed that listeners' performance with fixed babble noise training was significantly different across the test time period (time: $F_{4, 36} = 9.68$, $p < 0.001$). The following pairwise comparison demonstrated that listeners' performance did not rise immediately, and demonstrated no sharp improvements from the pre-test to the day 1 training sessions ($p > 0.05$). This was also true for listeners' performance between day 3 and the post-test ($p > 0.05$). For training session performance, there was no significant difference between day 1 and day 2 ($p > 0.05$). However, a significant increase was found between day 2 and day 3 ($p < 0.05$).

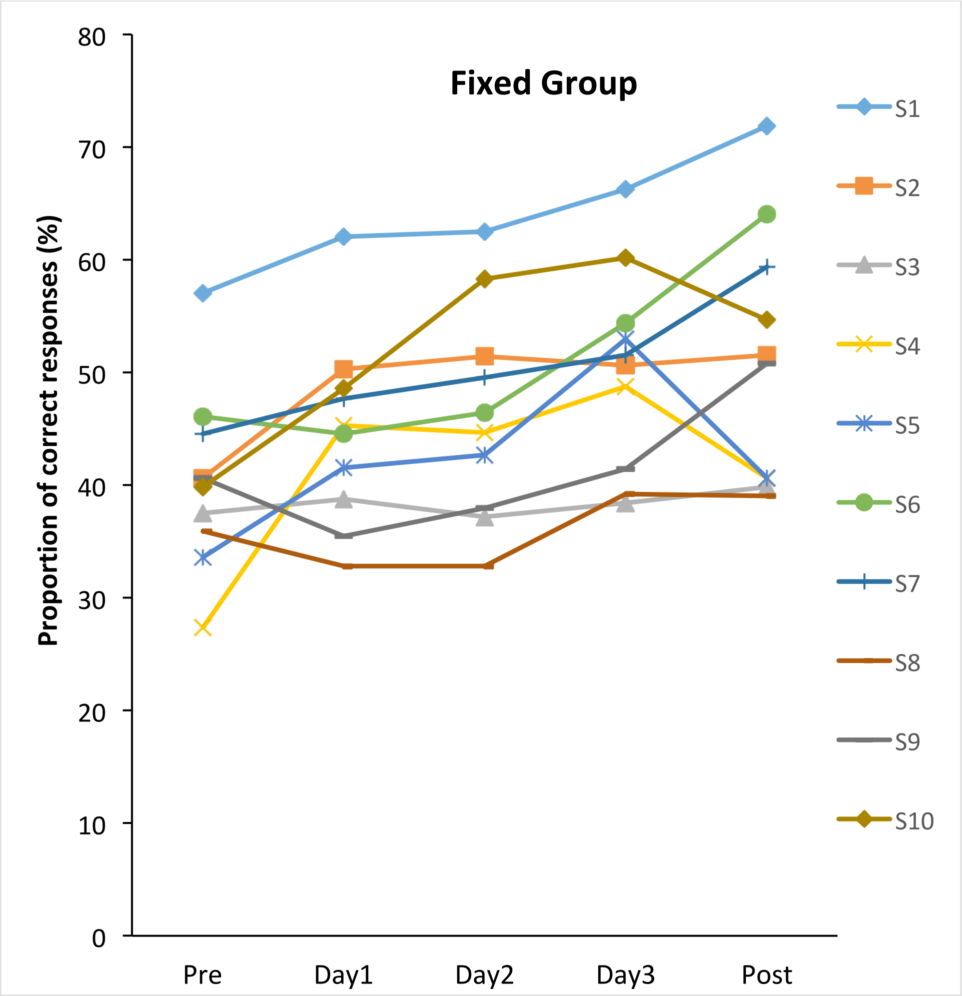


Fig. 5.7 The proportion of correct responses for the fixed babble noise group. Individual performance from the pre-test, the fixed babble noise training period and post-test sessions (average across speakers and eight consonants /d,f,g,k,m,n,b,p/).

As shown in Fig. 5.8, participants showed an overall improvement across the test sessions (pre-test, training sessions and post-test). During the training session, listeners' consonant identification performance improved gradually. Regarding the random trained listeners' performance during the test session, a one way ANOVA showed that there were significant differences for listeners' performance with random babble noise training across test time periods (time: $F_{4,\ 36} = 31.54$, $p < 0.001$). The following pairwise comparisons demonstrated that listeners' performance showed immediate and sharp improvements from the pre-test to the day 1 training sessions ($p < 0.05$). But this was not true for listeners' performances between day 3 and the post-test ($p > 0.05$). For the performance of training session, there was a significant increase from day 1 to day 3 (all $p < 0.05$).

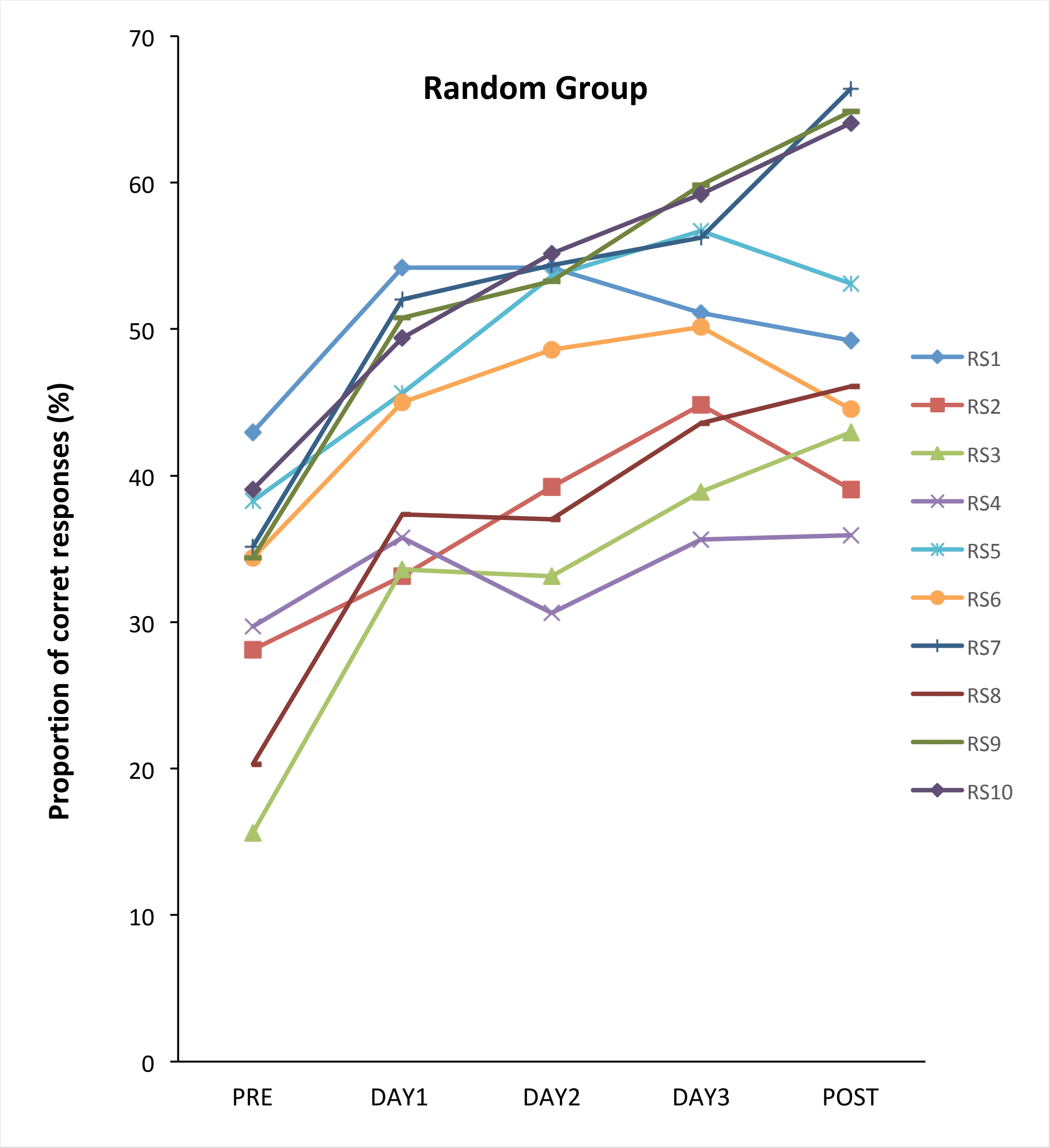


Fig. 5.8 The proportion of correct responses for the random babble noise group. Individual performance from the pre-test, the random babble noise training period and post-test sessions (average across speakers and eight consonants /d,f,g,k,m,n,b,p/).

5.2.3 “Don’t know” responses results

5.2.3.1 Pre- and post-test “Don’t know” performance

As shown in Fig. 5.9, across the interventions of test time (pre- and post-test), both the two-way ANOVA and ANCOVA test indicated that there was no overall “ Don’t know” performance difference between the pre- and post-test results for the fixed group and the random group (ANOVA: group, $F_{1, 18} = 0.99$, $p > 0.05$; ANCOVA: $F_{1, 17} = 2.67$; $p > 0.05$). However, the ANOVA showed that VCV “Don’t know” performance significantly decreased from pre-test to post-test session for both the fixed and random babble noise trained groups (time, $F_{1, 18} = 44.11.11$, $p < 0.001$). The following paired t-tests were conducted for each test group to compare each one’s pre-and post-test “Don’t know” performance. Both of the trained groups showed a significant decrease from their pre-test performance to the post-test results [fixed: t (9) = 6.17, $p < 0.001$; random: t (9) = 4.72, $p < 0.001$]. The ANOVA also showed that there a significant interaction occurred between the group and time period (group × time interaction, $F_{1, 18} = 4.62$, $p < 0.05$).

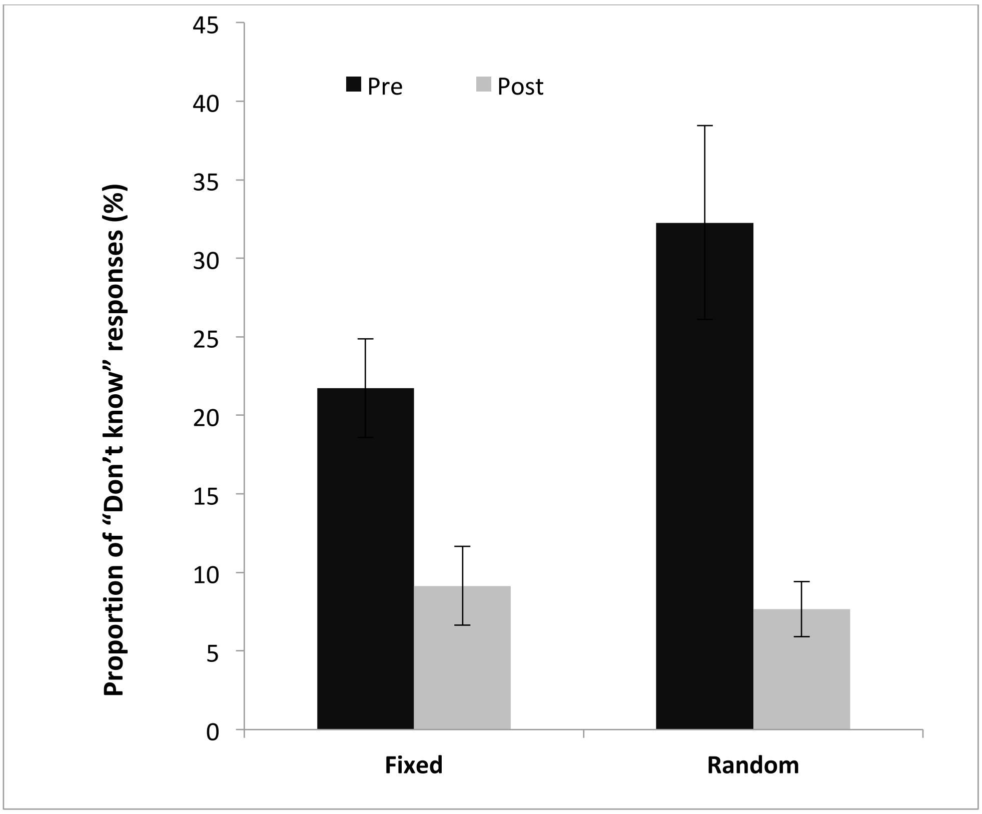


Fig. 5.9 Proportion of "Don't know" responses as a function of babble noise training (black: before training, grey: after training) plotted separately for the fixed (n=10) and random babble noise training groups (n=10). Error bars reflect ± one standard error of the mean.

An independent t-test showed that the amount of decreased "Don't know" performance from the random babble noise trained group was significantly larger than that of the fixed babble noise trained group [t (18) = 2.15, *p* = 0.046]. Therefore, the random babble noise trained group had a larger decreased amount of "Don't know" responses (Pre-test: M = 32.27%, SD = 19.49%, Post-test: M = 7.66, SD = 5.55%; Improvement: M = 24.61%, SD = 16.49%) than the fixed babble noise trained group (Pre-test: M = 21.72%, SD = 9.92%, Post-test: M = 9.14, SD = 7.94%; Improvement: M = 12.58%, SD = 6.45%).

5.2.3.2 Guess rate vs. Improvement in correct responses

As shown in Fig. 5.10, a two way ANOVA showed that comparing the two choices (guess rate vs improvement in correct responses) was significantly different between each other (choice, $F_{1,\,18} = 53.13$, $p < 0.001$). The following paired t-tests were conducted for each test group to compare each one's choice. Both of the two trained groups showed that the guess rate was significantly less than the amount of improvement in correct responses [fixed: t (9) = 5.87, $p < 0.001$; random: t (9) = 5.24, $p = 0.001$]. As shown in Table 5.3, the supposed guess rate for both the fixed (1.57%) and the random groups (3.08%) were less than the increased amount of correct responses for each group (fixed: 10.94%; random: 18.83%). Therefore, improvements in the percentage of correct responses indicated that the perceptual leaning processing was not due to changes in response criteria. This is confirmation that perceptual learning occurred for both the fixed and random training groups. Regarding the performance of each test group, there were also significant differences between them (group: $F_{1,\,18} = 5.84$, $p < 0.05$). However, the ANOVA showed that there was no significant interaction between the group and choice (group × choice interaction, $F_{1,\,18} = 3.44$, $p > 0.05$).

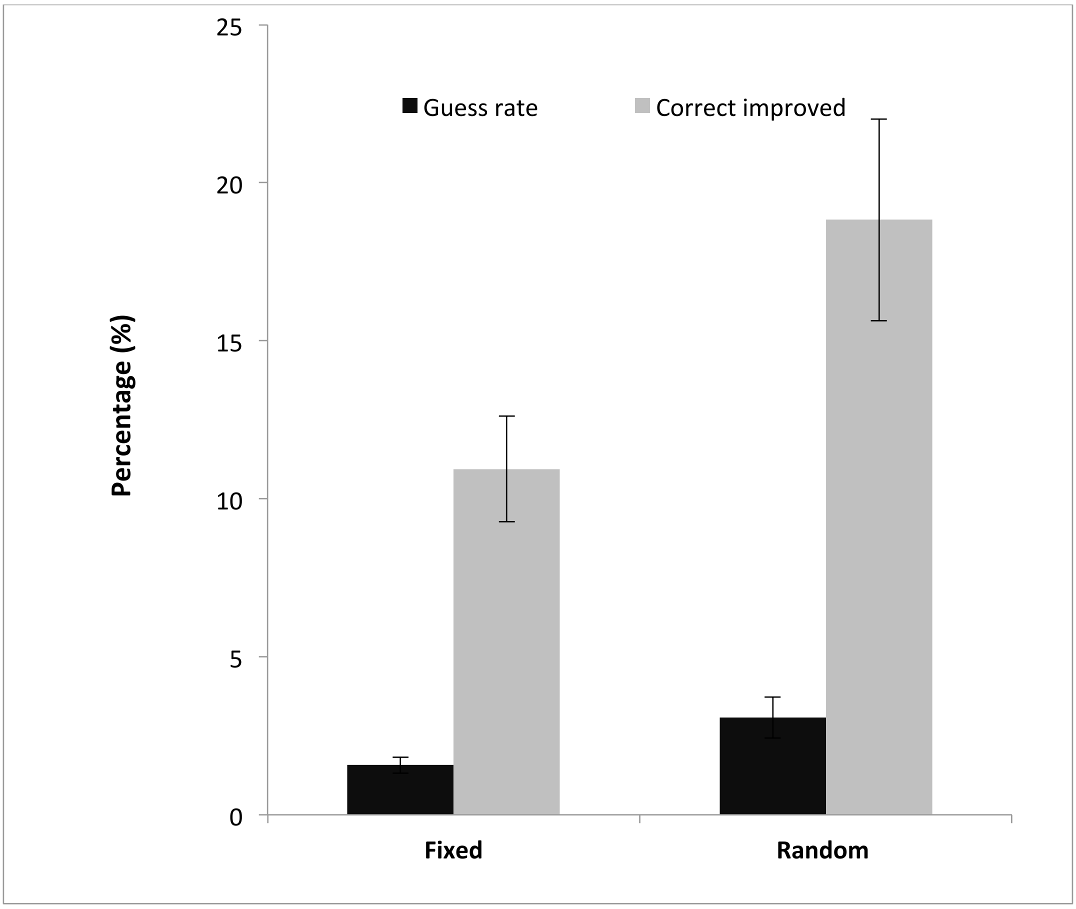


Fig. 5.10 Percentage in responses (Black: guess rate= decrease between pre- and post-test sessions in "Don't know" responses divided by 8; Grey: improvement in correct responses from pre- to post-test, named correct improved in the figure) as a function of babble noise training plotted separately for fixed (n=10) and random babble noise training groups (n=10). Error bars reflect ± one standard error of the mean.

Table 5.3 Data for decrease in "Don't know" responses, improvement in correct responses and guess rate for fixed and random group with SNR -24dB

SNR-24dB	**Decrease in "Don't know" responses**	**Improvement in Correct responses**	**Guess rate**
Fixed Group	12.58%	10.94%	1.57%
Random Group	24.61%	18.83%	3.08%

5.2.3.3 Day to day training "Don't know" performance

Fixed babble noise training group: Fig. 5.11 shows each individual participant's VCV "Don't know" performance for each test session during the fixed babble noise training experiment. There was considerable variability in the amount of "Don't know" responses for each listener. A one-way ANOVA showed that listeners' "Don't know" responses with fixed babble noise training was significantly different across the test time period (time: $F_{4, 36} = 22.56$, $p < 0.001$). The following pairwise comparison demonstrated that listeners' "Don't know" performance showed an immediate and sharp decrease from the pre-test to the day1 training sessions ($p < 0.05$). However, there was a sharp increase for listeners' "Don't know" performance between day3 and the post-test ($p < 0.05$). Regarding the performance of the training sessions, there was no significant difference between days, day1 to day3 (all $p > 0.05$).

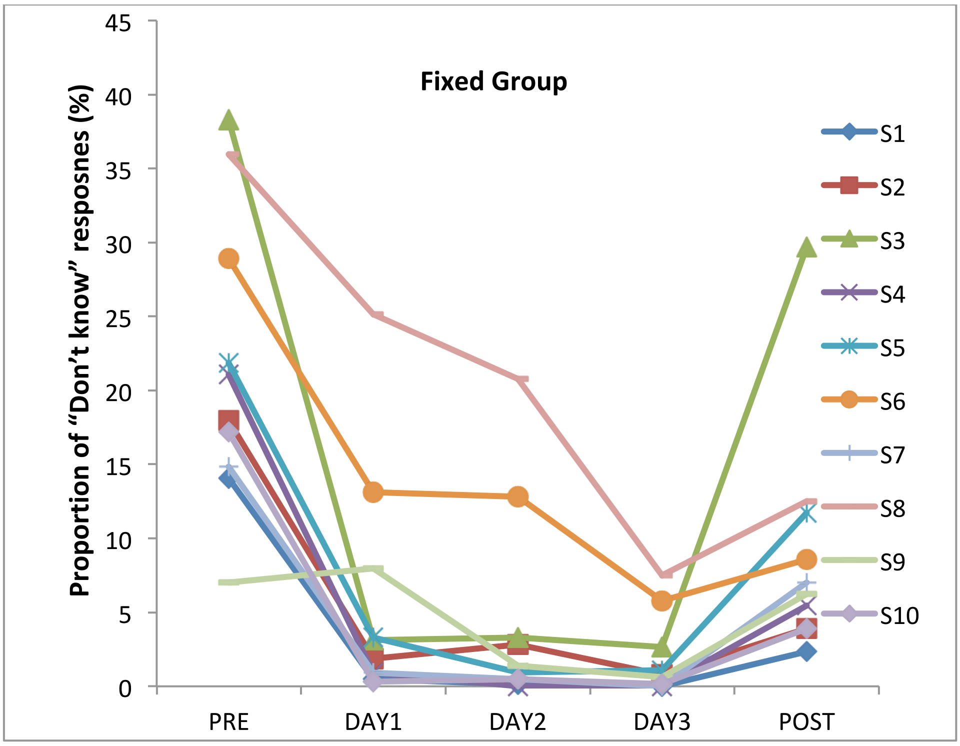


Fig. 5.11 The proportion of "Don't know" responses for the fixed babble noise group. Individual performance from the pre-test, the fixed babble noise training period and post-test sessions.

Random babble noise training group: As shown in Fig. 5.12, random trained participants showed an overall decrease in "Don't Know" responses across the sessions (pre-test, training sessions and post-test). A one way ANOVA showed that listeners' "Don't know" performance with random babble noise training was significantly different across time periods (time: $F_{4,36} = 19.34$, $p < 0.001$). The following pairwise comparisons of changes in "Don't Know" responses between consecutive days demonstrated that listeners' performance showed an immediate and sharp decrease from the pre-test to the day1 training session ($p < 0.05$). But this was not true for listeners' performance between day3 and the post-test ($p > 0.05$). During the training sessions, there was no significant difference in Don't Know responses between day1 and day2 or between day2 and day3 (all $p > 0.05$).

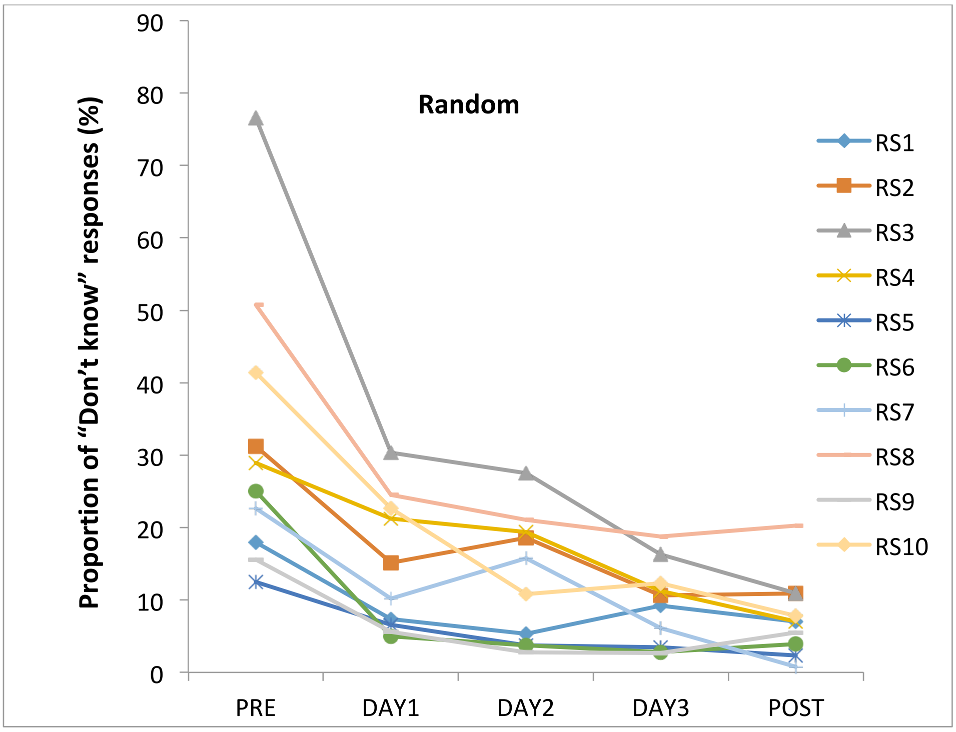


Fig. 5.12 The proportion of "Don't know" responses for the random babble noise group. Individual performance from the pre-test, the fixed babble noise training period and post-test sessions.

5.2.4 Learning outcomes

- ***The VCV correct responses and "Don't know" choice results***

Results from VCV study one showed that a learning effect was found from training listeners to recognise the stimulus sound against a fixed or random background noise. The random group showed larger learning effects (18.8%) than the fixed group (10.9%). However, it was noticed that in the pre-test results the fixed group outperformed the random group (the same tasks were given to these two groups in the pre-test, yet there were significant differences in their pre-test results). So the much larger improvements were due to the differences at the pre-test. These results could not therefore confirm whether fixed training is better than random training.

“Don’t know” responses decreased for both the fixed and random groups from pre- to post- test sessions. The amount of decreased “Don’t know” performance for the random group was larger than for the fixed group (Random vs Fixed: 24.61% vs 12.58 %). As the guess rate (Decrease in “Don’t know” responses divided by 8) was significantly less than the improvement in correct responses, it confirmed that any improvement in participants’ performance was not because of changes in the response criteria, but was due to the listeners’ perceptual learning.

- ***Individual consonant stimuli identification performance***

Regarding individual consonant stimuli identification performance for the fixed and random babble noise trained groups, test results suggest that the eight consonants can mainly be attributed to two categories (this held for both speakers). The first category consists of lower scored consonants /'m', 'n', 'b', 'p'/. These four consonants yielded generally lower identification performances than the other four consonants. There did not seem to be any substantial difference in either the pre-test or post-test performance between the fixed and random groups across these four consonants. The second category consists of the remaining four consonants /'f', 'g', 'k',’d’/. They yield generally much higher identification scores than the four consonants /'m', 'n', 'b', 'p'/, and for these higher scored consonants, the fixed group substantially outperformed the random group in the pre-test. This is the reason why the overall improvement from fixed babble training was much less than that of the random babble noise training method. The lower scored /'m', 'n', 'b', 'p' / and higher scored /'f', 'g', 'k', 'd'/ consonants for both fixed and random babble training groups were also analysed. The results indicated that the harder consonants /'m', 'n', 'b', 'p'/ for the fixed and random groups were similar in the pre-test, and the harder consonants presented a similar improvement between fixed and random training. Results also showed that for the fixed training group, the harder the VCV training task was, the higher the identification performance improvement. But for the random training group, the easier the VCV training task was, the higher the identification performance improvement.

- Further steps

As the visual study from Schubö *et al.* (2001) showed, people can improve their detection performance by learning to ignore fixed (visual) “noise”, but no learning effect was found from random “noise” training. Results from my study indicated that random babble noise training showed larger improvement than fixed babble noise training. However, based on the results above, easier consonants in the fixed group substantially outperformed the random group in the pre-test, as a consequence this study cannot confirm which training method (fixed vs random) leads to the most VCV identification performance improvement (see Table 5.2). In order to reduce variation in participants’ performance at the pre-test session stage, it was decided to increase task difficulty to investigate whether making the test tasks harder would cause learning improvement differences between the fixed babble noise training condition and the random babble noise training condition for next step.

5.3 VCV study two (SNR-30dB)

Following the results of the perceptual learning study in section 5.2 (VCV study one with SNR-24 dB), VCV study two focused on exploring the effects of making the VCV task harder than VCV study one and whether this would result in any performance improvement difference between the random babble noise training and the fixed babble noise training methods. VCV study two made the task more difficult by reducing the SNR and investigating if this led to any performance improvement difference between random and fixed babble noise training. For this experiment, the background noise was increased, and part of the experiment was repeated with an SNR of -30 dB**.** This second VCV study also tested whether the learning effect from training listeners with VCV against fixed babble noise generalized to a VCV against random babble noise condition. In order to answer this question, a control group without any training was added at this stage to explore the generalization effect. Therefore, in addition to the two objectives of VCV study one, the third objective of this study was to compare the results of the pre and post VCV tests with random babble noise in order to investigate whether training with VCV against fixed babble noise generalizes to VCV against random babble noise.

5.3.1 Test methods

5.3.1.1 Participants

Thirty normal-hearing native English speakers (16 males and 14 females) participated in this experiment. All of the participants had no prior experience participating in psychoacoustic experiments, and their pure tone thresholds were less than 20 dB HL. The age range was from 18 to 40 years old. The participants were all volunteers recruited from the student and staff population of The University of Warwick.

5.3.1.2 Stimuli

The SNR for this present VCV study two was fixed at -30 dB through the test sessions. Except for the SNR used for this study was different from the VCV study one, all the other test materials and interface were same as the previous section 5.2.1.

5.3.1.3 Experimental procedure

As one more control group without training is added for this experiment, the experimental design of VCV study two differs VCV study one (in section 5.2.1). The flowchart of this new experimental design can be seen in Fig. 5.13.

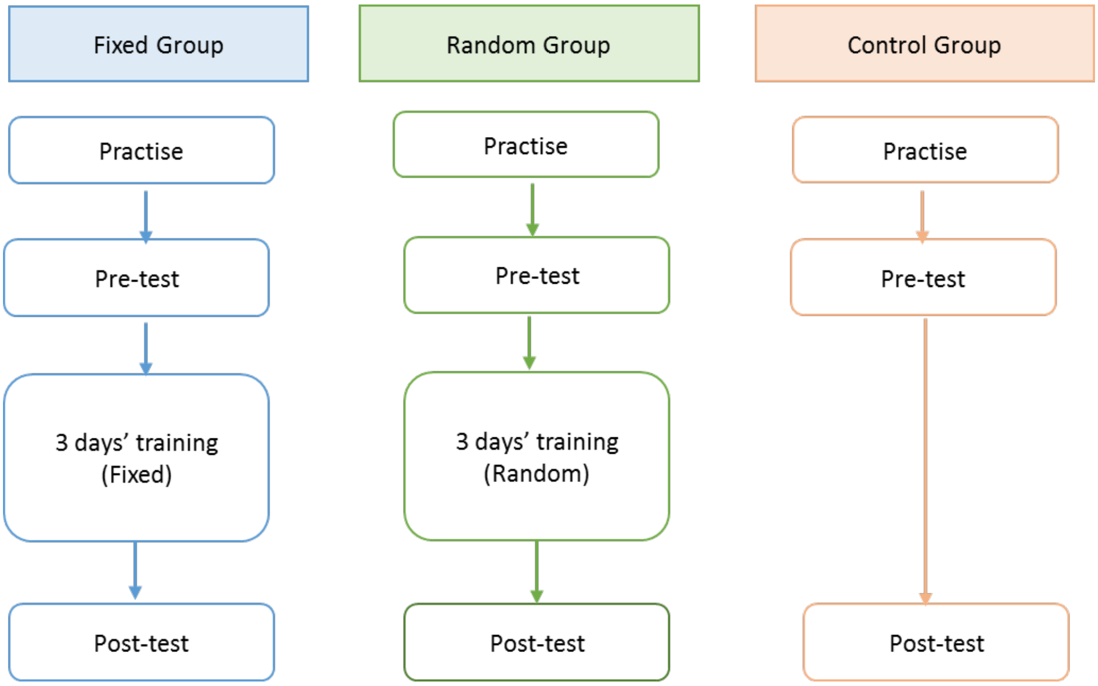


Fig. 5.13 Flowchart for the VCV training test with SNR-30dB

5.3.1.4 Data analysis

All participants produced pre-test performance values within two standard deviations of the mean, so no datasets were identified as outliers in this study, and all of the data were used for assessing the influence of training on participants' identification performance as follows: an analysis of covariance (ANCOVA), with pre-test performance as the covariate, was used to compare post-test VCV identification accuracy for the three groups; a mixed between-within analysis of variance (ANOVA); post hoc tests using Bonferroni correction for multiple comparisons; and t-tests were also used to analyse the results. Regarding the performance improvement for each test group, an ANOVA was used to investigate the learned values (improvement from pre- to post-test). Independent t-tests were used to compare the improvement for paired groups (fixed vs random, fixed vs control, random vs control). In order to explore the effect of training at the individual level, linear regression of the post-test performance on the pre-tests results for the two babble noise trained (fixed and random) groups with the control group was conducted to see the relationship between the pre- and post- test VCV identification performance across individual participants. Data analysis for "Don't know" performance was the same as that used in VCV study one (details in section 5.2.1.4). As participant 4 in the random test group did not choose any "Don't know" responses, this listener's data was excluded from the "Don't Know" analyses. See Appendix 6, Table 2 for the proportions of correct responses, incorrect responses, and don't know responses for each condition at each time point for SNR -30 dB.

5.3.2 Correct responses results

5.3.2.1 Pre- and post-test results

As show in Fig. 5.14, after adjusting for the pre-test results an ANCOVA showed that there was a significant difference between the three groups on post-test identification of VCV stimuli with babble background noise (group: $F_{2, 26}$= 17.25, $p < 0.001$). A mixed between-within ANOVA was conducted to assess the VCV identification performance of three different groups' VCV recognition against babble noise across the interventions of time period (pre- and post-test). The main effect of group was significant ($F_{2, 27} = 0.23$, $p < 0.001$). A significant difference was found between the pre- and post-training sessions (time: $F_{1, 27} = 112.95$, $p < 0.001$). There was a significant interaction between pre- and post-test session (time) and the three different test groups (time × group: $F_{2, 27}$=15.38, p <0.001). Paired t-tests were conducted for each test group to compare each one's pre-and post-test performance. As shown in Fig. 5.14, all three groups VCV recognition performance with random babble noise significantly improved between pre- and post -test [fixed: t (9)=4.64, p < 0.001; random: t (9) =10.24, $p < 0.001$; control: t (9)=3.28, $p < 0.05$)].

Regarding performance improvement from pre- to post-test for the three test groups, a one way between groups ANOVA showed that there was a significant difference between the three groups (group: $F_{2, 29} = 15.38$, $p < 0.001$). VCV identification improvements against babble noise for the three groups were 8.75% for the fixed training group, 18.36% for the random training group, and 5.07% for the control group. Independent t-tests indicated that among all three groups, the random babble noise trained group had the largest improvement (Pre-test: M = 11.88 %, SD = 5.70%, Post-test: M = 30.23, SD = 5.77%; Improvement: M = 18.36%, SD =

5.67%). It was greater than the learned values for both the fixed trained and control groups [random vs control: t (18) = 5.60, $p < 0.001$; random vs fixed, t (18) = 3.69, $p < 0.05$]. However, the improvement for the fixed training group (M = 8.75%, SD = 5.96%) was not significantly different from that of the control group (M = 5.08%, SD = 4.90%), [fixed vs control, t (18) =1.51, $p > 0.05$].

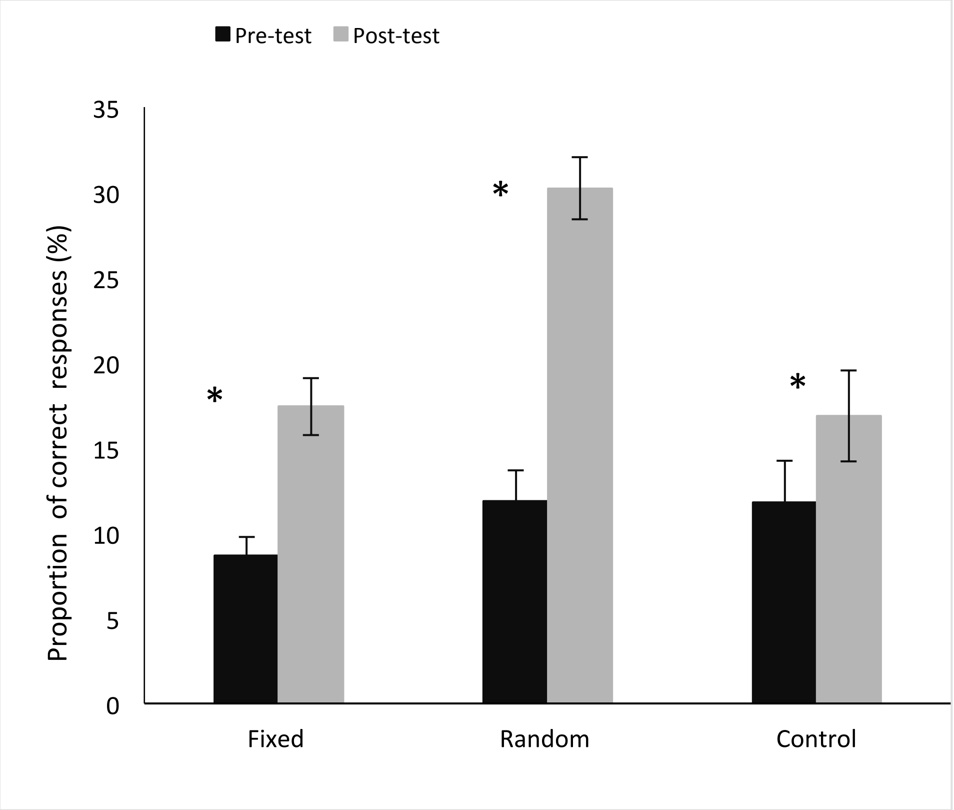


Fig. 5.14 Proportion of correct responses in the pre- and post-test sessions (all the average across all the eight consonants/d, f, g, k, m, n, b, p/), plotted separately for each of fixed (n=10), random babble noise training group (n=10) and control group (n=10). Error bars reflect ± one standard error of the mean.

5.3.2.2 Pre- and post-test results regression

Fig. 5.15 depicts the relationship between the pre- and post-test performance of the individual listeners for each test group (A: fixed group vs control group; B: random group vs control group; C: fixed group vs random group). For the babble noise trained participants (both fixed and random trained listeners), all the points were distributed above the positive diagonal (solid black no-improvement line), except one point from the fixed babble noise trained group (filled triangles), indicating that improvement was observed after multiple training sessions between the pre- and post-test for both the fixed and the random trained groups. Regarding the slopes of the regression lines fitted to the random babble trained (slope: 0.52; $r^2 = 0.26$, $F_{1,8} = 2.82$; $p = 0.13$) and the fixed babble noise trained groups (slope: 0.18; $r^2 = 0.01$, $F_{1,8} = 0.11$; $p = 0.75$), they were not significantly different from zero. The slope of the fixed babble noise trained group is quite shallow; indicating that most of the fixed babble noise trained listeners finished the test with similar post-test performance to each other.

For the control group (filled grey points), unlike the trained participants, the regression line fitted to these listeners was significantly different from zero (slope: 0.89; $r^2 = 0.67$, $F_{1,8} = 16.37$; $p = 0.004$) and had a slope approaching 1, indicating that there was a strong relationship between the pre- and post-test performance of the control group’s participants. Therefore, the control group also improved, but by a relatively constant amount regardless of the pre-test performance. Regarding the overall data, three points fell below the positive diagonal, one from the fixed babble noise trained group, and two from the control group, suggesting no improvement showed for these three participants between their pre- to post-tests.

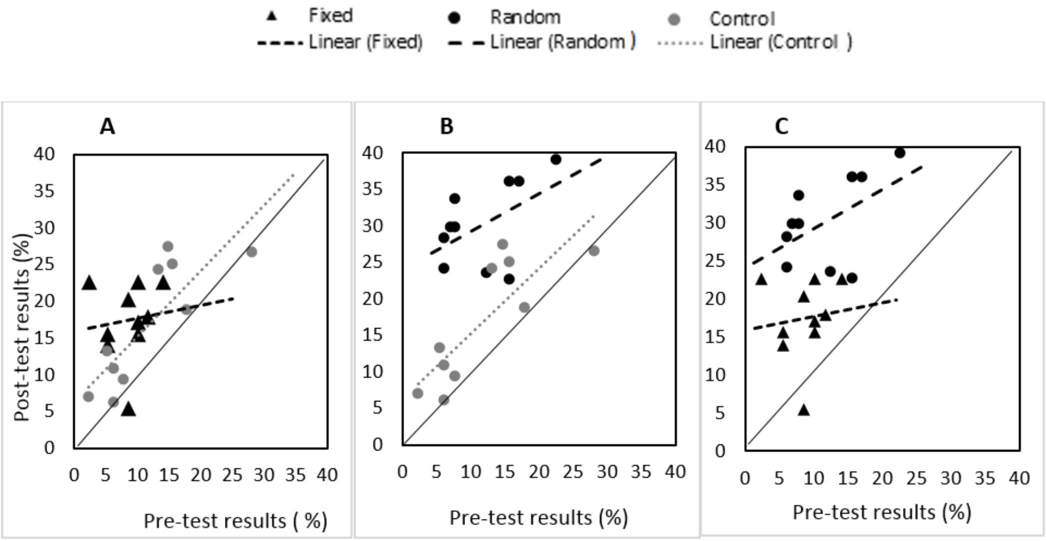


Fig. 5.15 The relationship between the pre- and post-test performance of individual listeners from fixed training, random training, and the control group (A: fixed group vs control group; B: random group vs control group; C: fixed group vs random group). For each of the test groups: the proportion of correct responses from pre-test (x axis) and post-test (y axis). Results are shown for the fixed babble noise (filled triangles), random babble noise training group (filled black points) and control group (filled grey points). The linear regression of the post-test performance on the pre-test performance was determined for each data set. Separate lines were estimated for fixed noise trained listeners (short black dashes), random babble noise trained participants (long black dashes) and the control group (short grey dashes).

5.3.2.3 Day to day training performance

Fig. 5.16 and Fig. 5.17 show individual trained participants' VCV recognition performance across experimental test sessions. There was considerable variability in improvement for fixed or random babble noise trained participants.

Fixed babble noise training group: For the fixed babble noise trained group, a one-way ANOVA showed that listeners' performance with fixed babble noise training was significantly different across the test time period (time: $F_{4, 36} = 6.4$, $p < 0.001$). The following pairwise comparisons between consecutive days demonstrated that listeners' performance showed immediate and sharp improvement from the pre-test to the day1 training sessions ($p < 0.05$). However, there was no significant change in listeners' performance between day3 and the post-test ($p > 0.05$). For the performance during training sessions, there were also no significant differences between day1 vs day2 or day2 vs day3 (all $p > 0.05$).

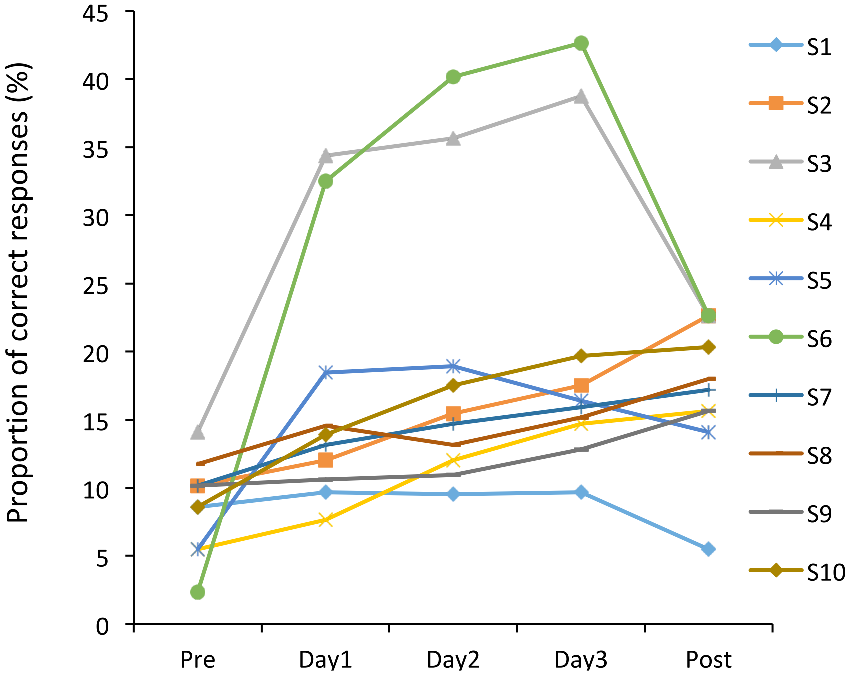


Fig. 5.16 The proportion of correct responses for the fixed babble noise group. Individuals' performance from the pre-test, the fixed babble noise training period and post-test sessions (average across speakers and eight consonants /d,f,g,k,m,n,b,p/).

Random babble noise training group: A one-way ANOVA showed that listeners' performance with random babble noise training was significantly different across the test time period (time: $F_{4, 36} = 45.86$, $p < 0.001$). The following pairwise comparisons demonstrated that unlike the fixed babble noise training group, both the average performance across participants from pre-test to day1, and day3 to post-test, showed significant improvements (all $p < 0.05$). For the performance of the training sessions, there were significant improvements between day1 vs day2 and day2 vs day3 (all $p < 0.05$). As shown in Fig. 5.16, all random babble noise trained participants showed an overall improvement across the sessions (pre-test, random babble training sessions and post-test), especially the listeners RS7 and RS10, who demonstrated a much larger VCV identification performance improvement from their pre-test to their post-test. Based on the test results across random noise training sessions, the learning performance from listener RS4 fluctuated from day1 to day3.

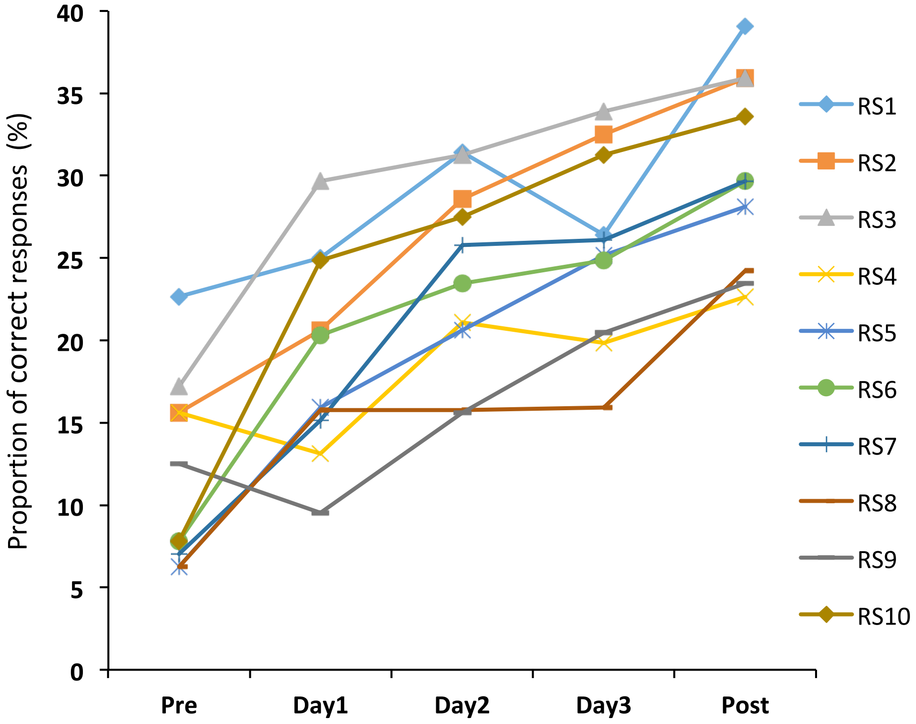


Fig. 5.17 The proportion of correct responsesfor the random babble noise group. Individual performance from the pre-test, the fixed babble noise training period and post-test sessions (average across speakers and eight consonants /d,f,g,k,m,n,b,p/).

5.3.3 "Don't know" responses results

5.3.3.1 Pre- and post-test "Don't know" performance

As shown is Fig. 5.18, after adjusting for the "Don't know" responses in the pre-test session, an ANCOVA showed that there was a significant difference amongst the three groups on post-test "Don't know" performance for VCV stimuli with babble background noise (group: $F_{2,\ 25} = 19.93$, $p < 0.001$). The following ANCOVA pairwise comparisons showed that the random training group had the lowest post "Don't know" responses after training (Random vs. Fixed: $p < 0.001$, Random vs. Control: $p < 0.001$). But there was no differences between the fixed and control group ($p > 0.05$). The mixed ANOVA was conducted to assess the VCV "Don't know" performance of three different groups' VCV recognition against babble noise across the interventions of the time period (pre- and post-test). The main effect of group was not significant (group: $F_{2,26} = 2.55$, $p > 0.05$). The results from the group analyses were in contrast between ANCOVA and the mixed ANOVA. It was noticed that the proportion of "Don't know" responses in the pre-test for the control group was the lowest among three groups (see Fig. 5.18). A one-way between groups ANOVA showed that there was a significant difference in the "Don't know" pre-test performance among the three groups (group: $F_{2,26} = 6.82$, $p < 0.05$). Post hoc tests using a Bonferroni correction for multiple comparisons revealed that there was no significant difference in pre-test "Don't know" performance between fixed and random group listeners' ($p > 0.05$). But the "Don't know" performance in the pre-test for control group was significantly lower than both fixed and random groups (all $p < 0.05$). Due to the proportion of "Don't know" responses in the pre-test for the control group being significantly lower than both the fixed and random groups, the ANCOVA test is a more reliable measure than the mixed-ANOVA here as it accounts for this difference. A significant decrease in "Don't know" responses was found between the pre- and post-training sessions (time: $F_{1,26} = 143.96$, $p < 0.001$). There was a significant interaction between pre- and post-test session (time) and the three different test groups (time × group: $F_{2,26} = 26.88$, $p < 0.001$). Paired t-tests were conducted for each test group to compare each one's pre-and post-test "Don't know" performance. As shown in Fig. 5.18, all three groups VCV "Don't know" performance significantly decreased between the pre- and the post-test [fixed: t (9) = 6.24, $p < 0.001$; random: t (8) =10.29, $p < 0.001$; control: t (9) =2.67, $p < 0.05$)]. However, results for paired t-tests can't be interpreted reliably due to the pre-differences.

Regarding the change in "Don't know" responses from pre- to post-test for the three test groups, a one way between groups ANOVA showed a significant difference amongst the three groups (group: $F_{2, 26} = 26.88$, $p < 0.001$). The reductions in "Don't know" responses for the three groups were 22.89% for the fixed training group, 46.70% for the random training group, and 7.90% for the control group. Independent t-tests indicated that among all the three groups, the random babble noise trained group had the largest decrease in "Don't know" responses (Pre-test: M = 57.12%, SD = 17.47%, Post-test: M = 10.42, SD = 6.71%; Improvement: M = 46.70%, SD = 13.62%). It was greater than the reduction for both the fixed trained and control groups [random vs control: t (17) = 7.70, $p < 0.001$; random vs fixed, t (17) = 4.12, $p < 0.05$]. The reduction in "Don't know" responses for the fixed training group was also significantly more than that of the control group, [fixed vs control, t (18) =3.18, $p < 0.05$]. However, it should be noted that the low Don't Know response rate of the control group in the pre-session could be contributing to the smaller reduction observed in the post-session for this group.

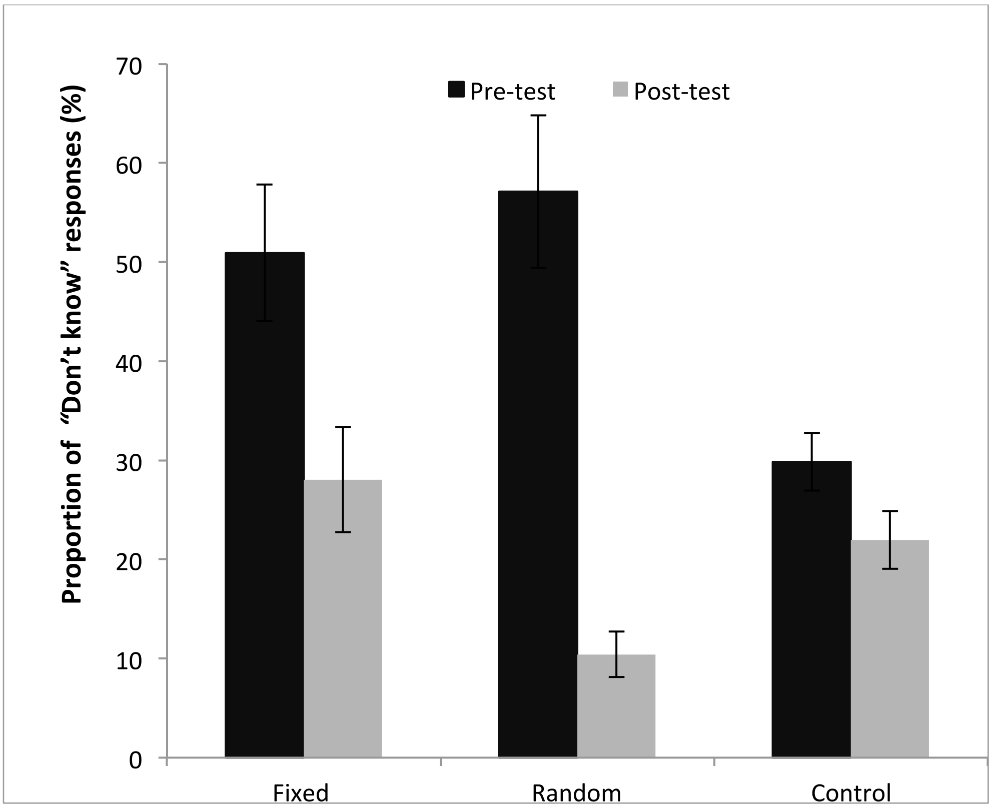


Fig. 5.18. The proportion of "Don't know" responses in the pre- and post-test sessions, plotted separately for each fixed (n=10), random babble noise training group (n=9) and control group (n=10). Error bars reflect ± one standard error of the mean.

5.3.3.2 Guess rate vs. Improvement in correct responses

As shown in Fig. 5.19, a two way ANOVA showed that the measures of guess rate vs correct improved responses were significantly different between each other (choice, $F_{1, 26} = 70.10$, $p < 0.001$). The following paired t-tests were conducted for each test group to compare each one's choice. All the three groups showed that the guess rate was significantly less than the amount of improved correct responses [fixed: t (9) = 3.25, $p < 0.05$; random: t (9) = 12.02, $p < 0.001$; control: t (9) = 2.31, $p < 0.05$; see Table 5.4]. Therefore, the improvements in percentage correct reflected perceptual processing, not changes in responses criterion. It confirmed that perceptual learning occurred for all the groups. Regarding the performance for each test group, there was also significant different between each other (group: $F_{2,\,26} = 30.28$ $p < 0.001$). The ANOVA also showed that there was a significant interaction between the group and choice (group × choice interaction, $F_{2,\,26} = 9.58$, $p < 0.05$).

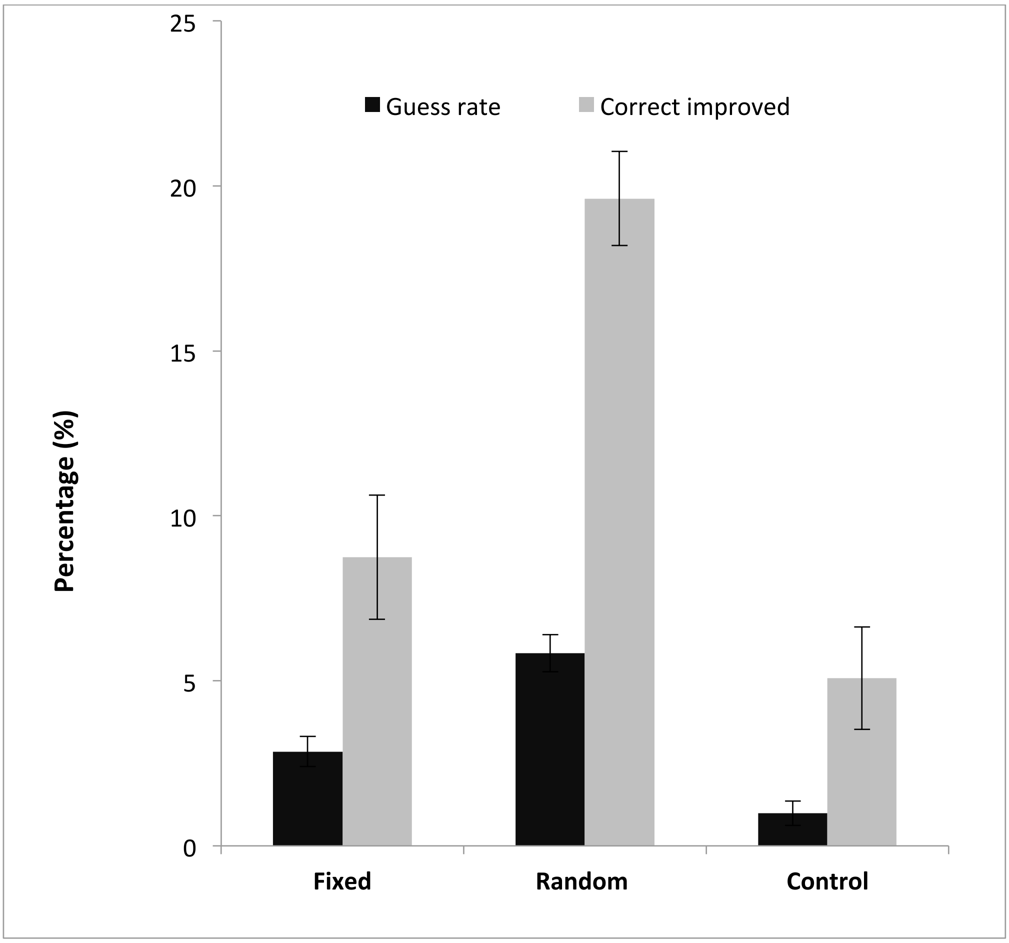


Fig. 5.19 Percentage in responses (Black: guess rate = decrease between pre- and post-test session in "Don't know" responses divided by 8; Grey: improvement in correct responses from pre- to post-test, named correct improved in the figure) as a function of babble noise training plotted separately for fixed (n=10) and random babble noise training groups (n=9). Error bars reflect ± one standard error of the mean.

Table 5.4. Data for decrease in "Don't know" responses, improvement in correct responses and guess rate for fixed, random and control group with SNR -30 dB (Excluded the participants RS04)

SNR-30dB	Decrease in "Don't know" responses	Improvement in Correct responses	Guess rate
Fixed Group	22.89%	8.75%	2.86%
Random Group	46.70%	19.62%	5.84%
Control Group	7.90%	5.08%	0.99%

5.3.3.3 "Don't know" performance for day to day training

Fig. 5.20 and Fig. 5.21 show individual trained participants' "Don't know" performance across test sessions. There was considerable variability in the decreased amount of "Don't know" performance for fixed and random babble noise trained participants.

Fixed babble noise training group: A one-way ANOVA showed that listeners' performance with fixed babble noise training was significantly different across the test time period (time: $F_{4,\ 36} = 19.25$, $p < 0.001$). The following pairwise comparisons demonstrated that listeners' performance showed an immediate and sharp decrease from the pre-test to the day1 training session ($p < 0.05$). However, listeners' "Don't know" performance between day3 and the post-test showed a sharp increase ($p < 0.05$). For the performance during training sessions, there was no significant change between day1 vs day2 or day2 vs day3 (all $p > 0.05$).

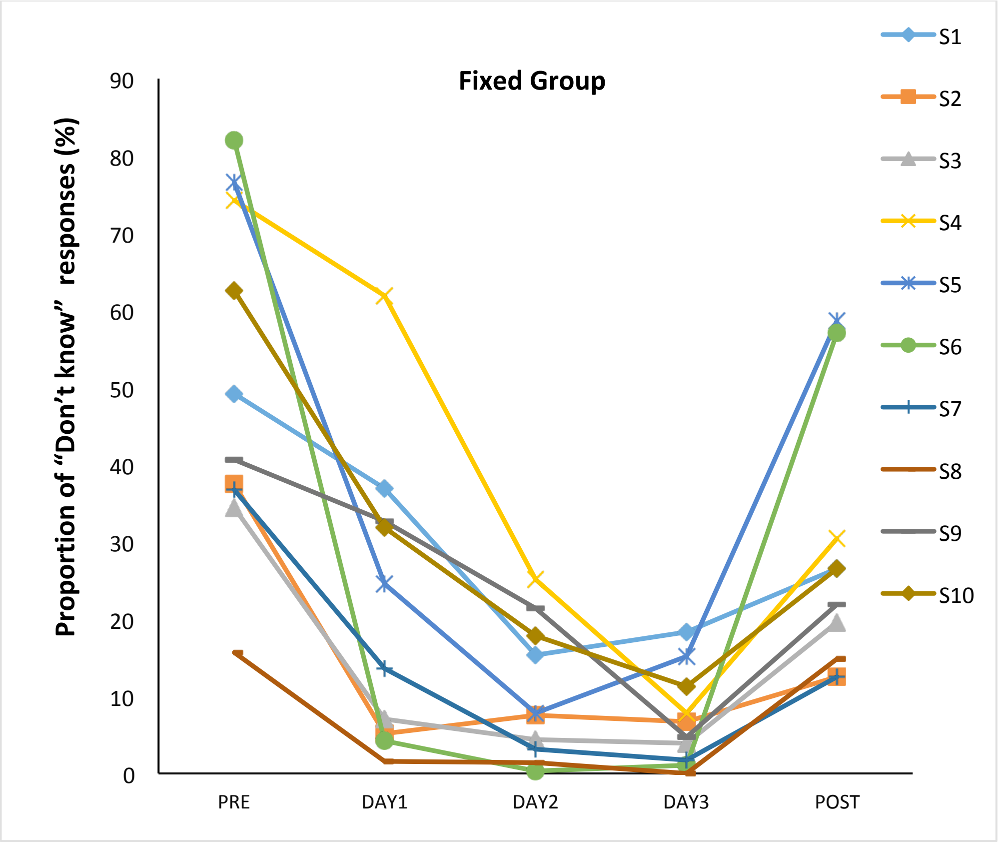


Fig. 5.20 The proportion of “Don’t know” responses for the fixed babble noise group. Individuals’ performance from the pre-test, the fixed babble noise training period and post-test sessions.

Random babble noise training group: A one way ANOVA showed that listeners’ “Don’t know” performance with random babble noise training was significantly different across test time period (time: $F_{4,\ 32} = 45.86$, $p < 0.001$). The following pairwise comparisons demonstrated that unlike the fixed babble noise training group, the average performance across participants from pre-test to day1 showed a significant decrease ($p < 0.05$), but no significant change from day3 to post-test ($p > 0.05$). For the performance during training sessions, there was a significant reduction between day1 and day2 ($p < 0.05$), but no significant difference between day2 and day3 ($p > 0.05$).

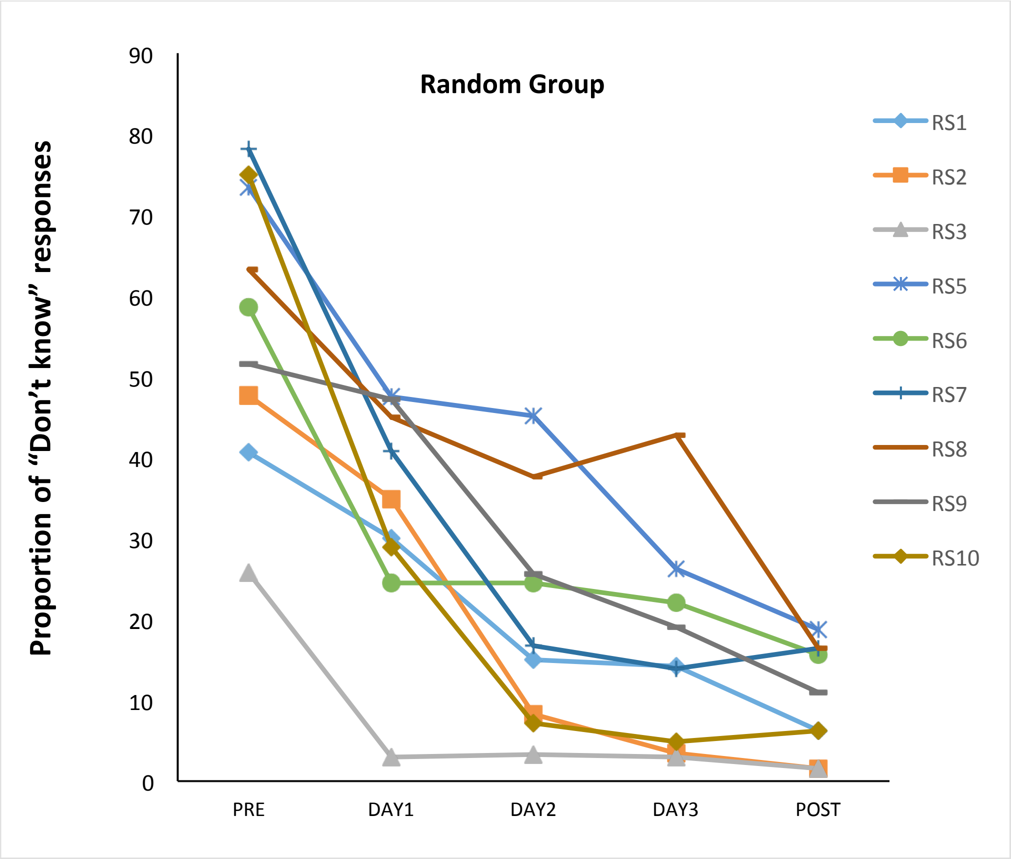


Fig. 5.21 The proportion of "Don't know" responses for the random babble noise group. Individuals' performance from the pre-test, the random babble noise training period and post-test sessions.

5.3.4 Learning outcomes

- ***The VCV correct results***

Based on the test results in section 5.3.2, it was found that participants' performance improved significantly for all three groups, regardless of whether they were trained or untrained. Regarding the generalized effect, the results showed that identification performance improvement for the control group was similar to the identification performance improvement for the fixed babble noise trained group. Therefore, no generalized effect occurred from training with the fixed babble noise condition to random babble noise environment. However, training with random babble noise produced better identification performance against a random-noise background than training with fixed noise.

- ***"Don't know" choice results***

"Don't know" responses decreased for all fixed, random and control groups (pre- to post-test session). Random was significantly lower than Control and Fixed, but Fixed was not significantly different to Control. Among the three test groups, the reduction in "Don't know" responses for the random group was the largest (Random vs Fixed vs Control: 46.70% vs 22.89% vs 7.90%). However, regarding the guess rate, the control guess rate may be unreliable due to the Pre-test "Don't know" response rate being much lower than the other groups. Apart from the control groups, the guest rate for the other groups have similar trends to VCV study one, the results from the guess rate (the amount of decrease in "Don't know" performance/8) and the improvements in the percentage of correct responses confirmed that participant's performance improvement was not due to changes in response criteria, but due to listeners' perceptual learning.

- ***Why not compare harder and easier consonant groups in VCV study two?***

The results of VCV study one cannot confirm whether fixed babble noise training was better than random babble noise training, but visual theory shows that fixed training is better than random training. Results in VCV study one showed that the fixed group was outperformed to random group at the pre-test stage. It was noticed that although the easier consonants /'f', 'g', 'k',’d’/in the fixed group was much higher than the random trained group, the easier consonants procured a similar post-test performance between fixed training and random training. However, the harder consonants / 'm', 'n', 'b', 'p'/ for both the fixed and random groups were similar in the pre-test, in that the harder consonants obtained a similar post-test performance between fixed and random training. Therefore, in order to reduce participants' performance variety at the pre-test sessions, it was decided to increase the task difficulty to investigate whether making the test tasks harder would lead to learning improvement differences between the fixed babble noise training and random babble noise training conditions. Accordingly, for VCV study two the SNR was decreased to make the task harder by increasing the task difficulty. However, the results showed that when the task is harder, random babble noise training led to better perceptual learning performance than the training with fixed babble noise conditions. So at this stage, there was no requirement to analyse the easier and harder consonants for VCV study two again.

5.4 Comparison of VCV studies one and two

5.4.1 Correct responses results: SNR -24 dB vs SNR-30 dB

Fig. 5.22 displays the VCV identification performance accuracy of participants for both studies of SNR -24 and SNR -30 dB over time periods (pre-test, day1, day2, day3 and post-test). A mixed analysis of variance (ANOVA) was conducted to assess the impact of two different training methods (fixed and random babble noise training) with two signal to noise ratios (SNR-24 dB and SNR-30 dB) on participants' performance across time periods (pre-test, day1, day2, day3 and post-test). The mixed ANOVA showed that after training, listeners' post-test VCV identification performances were significantly better than their pre-test results (time: $F_{4,\ 144} = 64.38$, $p < 0.001$). A significant interaction was also observed between training babble noise types and time periods (time × noise: $F_{4,144} = 4.96$, $p < 0.05$). However, no significant interaction was observed between the SNR and time periods (time × SNR: $F_{4, 144} = 0.49$, $p > 0.05$) or between test period, SNR, and babble noise type (time × SNR × noise: $F_{14,144} = 2.28$, $p > 0.05$). Results also showed that VCV identification accuracy at two different SNR levels was significantly different from each other (SNR: $F_{1,36} = 128.26$, $p < 0.001$), but there was no difference between the types of babble noise used in the training tests (noise: $F_{1,36} = 0.58$, $p > 0.05$). There was a significant interaction between trained noise and SNR levels (noise × SNR, $F_{1,36} = 3.04$, $p > 0.05$).

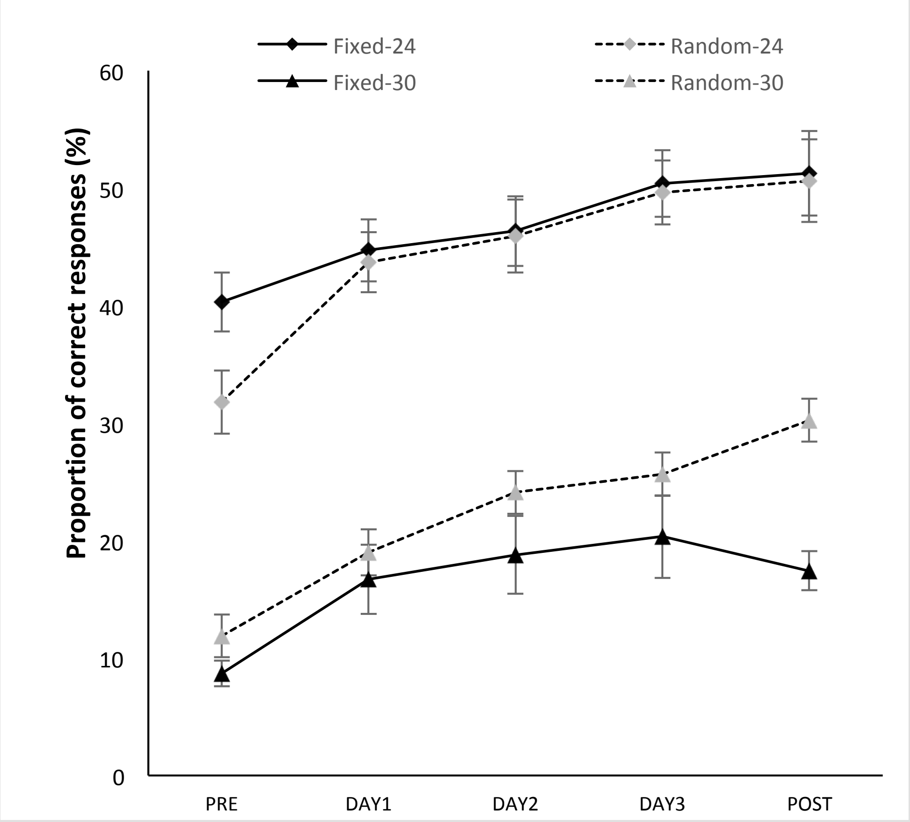


Fig. 5.22 Proportion of correct responses as a function of babble noise training (all the average across all the eight consonants /d,f,g,k,m,n,b,p/), plotted separately for each of fixed (n=10) and random babble trained with SNR -24 dB (n=10), fixed (n=10) and random babble trained with SNR -30 dB (n=10). Error bars reflect ± one standard error of the mean.

A mixed ANOVA was used to test listeners' improvement in correct responses across adjacent test sessions with different training methods (fixed and random babble noise) at the two SNR levels (SNR -24 and SNR -30). Result showed that there was significant decrease across time sessions (time: $F_{3,\,108} = 10.53$, $p < 0.001$). There was no significant interaction between SNR and time (SNR × time, $F_{3,\,108} = 0.80$, $p > 0.05$), between SNR and noise (SNR × noise, $F_{1,\,36} = 0.15$, $p > 0.05$) and between noise and time (noise × time: $F_{3,\,108} = 0.94$, $p > 0.05$). But there was a significant difference between fixed and random training methods (noise, $F_{1,\,36} = 15.48$, $p < 0.001$). However, no significant between different SNR (SNR, $F_{1,\,36} = 0.36$, $p > 0.05$) and also no significant interaction was observed between test period, SNR, and babble noise type (time × SNR × noise: $F_{3,108} = 3.18$, $p > 0.05$).

5.4.2 "Don't know" responses results: SNR-24dB vs SNR-30dB

A mixed between-within subjects analysis of variance (ANOVA) was conducted to assess the impact of two different training methods (fixed and random babble noise training) with two signal to noise ratios (SNR-24 dB and SNR-30 dB) on participants' "Don't know" performance across time periods (pre-test, day1, day2, day3 and post-test). Results showed that " Don't know" responses at the two different SNR levels were significantly different from each other (SNR: $F_{1,\,35} = 20.41$, $p < 0.001$). The following pairwise comparisons showed the "Don't know" response for the SNR-24 dB was lower than SNR-30 dB ($p < 0.001$). There was no difference between the types of babble noise used in the training tests (noise: $F_{1,\,35} = 4.17$, $p > 0.05$) and no significant interaction between trained noise and SNR levels (noise × SNR, $F_{1,\,35} = 0.23$, $p > 0.05$).

The mixed ANOVA also showed that after training, listeners' post-test "Don't know" performances were significantly lower across test sessions (time: $F_{4,\ 140}$ = 79.998, $p < 0.001$). A significant interaction was observed for training babble noise types and time periods (time × noise: $F_{4,\ 140}$ = 8.75, $p < 0.001$). The following comparison of the "Don't know" responses between fixed and random training method across time demonstrated that the "Don't know performance in pre and day1 were similar ($p > 0.05$), while the post session showed the random training group had lower "Don't know" responses than the fixed training group ($p < 0.01$). The day2 and day3 sessions also showed that the fixed training group had lower "Don't know" responses than the random training group (all $p < 0.01$). These comparisons indicated that the participants in the fixed training group are more confident (with lower don't know responses) than in the random training group at day2 and day3. Then in the post-test session, the random group showed more confidence. This reflected that changing background noise for the fixed training group at post-test session reduced participants' performance confidence.

There was also a significant interaction between the SNR and time periods (time × SNR: $F_{4,\ 140}$ = 9.46, $p < 0.001$). Comparison of the "Don't know" responses between SNR -24 dB and SNR -30 dB across time showed that there was no significant "Don't know" performance difference in pre, day3 sessions ($p > 0.01$). The day1, day2 and post sessions all showed that SNR-24 dB with lower "Don't know" responses than SNR -30 dB (all $p < 0.01$). However, no significant interaction was observed between test period, SNR, and babble noise type (time ×SNR ×noise: $F_{4,\ 140}$= 1.94, $p > 0.05$).

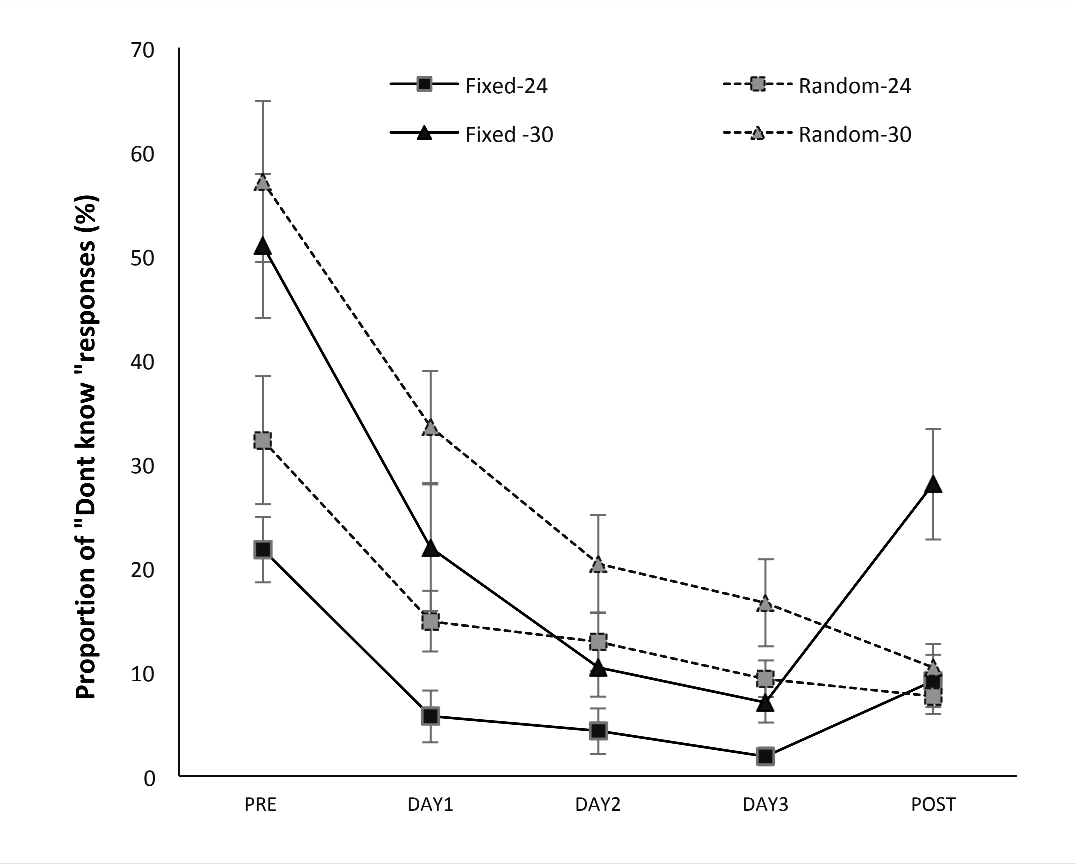


Fig. 5.23 Proportion of "Don't known" responses as a function of babble noise training , plotted separately for each of fixed (n=10) and random babble trained with SNR -24 dB (n=10), fixed (n=10) and random babble trained with SNR -30 dB (n= 9). Error bars reflect ± one standard error of the mean.

A mixed ANOVA was used to test listeners' reduction in "Don't know" responses across test sessions with different training methods (fixed and random babble noise) at the two SNR levels (SNR -24 and SNR -30). Results showed that there was a significant decrease across time sessions (time: $F_{3,\,105} = 32.74$, $p < 0.001$). There was also a significant interaction between both SNR and time (SNR × time, $F_{3,\,105} = 3.70$, $p < 0.05$) and between noise and time (noise × time: $F_{3,105} = 3.70$, $p < 0.05$). However, no significant interaction was observed between test period, SNR, and babble noise type (time ×SNR ×noise: $F_{3,105} = 1.94$, $p > 0.05$).

5.4.3 Learning outcomes

According to the comparison between VCV study one and study two, the outcomes can be summarized under the following three main points:

- Correct responses for pre- and post-test: Participants' performance improved significantly for both fixed and random babble noise training at the two noise ratios (SNR-24dB and SNR-30dB) from pre-test to post-test. VCV identification accuracy at the SNR -24 was much better than at the SNR -30, and this was true for both fixed and random training groups. Therefore, SNR makes a VCV identification performance difference for both types of the training methods used.

- Improvement in correct responses from pre- to the post-test session: The VCV identification correct responses improvement (from pre-test to post-test session) for the random babble noise training was much higher than for the fixed babble training group, this was true at the SNR -30dB level used in the test. Therefore, random training leads to better VCV identification performance than a fixed training method for the SNR-30dB. Comparing the improvement for the fixed training (M=8.75%) and control group (M=5.08%), there was no significant difference between these groups. It is therefore suggested that training with fixed noise has no transfer learning effect when testing with random noise. In contrast, the random training group (M=18.36%) had the largest improvement compared to the performance improvement for both the fixed and control groups. It indicated that changing background noise from random (in pre- and post test sessions) to fixed (training session) did affect the perceptual learning process. It further confirmed that stimulus differences across test sessions lead to different performance improvements.

- Training session results with different SNR levels (day1-day3): Participants' correct response performance improved significantly for both fixed and random training methods at the two signal to noise ratios (SNR-24dB and SNR-30dB) from day 1 to day 3. VCV identification accuracy at the SNR - 24 was much better than it at the SNR-30 from day1 to day3 and this was true for both fixed and random training groups. Therefore, SNR makes a VCV identification performance difference in the training session for both types of training methods used.

- "Don't know" responses: The "Don't know" responses also demonstrated that listeners' post-test "Don't know" performances had decreased compared to their pre-test "Don't know" results for both fixed and random babble noise training at the two noise ratios (SNR-24 dB and SNR-30 dB). The "Don't know" performance also demonstrated that the random trained had larger decrease in "Don't know" responses than the fixed trained group for both SNR levels (SNR -24 dB: random 24.61% vs fixed: 12.58 %; SNR -30 dB: random 46.70% vs fixed 22.89%). The decrease amount in "Don' t know" responses for the control group (7.90%) was lower than that for both of the fixed and random groups. However, as the pre "Don't know" performance for control group was also the lower than for both fixed and random, the comparisons of the reduction in "Don't know" performance for control with the fixed and random can not be interpreted reliably in the chapter.

5.5 General discussion of the VCV studies

These two studies demonstrated that listeners' performance significantly improved between pre and post VCV random babble noise tests for both training methods used. Comparing the performance improvement from pre-test to post-test between the fixed babble noise trained group and the random babble noise trained group, the results in the first VCV study showed that random babble noise training showed larger improvement than fixed babble noise training. However, there was a poorer pre-test identification performance in the random trained group compared to learning with the fixed sample of babble noise trained group. Therefore, these results cannot confirm whether fixed training or random training is better for identification performance improvement.

The second VCV study made the task more difficult by reducing the SNR (from -24 dB to -30 dB) and investigated if this led to any performance improvement difference between random and fixed babble noise training. It also added a control group without any training to explore whether the learning effect from training listeners with VCV against fixed babble noise generalized to a VCV against random babble noise condition. Results from the VCV study two confirmed that random babble noise training produced better identification performance than learning with a fixed sample of babble noise. The "Don't know" responses from both VCV studies also demonstrated that the random trained had larger decrease in "Don't know" responses than the fixed trained group. From VCV study two, the improvement in correct responses (from pre- to post-test) for the fixed training group was not significantly different from the improvement for the control group. Therefore, the listeners' VCV identification performance from training with fixed babble noise did not generalize to the VCV identification condition with random babble noise. However, the performance improvement (from pre- to post-test) for the random trained group was significantly greater than the improvement for the control group. Therefore, training with random babble noise was more effective than training with fixed babble noise. Results from the comparison between guess rate (the amount of decrease in “Don’t know” performance / 8) and the improvements in the percentage of correct responses confirmed that participants’ performance improvement was not due to changes in response criteria (stimulus detection and identification), but due to listeners’ perceptual learning. More information about stimulus detection and identification will be discussed in the following sections.

5.5.1 Performance improvement for test groups in terms of learning types

As mentioned previously in Chapter 2, there are three different types of learning that lead to performance improvement on auditory tasks: stimulus learning, task learning and procedural learning (Ortiz & Wright, 2009). Stimulus learning is defined as learning specific characteristics of the stimulus used during training (purely learning from trained and evaluated using the same stimulus), which can also be obtained from a stimulus that shares a particular feature with the training stimulus (transfer learning or generalization, trained and evaluated using different stimuli) (Ortiz & Wright, 2009). Task learning is considered to be learning associated with the particular auditory decision (e.g. sound detection or discrimination) to be made (Robinson & Summerfield, 1996; Ortiz and Wright, 2009). Procedural learning is the learning of the test components (i.e. experimental setting, the test methods, test response requirements and general strategies for doing the test tasks), but excluding the experience resulting from trained tasks and stimuli (Robinson & Summerfield, 1996; Ortiz & Wright, 2009).

In order to explore the three types of learning, Ortiz and Wright (2009) evaluated participants' performance on a target ITD discrimination task (two 300 ms 0.5 kHz tones) with either two hours training (three groups) on the target ITD task that shared some features, or no training (one group). In their experiment, they had four groups, 1) the ITD trained group (n=14), trained on the target ITD discrimination task and named target trained group; 2) the interaural level difference (ILD) discrimination trained group (n=18), trained on two 300 ms 4 kHz tones that shared both procedural properties and the lateralization task, but not stimulus with the target trained group; 3) The temporal-internal discrimination group (n=17), trained on two 15 ms brief 4 kHz tones with temporal interval of 100ms or random interval with 100ms plus some positive variable interval. The task setting for this group was different from the target trained group in both task and stimulus, but in common with procedural; 4) no trained group (n=94). Results showed that both stimulus and procedural learning contributed to performance improvements on ITD discrimination (Interval-trained group vs no trained group). But task learning did not contribute to improvements on ITD discrimination (ITD trained group vs ILD trained group).

In the current VCV experiments, there are two components of stimulus learning: one is the target (VCV) stimulus learning and the other is the babble background noise learning (Fixed or Random babble noise depends on attribution of the trained groups). However, Ortiz and Wright (2009) used two types of tasks, only one type of task used for the current VCV experiment. The task learning here refers to the identification of the nonsense syllables in babble noise task. Also, the procedures in the VCV experiment (familiarity with the lab environment, keyboard, PC screen, using the mouse and familiarity with the test requirement) were common across all groups of participants.

The VCV study two in this chapter revealed that there were significant changes in correct responses that occurred from pre- to day1 for fixed and random training groups (all $p < 0.05$, see section 5.3.2.3), and pre to post for control group ($p < 0.05$, see section 5.3.2.1). They were all likely attributed to rapid learning. This rapid learning was likely due to both procedural learning (such as the experimental setting, test method, test requirements and so on) and stimulus learning (Demany, 1985; Hawkey *et al*., 2004). As both stimulus and procedural learning contribute to rapid learning, once procedural learning has been completed, the stimulus learning will dominate the learning process (Hawkey *et al*., 2004). In order to separate stimulus and procedural learning during rapid learning, Hawkey *et al.* (2004) included a trained group with a different procedure to their target trained group (trained on target task). However, due to the design of current VCV experiments (no different procedure group), it is not possible to state how much rapid learning belongs to procedural learning or stimulus learning.

In the second VCV experiment design, all of the three groups received the same pre- and post-test session (VCV in random babble noise), the control group received no training and only completed the pre and post trials. Therefore, any improvement in performance from the control group can be considered to be a baseline level of learning from completing the pre and post tasks. Improvement in performance especially due to training sessions (for fixed and random trained groups) should be considered to be any improvement over and above that shown by the control group.

Although it is not possible to separate out procedural, task and stimulus learning within each condition, it is possible to consider what type of learning may have contributed to the performance improvement differences across three groups. However, based on this experiment design, as the same procedural and task are used across all the test sessions for all the participants, it is hard to separate the procedural, task and stimulus learning for the additional improvements between trained groups (fixed or random group) with the control group. Therefore, only the types of learning for additional performance between fixed and random group is analysed in the following session. Regarding the improvement differences between the fixed and random group, it was due to different training babble background noise-induced learning. Also considering the random and fixed group, performance improvement for the random group was much larger than for the fixed group. Thus, the observed additional improvements in performance between random and fixed group appears to be due to learning of random noise rather than to task familiarity. Details of the comparison about the performance improvements for each group in terms of learning types were listed in Table 5.5.

Table 5.5 Comparison of performance improvement differences for each group in terms of learning types (SNR-30)

	Group	Procedural learning[1]	Task learning	Stimulus learning			Results	
				VCV stimulus	Babble background noise		Increase in correct responses (pre-post)	Reduction in Don't know responses (pre-post)
					Random Babble	Fixed Babble		
Number of trials	Control	128 (Pre) 128(post)	128 (Pre) 128(post)	128 (Pre) 128 (post)	128 (Pre) 128 (Post)	0	5.08%	7.9% [2]
	Fixed	128 (Pre) 1920 (Training) 128 (post)	128 (Pre) 1920 (Training) 128 (post)	128 (Pre) 1920 (Training) 128 (post)	128 (Pre) 128 (post)	1920 (Training)	8.75%	22.89%
	Random	128 (Pre) 1920 (Training) 128 (post)	128 (Pre) 1920 (Training) 128 (post)	128 (Pre) 1920 (Training) 128 (post)	128 (Pre) 1920 (Training) 128 (post)	0	19.62%	46.70%
Actual Outcome	Fixed vs Control	-	-	-	-	-	Fixed not significantly different from Control ($P > 0.05$)	Larger reduction in Fixed compared to Control ($P < 0.05$)
	Random vs Control	-	-	-	-	-	Larger improvement in Random compared to Control ($P < 0.001$)	Larger reduction in Random compared to Control ($P < 0.001$)
	Random vs Fixed	No	No	No	Yes (Random > Fixed)	No	Larger improvement in Random compared to Fixed ($P < 0.05$)	Larger reduction in Random compared to Fixed ($P<0.05$)

Note* 1 Learning the computer interface and requirements of experiment; 2 Pre "Don't know" results significantly lower than those observed for fixed and random groups.

5.5.2 Consideration of “Don’t know” responses

The “ Don’t know” response was used in the experiment design, which is different from most of other auditory training research in the literature without this choice, such as psychoacoustic studies that use closed set tasks by forced choices (Amitay, *et al*., 2005; Fitzgerald & Wright, 2011) and speech studies that used open set sentences (Felty, *et al.* 2009; Fu, *et al.* 2005, Stacey & Summerfield, 2007). There are several advantages of including the “Don’t Know” response in the experiment design.

First, the “Don’t Know” response gives another measurement of performance that is not necessarily linked to correct responses. In order to explain reasons for performance difference, it is assumed that there are two processes for the perceptual learning of VCV in babble noise task, one is stimulus detection, which relates to detect a signal sound and the other is stimulus identification, which refers to making a decision. The two processes occur at the same time for the same stimulus, and operate concurrently. However, in the case where no stimulus is detected, participants were allowed to use the “Don’t know” response option that means they don’t have to guess. When the participants detect the VCV stimuli they will make choices among the eight consonants and choose less “Don't know” responses. Therefore the stimulus detection performance can be showed by the “Don’t know” responses. The stimulus identification can be reflected by how many correct decisions the participants made. As shown in both VCV studies (see Fig. 5.23), the “Don’t know” response was similar at both the pre and day1 for fixed and random group. But the “Don’t know” responses at the day 2 and day 3 showed that the fixed group had lower responses than the random group. The post session showed the random training had lower “Don’t know” responses than the fixed training group. This demonstrated that fixed group could detect more VCV stimulus than the random group and they are more confident than the random group at Day2 and 3.Then in the post-test session, the random group showed more confidence. This reflected that changing background noise for the fixed training group at post-test session reduced participants’ performance confidence.

Second, the “Don’t know” choices added extra information to understanding perceptual learning process. The “Don’t know” performance demonstrated that the random trained had larger decrease in “Don’t know” responses than the fixed trained group for both SNR levels (SNR -24 dB: random 24.61% vs fixed: 12.58 %; SNR -30 dB: random 46.70% vs fixed 22.89%). It was further noted that there were incorrect response differences between the fixed and random groups at each test session. It was observed that a lower percentage of “Don't know” responses, generally resulted in a higher percentage of incorrect responses. For example, the post-test results showed that the random group had less “Don’t know” responses than the fixed group, but the random group subsequently had more incorrect responses than the fixed group (See Appendix 6, Table 1 and 2). The levels of confidence in hearing the consonant differed between fixed and random group led to those two groups performing differently in making responses. It indicated that changing stimuli across test sessions affected listeners’ choices in responses. In addition, results from comparison between the guess rate (the amount of decrease in “Don’t know” performance/8) and the improvements in the percentage of correct responses from both VCV studies confirmed that participant’s performance improvement was not due to changes in response criteria, but due to listeners' perceptual learning.

Finally, the "Don't know" responses could give participants a choice for cases they did not detect the signal sound. As the listeners may feel at the beginning that it is hard to detect any VCV stimuli, it could potentially be more demotivating for them if they have to make random guesses. If participants are not motivated then they won't do the task to the best of their ability. Including "Don't know" choice in an experiment design can stop participants from being forced to guess when they cannot hear the signal. The stimuli used in this experiment had quite low SNRs (-24dB and -30 dB). Hence, although the instruction to participants was to guess if they are not sure, participants may not have noticed there was a signal sound when they did the initial few trials. Therefore, instead of forcing them to choose one of the eight consonants, they can use the "Don't know" response.

The main disadvantage of "Don't know" responses is that it may reduce the number of correct responses and increase participants' decision bias at initial stages of the learning process. The bias can be made between the eight consonant choices and the "Don't Know" response. That means participant may detect there is a signal, but they are not sure which one it is, then they may still choose the "Don't know" response. Similar to the open set tasks in which participants report words or sentences, participants may end up of using don't know responses when they missed the words or part of the sentences. Without providing a "Don't know" option in open set tasks, some of the participants who are very confident may guess what the words are and some of them who are not confident may end up with no answer (a zero score). In addition, psychoacoustics studies used closed set tasks in which participants are forced to make a choice with a guess if they could not detect the signal stimulus. It increased both the correct response rate and the response bias of guess rate. Further studies could investigate comparing forced choice tasks versus responses with a “Don’t know” option to explore which one is more effective and accurate for performance learning improvement.

5.5.3 Reasons for performance differences across groups

Comparing the improvement for the fixed training (M=8.75%) and control group (M=5.08%), there was no significant difference between fixed training and the control group. The fixed group had spent three days making 640 VCV identification responses each day to the same section of babble stimuli, but this didn’t produce any significant improvement over the control group in the post-test session. It suggested that training with fixed noise has no learning effect when testing with random noise. In contrast, the random training group (M=18.36%) had the largest improvement compared to that for the fixed and random training groups. The random group trained over three days making VCV identification to different, randomly selected sets of babble stimuli, with the improvement in the post-test session indicating that perceptual learning of stimulus identification is enhanced by VCV in random background noise training.

Across test time periods (pre-test, post-test, and training sessions), for the fixed babble noise group, although the target stimuli were the same, the training background noise stimulus (fixed babble noise) differed from the background noise stimulus used in the pre- and post-sessions (random babble noise). However, for the random group, both the target and background stimuli were the same in all three sessions (pre-, training and post-test session: VCV in random babble noise). As the consequence of the experiment design differences, it may lead to attention difference, which is caused by the trained listeners' attention differences by focusing on different elements when they did the VCV tasks with fixed or random background noise. One simple possibility is that fixed babble noise affords more opportunities than random noise for 'glimpsing' (listening in the comodulated or uncomodulated dips; Rosen *et al.*, 2013). It is possible that in learning the fixed babble listeners are better able to anticipate when dips in the background noise occur. Thus they can attend to these time points and spectral frequency regions that have a more favorable signal to noise ratio. The same section of background babble noise was presented in test trials for the fixed babble noise trained group, this might have made the test background noise easier to adapt to, so that the listeners in the fixed babble noise trained group could focus their attention effort on listening to the dip.

For the random training group, the background noise changed randomly across the test trials. In this case, trained listeners might focus on learning to be familiar with the stimulus cues of random background noise (i.e. noise cues such as number of speakers, frequency distribution of the voices, etc.) to allow them to achieve a better performance of identifying syllables in noise. Comparing fixed and random background noise, the noise patterns for the fixed training group across training sessions are more predictable sounds than the ones for the random training group. Repeated noise exposure induces rapid learning easier than unpredictable sounds (Agus, Thorpe and Pressnitzer, 2010). In their study, there were three types of noise,

1) running noise (N); 2) half duration of running noise repeated twice, repeated noise (RN); 3) one particular exemplar of RN reoccurred in trails, they named it reference repeated noise (RefRN). Participants were required to listen to detect which noise included a repeated sample. These three types of noise were randomly presented across trials within each experimental block. No feedback was given and the RefRN noise was not displayed on two consecutive trials during the test. Results showed that listeners were better at detecting the RefRN noise than both the N and RN sounds. It indicated that repeated noise exposure induced learning faster than the unpredictable noise. In this case, repeated exposure to the same background noise is likely to increase rapid learning process at the beginning, listeners in the fixed training group might feel it is easier to adapt to the fixed background noise than listeners in the random training group. Comparing the training session for fixed and random training group, the “Don’t know” response was similar at day1 for fixed and random group. But the “Don’t know” responses at the day2 and day3 showed that the fixed group had lower responses than the random group. However, comparing the correct responses across test sessions between fixed and random group, it showed that both group had similar correct response across day1 to 3.

The initial improvement found in both training groups might be an effect of procedural learning, but after listeners were familiar with the procedure, the later perceptual learning may be due to improved auditory ability to extract speech sounds from the background noises (Francis & Nusbaum, 2002). The auditory ability to extract speech signals from background noise is highly relevant to test stimuli. Results from the two VCV experiments showed that although both the random and fixed trained groups improved, the same improvements were not found between these two different training groups. As described in the previous test method sections (5.2.1 and 5.3.1), for the random babble noise trained group, although different stimuli background noise were presented through the whole experiment, there was no task variety or change in stimuli setting (background noise was always random babble noise). However, across the test sessions for the fixed babble trained group, there was no variety in the task, but there was a change in stimuli setting. So participants in the fixed trained group may need more time to fine-tune their listening ability to the change of auditory stimuli and detect the VCV stimuli in the post-test. Earlier perceptual speech training studies suggest that depending on the specific signal and background noise, auditory perceptual training may adjust the auditory system by increasing awareness of informative signal cues, decreasing the influence of less useful stimuli, or both (Francis, Nusbaum, & Fenn, 2007; Francis & Nusbaum, 2009). For random-noise trained listeners, doing the same task and using the same background noise stimuli setting (random section of babble noise per trial), VCV identification performance increased from both their pre-test to day one training session and day three training session to post-test. While the fixed-noise training participants' speech performance increased from pre-test to day one training session but decreased from day three training session to post-test (VCV with SNR-30). Therefore, it is easier for participants to switch from random babble noise to fixed babble noise, but harder for them to do the shift in the opposite direction.

5.5.4 Comparison with previous studies in the literature

Felty *et al*. (2009) used one short training session test to compare word performance with fixed and random babble background noises. They demonstrated that listeners obtained better word recognition performance with fixed babble noise than with random babble noise. The word recognition with fixed (named ‘frozen’ in their paper) babble noise (M= 57.7 %) was 9.7% larger than with random babble noise (M=48.0%) group. The experiments in this Chapter used multiple training sessions and evaluated with either same or different background to explore perceptual learning from both learning effect and generalization effect. The VCV experimental design in this study tested with random noise for both the pre- and post-test session. However, Felty *et al*. (2009) investigated the performance adaption differences between fix and random babble, which both training and evaluated background noise were the same. There were no pre-, post-tests or training sessions incorporated into their experimental design. Participants are both trained and tested on the same background noise (Fixed or Random). In contrast the VCV in noise task is investigating transferability between training on fixed and random babble noise. The VCV experiment design began with a random babble test for everyone, and this represented a certain amount of the initial learning period (affecting stimulus learning) that might have impacted on subsequent training session results observed for the fixed babble noise group. Otherwise, without this initial test better performance improvements might have been observed for the fixed babble group.

Another possible reason for the difference between these VCV studies and Felty *et al*. (2009) may relate to target sounds used in the two VCV studies. The open set word stimuli used in the study by Felty *et al*. (2009) were much more informative with communication meanings and also reduced the guess rate for responses. These reasons may have allowed for more learning benefits compared to the nonsense VCV stimuli used in this study. So it is possible that differences between VCV stimuli and the words used in this study led to identification differences in fixed and random noise conditions. In order to confirm this possibility, further experiments with purely fixed and random training should be conducted.

The outcome of this study also differed from perceptual learning findings in the visual domain, which demonstrates that visual perceptual learning ability could generalize from training with fixed noise to the identification of targets masked by random noise (Schubö *et al.*, 2001). However, signal processing in hearing is different from that in the visual domain. In particular, the signal and masking stimuli used in the visual study were separated in time (a backward or forward mask), while for the VCV study, the target sound and background noise were mixed together to be displayed at the same time for each trial. This may be a reason for the different findings of the visual study, and of this auditory perceptual learning study.

5.6 Summary

Two experiments examined auditory perceptual learning for nonsense syllable VCV stimuli following training with stimuli embedded in fixed and random babble noise. The results suggest that auditory training with random babble noise using a simple VCV task can improve people's stimulus identification performance in difficult listening conditions. As words and sentences are more informative than nonsense syllables, further training interventions may benefit more from training using real world sounds (such as words or sentences) with random babble noise. In addition, as normal hearing listeners participated in this experiment, it may be worthwhile to identify whether this training method could be applied in clinics to help train hearing-impaired people. For example, a perceptual learning study from Burk *et al.*

(2006) demonstrated that both older hearing-impaired people and young normal hearing people improved their speech in noise performance after training with repeated presentation of the test words. But older hearing people required a more advantageous SNR and more training time to improve their performance in the same way as that of the young normal hearing people. Therefore, the ability to understand speech under challenging listening conditions among hearing-impaired listeners, may also benefit from training speech tasks with similar random noise conditions with higher SNR and a longer training time.

Chapter 6 Single session study of nonsense stimulus recognition with fixed and random babble noise

6.1 Introduction

The VCV study in Chapter 5 investigated the training effects of fixed and random babble noises. The fourth experiment in this Chapter will look at a single session study to explore fixed and random noise adaption for the VCV method. As shown in a previous study from Felty *et al.* (2009), listeners achieved better word recognition performance with fixed babble noise than with random babble noise when they were trained and tested in a single session with the same patterns of background noise. In contrast, the study in chapter 5 has shown that training with fixed noise does not transfer as well to a random noise task as training with random noise.

Several experiment design differences were identified between this auditory perceptual training study and Felty *et al.* (2009) (these differences can be seen in Table 6.1). First, the test methods used are different; Participants in Felty *et al.* (2009) are both trained and tested on the same background noise (fixed or random noise), while the VCV experiment for the current project in Chapter 5 began with a random-babble test for everyone. This experimental difference could influence participants' initial learning performance. Second, there were differences in the test materials of these two studies. VCV stimuli were used as target sounds in the current study, which may have been too simple for listeners to learn from listening 'in the dips'. In contrast, Felty *et al*. (2009) used word stimuli as speech sounds, which are much more informative than nonsense syllables to produce a learning benefit of listening 'in the dips'. The concept of ‘dip listening’ can be found in Duquesnoy (1983). Duquesnoy (1983) demonstrated that normal hearing people can extract target information from background noise by listening to temporal dips in the fluctuating background noise. This process especially occurred at the condition when the mean target sound level was lower than the mean background noise level. Bernstein and Grant (2009) compared normal hearing and hearing-impaired listener’s speech intelligibility in noisy conditions (stationary noise, interfering male talker noise and speech-modulated noise), results indicated that there was an SNR dependence of the fluctuating-masker benefit for both NH and HI people. The more negative the SNR (SNR range: -30dB to 10dB) was, the more benefit from fluctuating masking noise was. So differences in SNR may be the third reason for the different findings between VCV study two in Chapter 5 (SNR -30dB) and Felty *et al.* (2009) (SNR 0, 5, 10dB).

Table 6.1. Experiment design differences: VCV vs Felty *et al.* (2009)

	Felty *et al.* (2009)	**VCV Experiment**
SNR(dB)	0, 5, 10	-24 & -30
Voice	Male	Male and Female
Materials	English Words	Vowel Consonant Vowel
SPL(dB)	77	65
Sessions	Single session with 357 trials	Three sessions with 2176 trials 128 (pre) + 640 (trained) * 3 days + 128 (post)
Evaluated method	Fixed group: words in fixed babble noise (same as tested) Random group: words in random babble noise (same as tested)	Fixed group: VCV in fixed babble noise (different from training session) Random group: VCV in random babble noise (same as the training session)

In this chapter, a follow-up study was carried out using a single session experiment with VCV stimuli in both fixed and random babble background noise. There were two objectives for this follow-up study, one was to investigate whether test method (multi-sessions: trained and tested with same noise for random group or different noises for fixed group vs. single session: trained and tested with same background noise for both fixed and random group) differences led to result differences in VCV study two (Chapter 5), and the study in this chapter. The other objective was to compare listeners' performance from the fixed babble noise and random babble noise group in order to find out which background noise led to better performance.

6.2 Test methods

6.2.1 Participants

Twenty-five volunteers (10 males and 15 females) with normal hearing participated in this experiment. All of the participants had no prior experience participating in sound experiments, and their pure tone thresholds were less than 20 dB HL. The age range was from 18 to 36 years old with a mean age of 23 years. The participants were all volunteers recruited from the student and staff population of The University of Warwick. Details of the participant' requirements are described in Chapter 3, section 3.2.

6.2.2 Experiment design

Twenty-five participants were randomly divided into a fixed babble noise (n=12) or a random babble noise group (n=13). As this was a short session study, only the female voice was used in the experiment. The rest of the experimental protocol in this chapter was similar to the practice and pre-test sessions of the VCV experiments in Chapter 5. A practice session was held to enable the participants to become familiar with the experiment before they participated in the main session. Participants were required to do a practice test for the VCV task (including eight consonants /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/) with female voices without any background noise first. Listeners then completed a test session lasting approximately half an hour, which included five female voice blocks with either fixed babble noise or random babble noise, depending on the group they were in. Each block of the VCV task contained 64 trials (including eight consonants /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/ in random order) presented in an /a/-consonant-/a/ format with female voices in fixed or random babble noise depending on the test sessions and groups. No feedback was provided, and participants were encouraged to guess if they were not sure. Participates were informed to select a “Don't know” response if they were really unsure of the correct answer. The proportion of correct responses for the VCV task with babble noise was calculated as a measurement of participants’ performance.

6.2.3 Stimuli

The SNR for this present VCV single-session study was fixed at -30 dB throughout the test sessions. Except for the SNR and female only voice used for this study, all the other test materials and the interface were the same as the previous experiment in section 5.2.1, Chapter 5.

6.2.4 Data analysis

In this study, no datasets were removed (i.e. identified as outliers) from the collected data. A mixed ANOVA was carried out to investigate both the VCV identification accuracy and the VCV "Don't know" responses for fixed and random babble noise groups across the five blocks. A table showing the proportions of correct, incorrect and don't know responses for each condition is provided in Appendix 6, Table 3.

6.3 Test results

Figure. 6.1 displays the VCV identification performance accuracy of participants, comparing fixed and random babble background noise over a 64-trial window for each block. The first point represents the average performance of listeners from trials 1 to 64, the second point from trials 65 to 128, and so on. A mixed ANOVA was used to assess the VCV identification performance of the two different groups in babble noise across the intervention of time period (block1 to block5). The main effect comparing the overall VCV identification correction between the two groups was not significant (group, $F_{1,23} = 0.015$, $p > 0.05$), indicating that the fixed babble noise group (mean= 16.90%, SD= 7.52%) had a similar VCV identification accuracy to the random babble noise group (mean=17.21%, SD=6.42%). A significant difference was observed among the test blocks (time: $F_{4,92} = 28.85$, $p < 0.001$). In a comparison of adjacent blocks, only block1 to block2 showed a significant difference ($p < 0.05$), all the other blocks were not significantly different from each other (all $p > 0.05$). These results suggest that although identification accuracy was improved across the test blocks of the experiment, there was no significant improvement from one block to the next (except block1 to block2). There was no significant interaction between time period and group (time × group interaction, $F_{4, 92} = 0.35$, $p > 0.05$).

Table 6.1 The proportion of correct responses and "Don't know" responses from test Block1 to Block5 for the fixed and random babble noise groups (averaged across eight consonants /d,f,g,k,m,n,b,p/).

Unit (%)	Group	Block1	Block2	Block3	Block4	Block5
"Don't know"	Fixed	41.28	16.02	12.76	13.67	12.11
	Random	57.57	50.24	43.39	42.07	36.66
Correct	Fixed	8.98	13.02	18.10	20.96	23.44
	Random	9.62	14.42	16.71	20.43	24.88

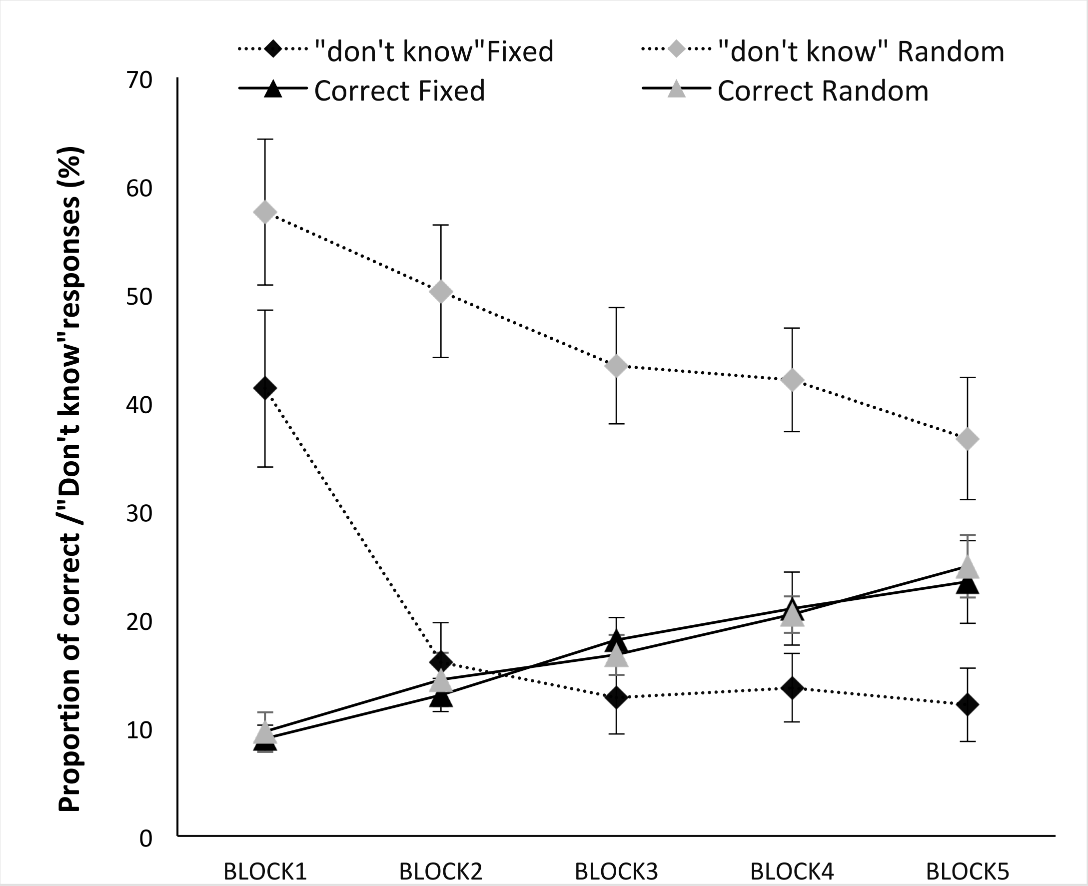


Fig. 6.1 The proportion of correct responses or "Don't know" responses from test Block 1 to Block 5 for the fixed and random babble noise groups (averaged across eight consonants /d, f, g, k, m, n, b, p/). Each point corresponds to the mean correct percentage correct for all subjects in the respective condition over a 64 trial window for each of the test blocks (Block 1: 1-64; Block 2: 65-128; Block 3: 129-192; Block 4: 193-256; Block 5: 257-320). Error bars reflect ± one standard error of the mean.

A mixed ANOVA was used to assess the VCV "Don't know" responses of the two different babble noise groups across the intervention time period (block1 to block5). The main effect comparing the overall VCV "Don't know" responses between the two groups was significant (group, $F_{1,\,23} = 17.01$, $p < 0.001$), indicating that the fixed babble noise group (mean= 19.17%, SD= 12.45%) had a much lower VCV "Don't know" response than the random babble noise group (mean= 45.97%, SD=7.23%). A significant difference was observed among the test blocks (time: $F_{4,}$

$_{92}$ = 26.21, $p < 0.001$). There was also a significant interaction between time period and group (time × group interaction, $F_{4,\ 92} = 3.13$, $p < 0.05$).

Fixed babble noise group: A one way repeated ANOVA was carried out to analyse VCV "Don't know" responses for the fixed babble noise groups across block1 to block5. Results showed that listeners' performance with fixed babble noise was significantly different from block1 to block5 (time: $F_{4,\ 59} = 12.82$, $p < 0.001$). The following pairwise comparison indicated that there was a sharp decrease for VCV "Don't know" response between block1 and block2 ($p < 0.05$), all the other following paired blocks showed no significant difference between each other (block2 vs block3, block3 vs block4, block4 vs block5, all $p > 0.05$). This indicated that after block 2, listeners might have achieved their asymptotic performance during the training session.

Random babble noise group: Listeners' "Don't know" responses with random babble noise were significantly different from block1 to block5 (time: $F_{4,\ 64} = 1.97$, $p < 0.05$). The following pairwise comparison showed no significant difference between consecutive blocks (block1 vs block2, block2 vs block3, block3 vs block4, block4 vs block5, all $p > 0.05$), but there were significant differences between block1 and block4, block1 and block5, block2 and block5 (all $p < 0.05$). It is therefore suggested that for the random group, the "Don't know" response rate was reducing steadily across test blocks.

6.4 Discussion

This study demonstrated that listeners' VCV identification performance significantly improved across test sessions from block1 to block5 for both the random and fixed test groups. Comparing the average of VCV recognition performance between the fixed babble noise group and the random babble noise group, the listener's performance from fixed babble noise was similar to the VCV identification condition with random babble noise. Therefore, this information confirmed that test method differences (i.e. multi-sessions: trained and tested with same noise for random group or different noises for fixed group vs. single session: trained and tested with same background noise for both fixed and random group) led to differences in the findings between VCV study (Chapter 5) and the study in this chapter.

Although nonsense syllables can be considered as the building blocks of language, research shows that comprehension of speech consistent with a lexically driven way of learning. Word-based training strategies provide more perceptual information than phoneme-based strategies (Davis Hervais-Adelman, Taylor, McGettigan, & Jonsrude, 2005). Auditory perceptual training and perceptual learning studies in hearing have demonstrated that training with word and sentence stimuli created better outcomes than training with nonsense syllables (Stacey & Summerfield, 2008). Speech perceptual differences between nonsense syllables and words may cause the auditory system to process such test tasks in different ways. Cooke (2006) and Vestergaard *et al*. (2011) demonstrated that 'dip listening' is considered as a masking release from the auditory system in fluctuating noise. Compared with faster fluctuating background noise, 'dip listening' is usually greater when the background noise fluctuates at a slower speed (Gustafsson & Arlinger, 1994; Bacon *et al*., 1998). Meaningful target information, which included more speech cues, can increase the probability to catch fluctuating background noise. As words are more informative than nonsense stimuli, learners may learn to 'listen in the dips' more effectively with words in fixed babble noise than in random babble noise as found by Felty *et al.* (2009). In the fixed noise condition, dips occurred in the same place on every single trial, so people could learn to listen at exactly those moments - i.e., they could adapt to fluctuation in the noise, and focus on the target sound because the background noise was fixed and predictable. However, nonsense syllables were too simple to make use of the benefits of ‘listen in the dips’ in babble noise conditions. Therefore, the identification performance of nonsense syllables in fixed and random babble noise conditions is similar. Hence, the target sound difference is a possible cause of the differences in results for the experiment that used nonsense syllables, compared to the study from Felty *et al.* (2009) which used word-based stimuli in their experiment.

The “Don’t know” responses from this study showed that the fixed group had much lower “Don’t know’ responses than the random group across test blocks. For the fixed group, there was a sharp decrease for VCV “Don’t know” response between block1 and block2. After block 2, listeners may have achieved their asymptotic performance for “Don’t know” responses. However, the amount of VCV “Don’t know” responses for the random group gradually decreased across test blocks. The random group had a higher proportion of “Don’t know” responses from the beginning than the fixed group, while the rate of decrease of “Don’t know” responses was stable across test blocks. Therefore, the “Don’t know” response pattern differences between fixed and random groups indicated listeners' confidence in what they heard and what responses they made. i.e. for fixed babble noise, participants rapidly became confident in what they thought they'd heard and the "Don't know" responses dropped off after the first block. On the other hand, the "Don't know" responses in the random babble noise group showed that the confidence in what they thought of they'd heard increased much slower. This is despite there being no difference in the actual correct response between both groups.

As the correct responses for both groups were similar, the resulting incorrect response rate had increased where the number of "Don't know" responses decreased. Results showed that the proportion of "Don't know" responses across test blocks for the fixed group was much lower than the random group. The number of incorrect responses across test blocks for the fixed was therefore higher than for the random group, due to both conditions having similar proportions of correct responses across blocks (See Appendix 6, Table 3). This indicated participants in the random group would rather not guess the consonant and make a wrong response, they preferred to choose the "Don't know" response. In contrast, the fixed group had higher levels of confidence in making a response, but this resulted in a greater number of incorrect responses. This result highlights that levels of confidence in hearing the consonant differed due to the background noise pattern differences between fixed and random noise, which led to those two groups to perform differently in making responses.

Based on the findings from Chapter 5, Felty, *et al*. (2009), and the results in this chapter, it is suggested that due to the lack of generalization from training fixed noise to random babble noise, if a study's purpose is to train people to understand stimuli in babble noise, then training with random babble noise is more effective. If the study's purpose is to train people to 'listen in the dips' during babble noise, which is a valuable listening strategy, then it's best to train them with fixed babble noise. But it's important to check if this then generalizes to random noise. A visual study conducted by Schubö *et al*. (2001) suggested that visual perceptual learning ability could generalize from training with fixed noise to the identification of targets masked by random noise, but a similar set up would have to be tested in the auditory domain, with temporally separated noise and target sound during the test.

6.5 Summary

In summary, based on this present study, VCV recognition performance with random babble noise is similar to the identification of VCV from fixed babble noise. The results confirmed that test method differences led to differences in the findings of VCV study two (multi-sessions, trained and tested noises same or different depend on random or fixed group in Chapter 5) and the study in this chapter (single session with trained and tested noise same for fixed and random group in this Chapter). For the next step in this thesis, it is suggested to use random babble noise to explore whether training with random babble noise can be generalized to other kinds of environmental background sounds, such as cars or rain.

Chapter 7 Generalization resulting from training of speech in babble noise to other background noises

7.1 Introduction

Auditory perceptual training has the potential to optimise the performance of hearing aid and cochlear implant users, and help them make full use of their prosthetic device (Sweetow & Sabes, 2006; Moore & Shannon, 2009). Apart from the training effect obtained from hearing perceptual learning studies, several studies suggest that auditory perceptual generalization has also been observed in both human (Nygaard & Pisoni, 1998; Delhommeau, Micheyl, Jouvent, & Collet, 2002) and animal studies (Delay, 2001). It can be explored from training with one set of familiar stimuli to untrained novel stimuli, such as amplitude modulated sounds, speech sounds, or real-world environmental stimuli (Tremblay & Kraus, 2002; Wong & Perrachione, 2007; Davis *et al.,* 2005). Better understanding of how and when auditory perceptual training generalises with normal hearing people will help devise better training for people with hearing impairment (Loebach *et al.,* 2009).

Although studies in auditory perceptual learning show a generalization effect from trained to untrained stimuli, they are mainly focused on changing the target stimuli using amplitude modulated sounds or speech stimuli (Hervais-Adelman *et al.,* 2011; Clarke & Garrett, 2004). Different studies use different background stimuli, babble noise or speech-shaped noise is commonly used. However, it is not clear whether training generalises to other types of noise, in particular real-world environmental noise such as car and rain noise. It has been argued that noise can limit the ability of the auditory system to process sounds (Corey & Hunspeth, 1983; Harris, 1968). However, Wiesenfeld and Moss (1995) demonstrated that the auditory system's sensitivity to weak signals can be increased by the addition of an appropriate amount of noise. Zeng *et al*. (2000) showed that noise could enhance listeners' ability to identify nonsense sounds. However, background noise is constantly changing in the real world. Therefore, the ability to detect speech signals in a noisy environment is critical in people's daily communication. Here we are interested in the effect of changing the background noise on auditory perceptual learning.

Evidence shows that different background noise can change the amount of perceptual learning. For example, Van Engen (2012) trained participants on English sentences recognition in three different background noise conditions: speech-shaped noise (SSN), Mandarin babble, and English babble. 56 participants were randomly assigned into 4 groups (three training groups: SSN Group, English babble group, Mandarin babble group and one control group without training). The training sessions took place over 2 days (30 minutes per day with four lists of BKB sentences) for three training groups. The control group just attended the post-test at the same day after the pretesting. The post-test was done with four lists of BKB sentences (two lists are training talker and two lists are novel talker) embedded with two types of background noise (English babble and Mandarin babble). Results showed that there were differences in the amount of performance observed after training across the different conditions. The post-test was only carried out with Mandarin or English babble. The post-performance for all the four groups was always better when tested with Mandarin babble than with English babble.

However, considering the post-performance for the English babble test only, the English training group had better performance than those who had trained with Mandarin or SSN. The results demonstrated that in order to improve people's speech perception in speech-in-speech environments, it is better to train them with speech informative sounds as the masking background noise than with non-speech stimuli. But also the noise 'structure' should be representative of the actual environment – i.e. training on mandarin babble and testing on English babble resulted in poorer performance. From perceptual learning in the visual domain, Schubö *et al*. (2001) showed that constant or random visual interferers (visual backward mask) could affect generalisation of visual perceptual learning, but less research has addressed generalisation of auditory perceptual learning. Felty *et al.* (2009) demonstrated that listeners obtained better word recognition performance with fixed babble noise than random babble noise. In contrast, the previous VCV trained experiments in Chapter 5 have shown that the VCV tasks identification performance against a random-noise background noise produced better learning effect (more improvement on VCV test with random noise) than against a fixed noise condition.

To date, even though previous studies have looked at perceptual training with environmental stimuli, these experiments have been done with environmental stimuli as target sounds such as footsteps, slamming door, air conditioner, dishwasher etc. (Reed & Delhorne, 2005; Kidd *et al.,* 2007; Burkholder, 2005), not as environmental background sounds. A study by Loebach and Pisoni (2008) trained 150 normal hearing participants to listen to cochlear implant simulated stimulus. They divided the listeners into five trained groups. Each group was trained with one set of test materials first and then tested participants' performance with all five (1.

words simple: Modified rhyme test; 2.words complex: Phonetically balanced words; 3. sentence meaningful: Harvard/IEEE sentences; 4. Sentence anomalous: Anomalous Harvard/IEEE sentences; 5. Environmental sounds) stimulus materials. Results showed that all five groups obtained significant improvement after training. The perceptual learning did not transfer from training on speech to the recognition of environmental sounds; however, the learning effect transferred from training on environmental sounds to both untrained environmental and speech sounds. This finding suggested that there are differences between the transfer of learning from speech and environmental stimuli. Following the study by Loebach and Pisoni (2008), Shafiro *et al.* (2012) did another experiment and they showed that perceptual learning generalized from environmental sounds to speech and novel environmental sounds in both the patterns of exposure (repeated short test) and training. However, the performance improvement was larger from the pattern of training than it was to the pattern of exposure alone (repeated short test). In addition, results from Shafiro *et al.* (2012) also demonstrated that the benefit of rapid performance improvement from auditory training could be retained over longer time periods. The sustained training effect was in line with earlier perceptual learning studies (Schwad *et al.* 1985; Francis *et al.,* 2007). Therefore, it will also be interesting to explore the lasting effects of perceptual learning.

There are two aims for this study: one is to investigate whether the transfer of perceptual learning will be observed from training with speech sounds against babble background noise to the perception of speech sounds against other real life environmental background noises such as car and rain noise. The other aim is to explore whether this transfer of perceptual learning is sustained over several weeks' gap. Because auditory perceptual learning studies have demonstrated that training outcomes are better with word and sentence stimuli than with nonsense syllables phonemes (Stacey & Summerfield, 2008), BKB sentences were used as the target sound in our experiment.

7.2 Test methods

7.2.1 Participants

All participants gave informed consent before participating in this study. Ethics approval for this experiment was given by the Biomedical and Scientific Research Ethics Committee (BSREC) of the University of Warwick. Twenty-four normal-hearing English native speakers (8 males and 16 females) participated in this experiment. All of the participants had no prior experience participating in psychoacoustic experiments, and their pure-tone thresholds (assessed by pure-tone audiometry) were less than 20 dB HL (BSA, 2011). The participants' age range was from 18 to 33 years old. The participants were all volunteers recruited from the student and staff population of the University of Warwick (for details of the participant requirements, see Chapter3, section 3.2).

7.2.2 Test stimuli

Bamford-Kowal-Bench (BKB) (Bench *et al.*, 1979) sentences recorded by a female British speaker were used as speech material. The speech material includes 21 lists and each list has 16 sentences containing a total of 50 target words. The sentences were centrally embedded in two seconds of background noise. The signal to noise ratio for each token was determined by comparing the root mean square average amplitude of the signal file with the background noise file (just the portion that actually overlapped with the sentence). The root mean square intensity was normalized to the same fixed value for all background noise.

Calibration was carried out before the main test took place. An IEC 711 acoustic coupler and a precision microphone were used to calibrate the output of the BKB sentence test. Then the maximum sound pressure levels from the PC were controlled to make sure the output from the software (MATLAB) was within exposure action value (65 dB SPL). The signal to noise ratios (SNR) used for this study were fixed for a given noise type, but varied for each noise condition: babble noise -20 dB, car noise -12 dB, rain noise -15 dB. The SNR noise levels were selected from a pilot study (n=8), in an attempt to obtain 50% correct target words identification with each of the background noises. Fig. 7.1 shows the waveforms and spectrums for an example sentence "The clown had a funny face" with three kinds of background noise. (a) example sentence in babble noise with SNR -20 dB; (b) example sentence in car noise with SNR -12 dB; (c) example sentence in rain noise with SNR -15 dB.

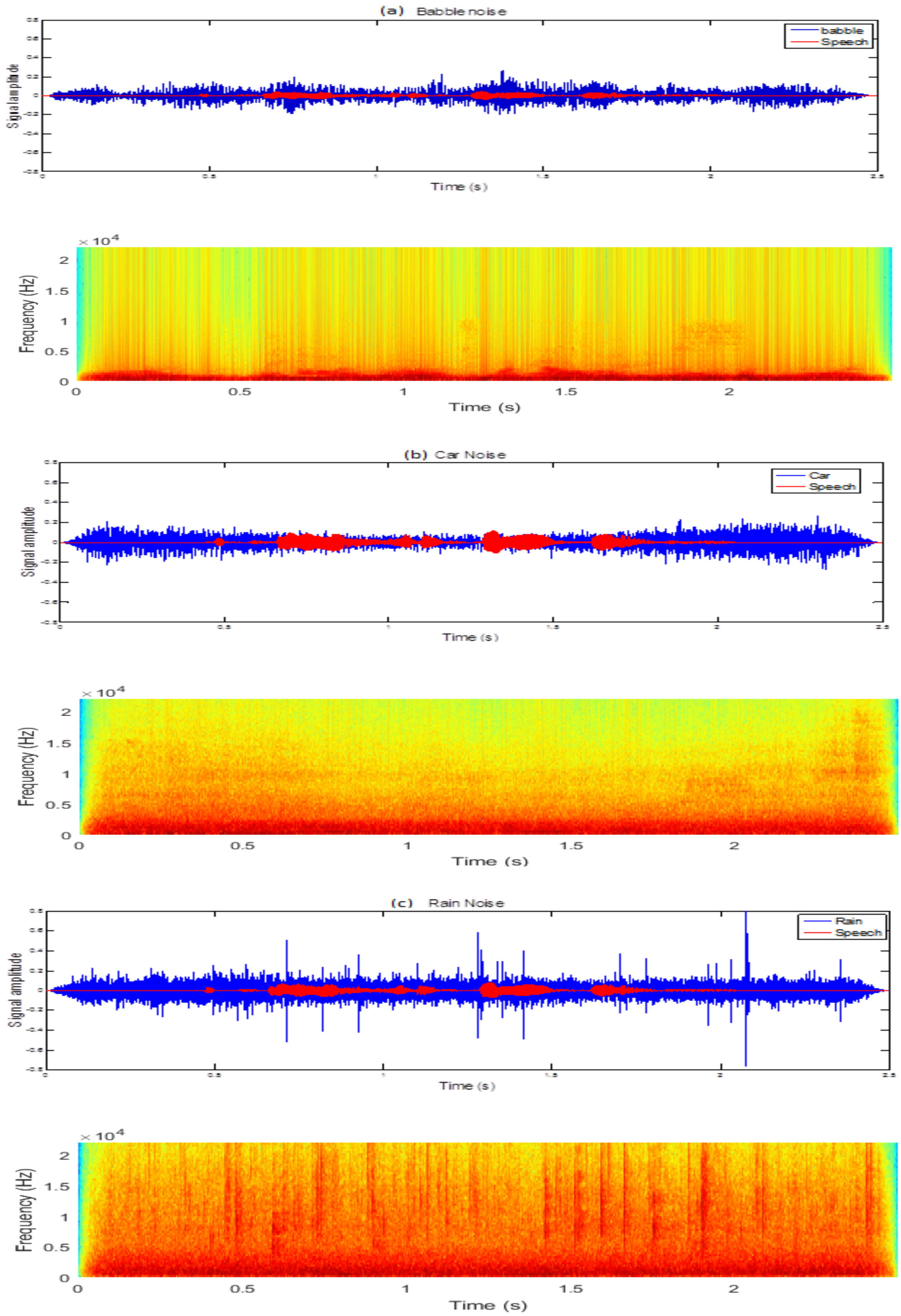


Fig. 7.1 Examples of a target sentence ("The clown had a funny face") in the background noise of babble, car and rain. The waveforms and spectrums are shown for (a) target sentence in babble noise with SNR -20 dB; (b) Target sentence in car noise with SNR -12 dB; (c) target sentence in rain noise with SNR -15 dB.

7.2.3 Experiment procedure

All tests were carried out in a sound-attenuating room. Participants were randomly assigned to either a control (n=12) or training (n=12) group. Before the test, a pure tone audiogram test was carried out to make sure the participant qualified to participate in the study. After that, the instructions for the experimental tests were given to the participant to read and ensure they understood the procedure. In order to let the participants be familiar with the experimental process, a practice session was given before the participants did the experiment trials. During the practice session, one example sentence was presented without background noise and the participants were required to repeat the speech sentences they heard. After that, they were required to listen to the babble, car and rain background noise samples, separately.

During the experiment trials, participants were told that the speech sounds would be softer than the background noise. Both groups were required to attend a pre-test and post-test session lasting approximately 5 minutes. They were encouraged to guess even if the sentences they repeated would result in a nonsense or incomplete sentence. The pre- and post-test session included one BKB sentence list (each list includes 16 sentences) with random babble noise, one with random car noise and one with random rain noise. The order of the three noise conditions was randomised in the pre- and post-tests but the BKB sentence list was the same across test participants. The training group attended three consecutive daily (half an hour) training sessions with BKB sentences presented amid random babble noise between the pre- and post-sessions. No feedback was given across the pre-, training and post-test sessions. Different BKB sentence lists were used for the pre-test (list 1-3), training (list 4-15) and the post-test (16-18) sessions.

The final follow-up evaluation session was carried out to test how the learning and generalization effect was retained. Participants were recalled back 8 to 18 weeks (the t-test showed that there was no significant difference for the time gaps between trained and control group) after the post-test session completed. Apart from the BKB sentence lists (list 19-21 used in the follow-up session), the procedure in this test session was the same as the pre- and post-test session. As some of the participants had already left the university, not all the listeners attended the follow-up study. Nine listeners from test group and seven from the control group came back and participated in the follow-up test session. Details of the experiment can be seen in the following Fig. 7.2.

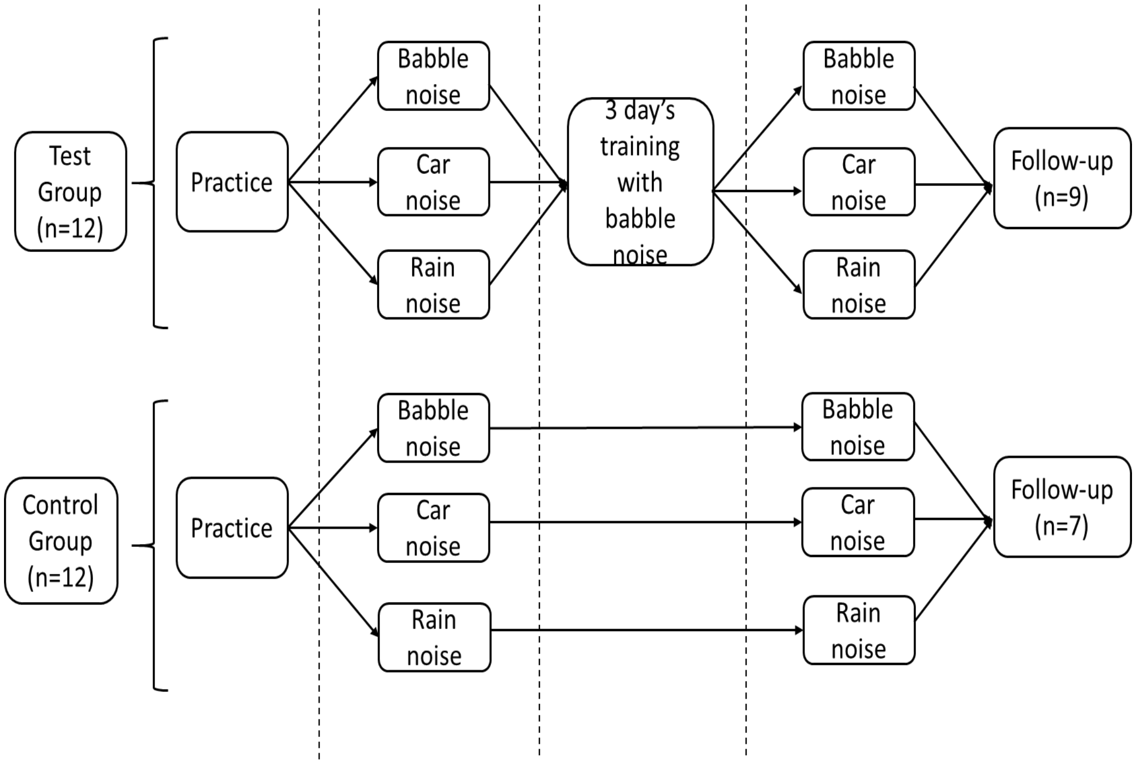


Fig. 7.2 Experiment design for the BKB test

7.2.4 Data analysis

The outcome measure was the number of BKB keywords that were correctly identified. Pre-test word identification was within two standard deviations from the mean for all participants, so no datasets were identified as outliers and removed from the analysis. Analysis of covariance (ANCOVA) was used to investigate whether post-test performance was better for the trained or control groups, and whether this differed across noise conditions (babble, car, rain). Pre-test performance was used as a covariate to control for baseline differences in pre-test performance. Post-hoc ANCOVAs were also carried out to investigate the difference between the post-test scores for the training and control groups in the different noise conditions.

Then a two-way analysis of variance (ANOVA) was used to look at whether improvement from pre to post test was greater for the trained group than the control group, and whether this differed across noise conditions (babble, car, rain). Independent t-tests were used to investigate whether the training group showed greater improvement than the control group in each of the three noise conditions (babble, car, rain). In order to explore the effect of training at the individual level, the linear regression of the post-test performance on the pre-tests results for the three noise conditions between trained (babble-trained, car-trained and rain-trained) and control (babble-control, car-control and rain-control) groups were conducted to see the relationship between the pre- and post-test performance across individual participants. The slopes of the regression-lines, which showed the relationship between pre- and post-performance across each condition and group in paired groups (babble trained vs babble control, car trained vs car control, rain trained vs rain control), were compared as well.

For individual day to day babble noise training performance, a one way repeated-measure ANOVA was carried out to test the learning results. Regarding the test results from the follow-up test session, mixed analysis of variance (ANOVA) was used to explore listeners' learning performance. Mixed ANOVA was also used to compare the learned values (improvement from pre- to post-test) for both groups with two time periods (period from pre- to post-test and pre-test to follow-up). Post hoc tests using a Bonferroni correction for multiple comparisons were carried out to explore the interaction between time and noise condition. In order to investigate whether listeners' learning performance was sustained and still measurable after several weeks, independent t-tests were conducted to compare the improvement from the pre-test to follow-up session for the training and control groups.

7.3 Test results

7.3.1 Pre- and post-test results

Fig. 7.3 displays the performance of correct responses for both of the test and control groups across pre and post-test with three different background noises: babble noise, car noise, and rain noise. Across the intervention of time period (pre- and post-test), ANCOVA showed that the main effect comparing the test and control groups' BKB performance was significantly different between each group (group, $F_{1, 19} = 41.07$, $p< 0.001$) and it indicated that the test group had significantly higher post-test scores than the control group. There was no significant difference among the three different background noise conditions (noise, $F_{2,38} = 2.55$, $p > 0.05$), but there was a significant interaction between the group and noise conditions (noise × group interaction, $F_{2,38} = 3.94$, $p < 0.05$). Post-hoc ANCOVAs investigated the difference between the post-test scores for the training and control groups in the different noise conditions. They showed that the training group scored higher than the control group in all three conditions. The amount by which the training group out-scored the control group was greater in the babble noise condition than the car noise condition ($F_{1,20} = 6.15$, $p < 0.05$) and the rain noise condition (F 1,20 = 4.98, $p < 0.05$) but there was no difference between the car noise and rain noise conditions ($F_{1,20} = 0.96$, $p > 0.05$).

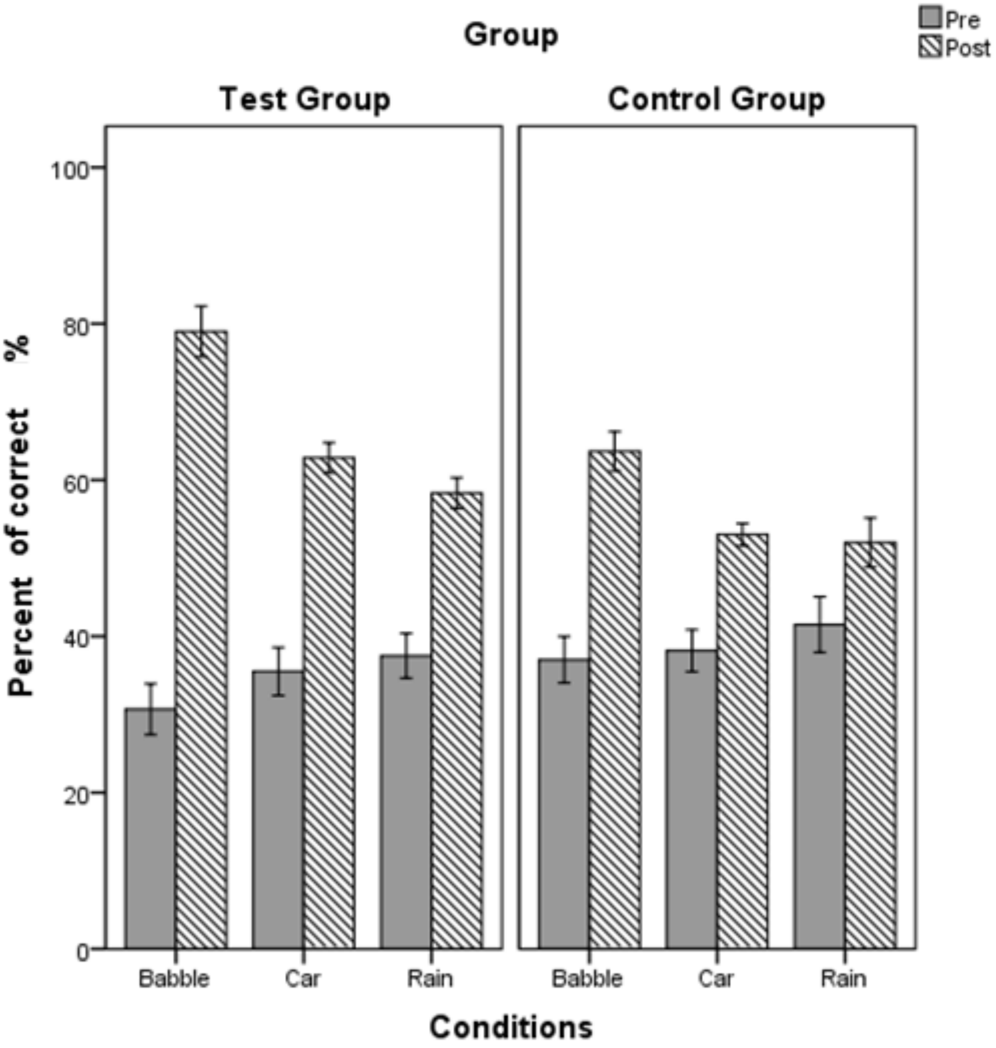


Fig. 7.3 Mean percent of correct responses for test (n=12) and control (n=12) groups with three different background noises: babble noise, car noise, rain noise. The pre- tests filled in grey and the post-test filled in lines. Error bars reflect ± one standard error.

A two-way ANOVA showed that the main effect comparing the learned values (the learning improvement from pre- to post-test) between the two types of intervention group was significant (group: $F_{1, 22} = 48.61$, $p < 0.001$). It indicated that the trained group improved more than the control group. Results also showed that there was a significant difference between the three noise conditions (noise: $F_{2, 44} = 39.82$, $p < 0.001$), and there was a trend towards a significant interaction between group and noise condition (group × noise: $F_{2,44} = 2.79$, $p = 0.072$). Independent t-tests were conducted for the three noise conditions to compare each test and control group's improvement. All of the three noise conditions showed a significant difference between the trained group's performance and the control group's performance [babble: $t(22) = 6.32$, $p < 0.001$; car: $t(22) = 4.07$, $p < 0.005$; Rain: $t(9) = 3.28$, $p < 0.05$)]. As shown in Fig.7.2, among all the background noise (babble noise, car noise and rain noise) conditions for BKB sentence performance, participants had the largest improvement in the babble noise condition (Pre-test: M = 30.67 %, SD = 11.29%, Post-test: M = 79%, SD = 11.27%; Improvement: M=48.33%, SD= 10.61%).The improvement for the control group in the rain noise condition was the smallest (Pre-test: M = 41.5%, SD = 12.33%, Post-test: M = 52%, SD = 10.85%; Improvement: M =10.5%, SD =11.48%).

7.3.2 Pre- and post-test results regression

Fig.7.4 depicts the relationship between the pre- and post-test performance of the individual listeners for each group (test group: filled with black points; control group: filled with grey points) and noise conditions (babble: Fig. 7.4.A, car: Fig. 7.4.B, rain: Fig. 7.4.C). For all the trained group participants, in all three noise conditions, the points were all distributed above the positive diagonal (solid black no-improvement line), indicating improvement was obtained between the pre- and post-test for the trained group in all three noise conditions. The slope of the regression line fitted to the car noise condition (slope: 0.44; $r^2 = 0.51$, $F_{1,10} = 10.42$; $p < 0.05$) was significant different from zero. Unlike the car noise condition, the babble noise (slope: 0.56; $r^2 = 0.31$, $F_{1,10} = 4.51$; $p = 0.06$) and rain noise conditions for the trained group (slope: 0.29; $r^2 = 0.17$, $F_{1,10} = 2.11$; $p = 0.18$) did not differ significantly from zero. The slope from rain noise condition is quite shallow, indicating that listeners in this condition finished the test with similar post-tests performance.

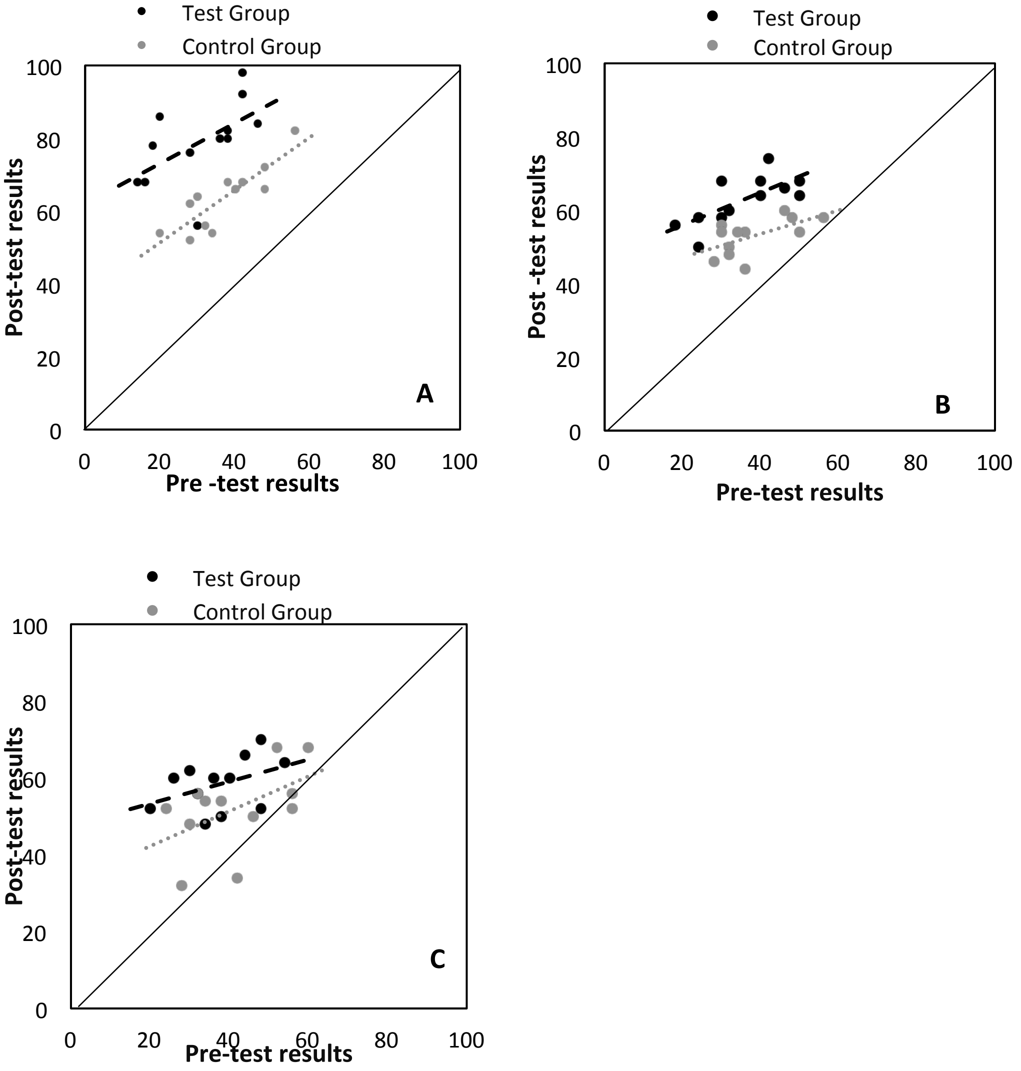


Fig. 7.4. Percent of correct responses in the pre-test (x axis) and post-test (y axis) for the babble (A), car (B), and rain (C) noise conditions. Data are shown for the test group (black points) and control group (grey points).

For the babble control group (filled grey points in Fig. 7.4.A), unlike the babble training group, the regression line was significantly different from zero (slope: 0.72; $r^2 = 0.73$, $F_{1,\,10} = 27.26$; $p < 0.001$). Its slope approaching 1, suggesting that there was a strong relationship between the pre- and post-test performance of the babble-noise control group participants. Therefore, the babble noise control group also improved, but by a relatively constant small amount regardless of the pre-test performance. For the overall control group data (filled with grey points in Fig. 7.4. A, Fig. 7.4. B, Fig. 7.4. C), two points from the rain noise control group fell below the positive diagonal, suggesting no improvement for these participants from their pre- to post-tests.

7.3.3 Individual day to day training performance

Fig. 7.5 shows individual training-group participants' BKB sentence recognition performance for each test session during the training experiment. A one-way repeated-measures ANOVA showed that for the babble noise condition, there is an overall improvement from the pre-test, training sessions and post-test sessions [$F_{4, 8} = 453.19$, $p < 0.0005$]. The following pairwise comparisons demonstrated that all the test sessions showed performance improvement from their previous test session (pre vs day1, day1 vs day2, day2 vs day3, all $p < 0.005$), expect for day 3 training session to the post-test performance ($p > 0.05$). There was considerable variability in the amount of performance improvement during the test sessions. For the babble noise training group, all of the listeners showed immediate and sharp improvement from the pre-test to the day 1 training session [$t(11) = 12.40$, $p < 0.05$]. While no significant difference was observed between the day3 and the post-test results [$t(11) = 1.70$, $p > 0.05$]. All of the participants showed overall improvement from the pre-test, training sessions and post-test, except the listener P02 who showed a slight drop down from the day3 training session to post-test.

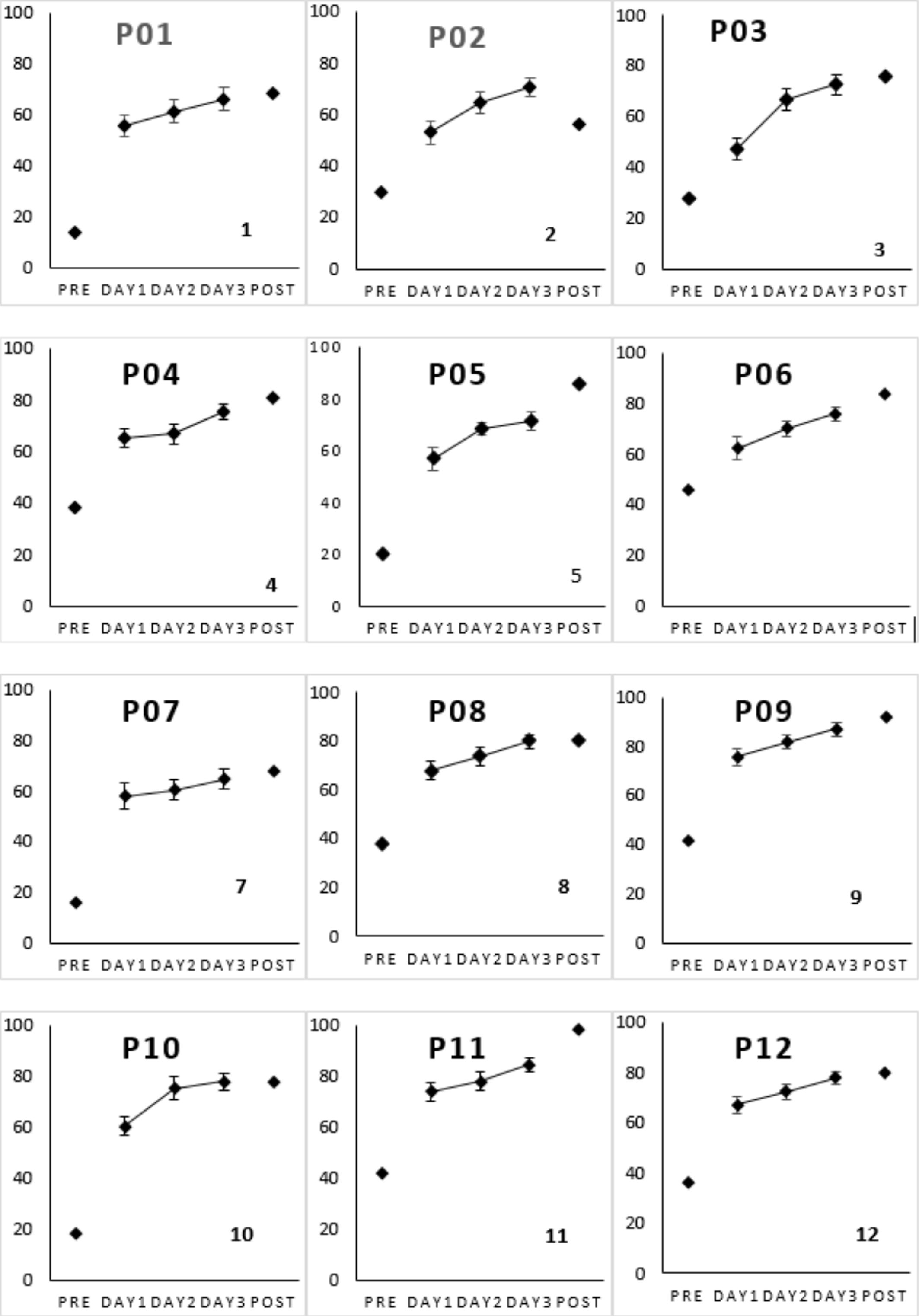


Fig. 7.5 Percent correct (y axis) performance from pre-test, training (day1, day2, day3; babble noise training) and post-test (x axis). Error bars reflect ± one standard error.

7.3.4 Results for follow-up test

Fig. 7.6 shows participants' performance across three test sessions (pre-test, post-test, and follow-up test session) with three different background noises (babble, car and rain noise). A mixed ANOVA was used to analyze the performance of both groups (trained and control) across three test time sessions (pre-test, post-test and follow-up) with different noise conditions (babble, car and rain). As we found before, the main effect comparing the listeners' speech performance with babble, car, and rain noise conditions was significant (noise: $F_{2, 28} = 8.72$, $p < 0.05$). Follow-up pairwise comparisons demonstrated that the babble noise performance was significantly different from both the car and the rain noise (both $p < 0.001$) but there was no significant difference between the car noise and the rain noise conditions ($p > 0.05$). There was a tendency towards statistical significance for the main effect comparing the two types of intervention –training group and control group (group, $F_{1, 14}= 4.54$, $p = 0.051$).

Regarding the main effect for time line, there was a significant difference across the three different test periods (time, $F_{2.56} = 133.81$, $p < 0.001$). Follow-up pairwise comparisons showed that there was a significant difference between performance in all three sessions (pre-test vs post-test; pre-test vs follow-up; post-test vs follow-up, all $p < 0.001$). There was also a significant interaction between time and noise (time × noise, $F_{4, 56}= 30.18$, $p <0.001$). The following post hoc tests using the Bonferroni correction for multiple comparisons revealed that most of the time periods in pairs are significantly different (pre-babble vs post-babble, pre-car vs post-car, pre-rain vs post-rain, post-car vs follow-up car, post-rain vs follow-up rain, pre-babble vs follow-up babble, all $p < 0.0056$). However, there was no significant interaction between time and noise condition in post-babble vs follow- babble, pre-car vs follow-car and pre-rain vs follow-rain (all $p > 0.0056$). This suggests that in the babble noise condition there was some retention of learning between the post-test and follow-up, whilst, the follow–up performance was almost the same as the pre-test results in both the car and rain noise conditions. The interaction between the three times and two groups was significant (time × group, $F_{2,56}=12.81$, $p < 0.001$), but no significant interaction from times, groups and three noise conditions (group × time × noise interaction, $F_{4,56} = 0.31$, $p > 0.05$). Details of the post hoc tests for the interaction between time and group will be analysed in the following session.

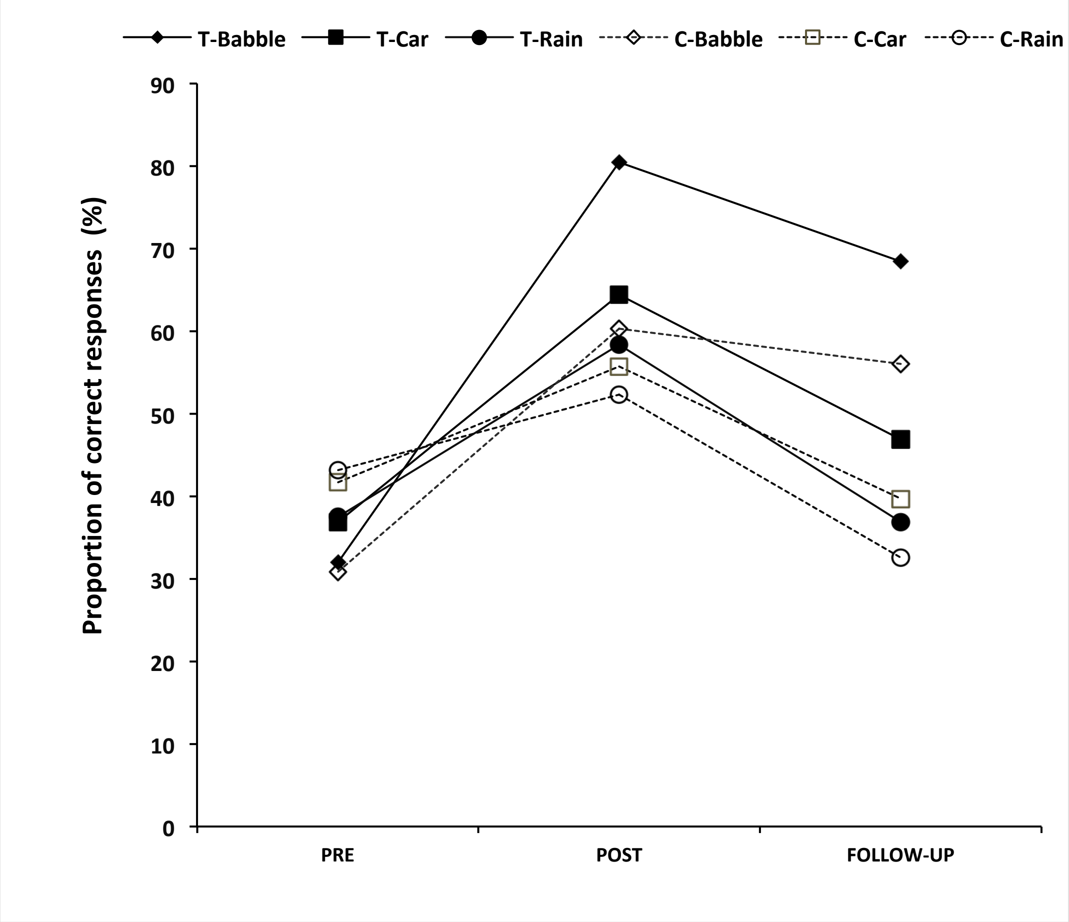


Fig. 7.6 Mean percentage of correct responses (words correct from BKB sentence tasks) for the test (n=9) and control group (n=7) from three test sessions (pre-test, post-test, and follow-up test session) with three different background noises (babble, car and rain noise).

Fig. 7.7 displays participant's performance for two gap periods (pre-test to post-test and pre-test to follow up session) with three different background noises (babble, car and rain noise). A mixed ANOVA was used to investigate the learned values (improvement from pre- to post-test and pre-test to follow-up) for the test and control groups in the different noise conditions to explore the sustained learning effect across three time periods. The main effect comparing the two types of intervention groups was significant (group: $F_{1,14} = 27.38$, $p < 0.001$). It suggested that the trained group improved more than the control one, regardless of the period of pre- to post-test and pre- to follow-up session. Results also showed that there was a significant difference among the three noise conditions (noise: $F_{2,28} = 50.08$, $p < 0.001$). Follow-up pairwise comparisons demonstrated that the babble noise improvement was significantly better than both the car and the rain noise conditions (both $p < 0.001$) but there was no significant difference between the car noise group and the rain noise group ($p > 0.05$). The improvement from the pre- to post- test period was significantly higher than the improvement from pre-test to follow-up period (time: $F_{2,28} = 87.94$, $p < 0.001$).

There was a significant interaction between noise and time periods (noise × time: $F_{2,28} = 7.95$, $p < 0.001$). Post hoc tests using a Bonferroni correction for multiple comparisons revealed that the improvement for the two periods (pre-post vs pre-to follow- up) were significantly different from each other, and this was true for all three noise conditions (babble: $p = 0.017$, car: $p < 0.00001$, rain: $p < 0.00001$). However, there was no significant interaction between group and noise condition (group × noise: $F_{2,44} = 0.20$, $p > 0.05$) and no three-way interaction between group, noise and time (group × noise ×time: $F_{2,28} = 0.42$, $p > 0.05$).

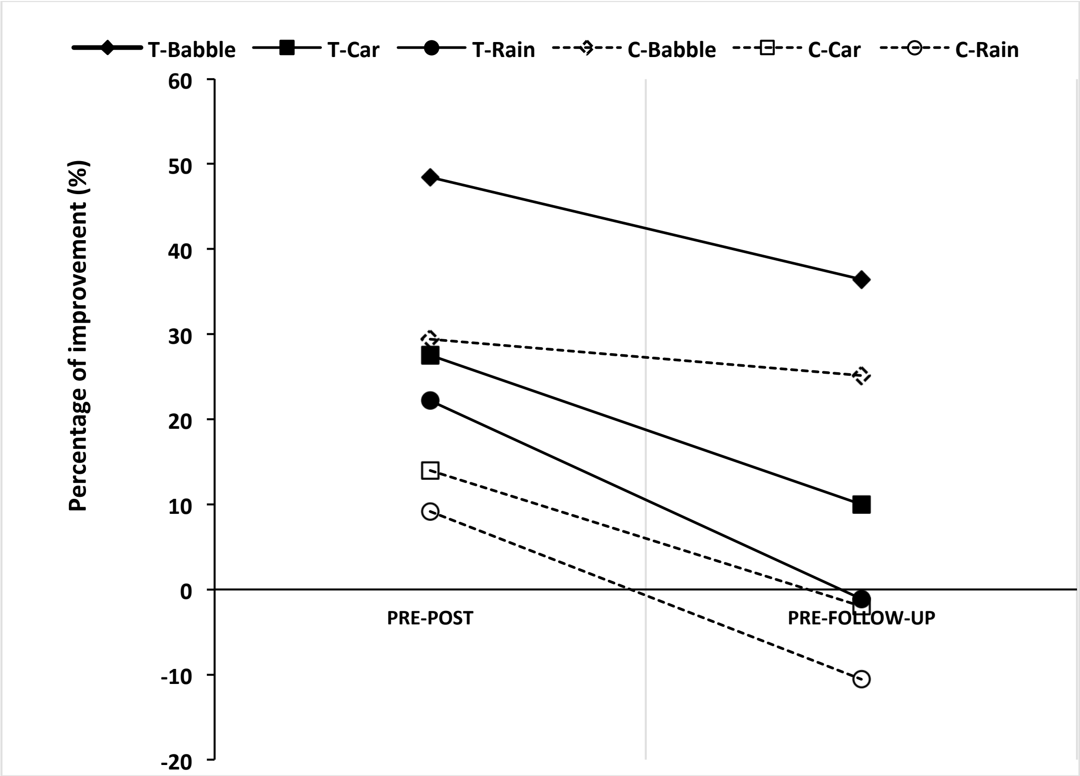


Fig. 7.7. Mean percentage of correct responses improvement (words correction from BKB sentence tasks) for the test (n=9) and control group (n=7) for two gap periods (pre-test to post-test and pre-test to follow up session) with three different background noises (babble, car and rain noise).

7.4 Discussion

Taken together, the present findings indicate that all participants from trained groups showed positive learning performance between the pre- and post-test in the three noise conditions. Both the post-test performance and learned values (improvement from pre-test to post-test) from the trained group were much higher than the results for the control group, regardless of the three noise conditions. It suggests that people learned to understand speech in babble background noise (learning effect), and this learning effect generalized to identify speech sounds against car and rain background noise conditions (generalization effect). Regarding the amount by which the training group out-scored the control group for the three noise conditions, it was found that speech training with babble noise improved performance more in the subsequent babble noise condition than in the car and rain noise conditions.

Listeners’ performance in the follow-up test (several weeks after the initial experiment) showed that the performance of the trained group was higher than the performance of the control group, regardless of the noise conditions. Both the babble noise trained and control groups showed that listeners who were explicitly subjected to babble noise demonstrated a slight failure to recall what they had learned from their pre-test to their post-test, only part of the learning effect was sustained after several weeks. However, although auditory generalization effects were observed from training speech sound against babble noise to speech sound against rain and car noise in the post-test session, the listeners’ performance improvement for both the car and rain noise conditions was not as much as the learning improvement for the babble noise trained condition. The follow–up performances were almost the same as the pre-test results for both the car and rain noise conditions. The potential reasons for the learning and generalization effect from this experiment are discussed below.

7.4.1 Test materials

In this study, the babble noise trained group obtained a larger training benefit in babble noise condition than the in car and rain noise conditions. This may be because the test stimuli (both target and background sounds) presented during training were the same for the babble group listeners at pre- and post-test sessions. While for car and rain noise conditions, the target sounds (BKB sentences) were the same from the training to the pre- and post-test sessions, but background noise were different from the training session to these two test sessions. Morris *et al*. (1977) and Roediger *et al*. (1989) suggest that listeners' learning and memory performance will be maximized if the stimuli used during the training and testing is the same. The similarity between the trained and tested sound determines the learning extent and amount of how much can be transferred (Borries *et al*., 2012). Therefore, although the learning effect from training speech sounds against babble noise generalized to speech sounds against rain and car noise, improvement of listeners' performance for both the car and rain noise conditions were not as much as the learning improvement for the babble noise condition.

7.4.2 Speech cues

Babble noise is one kind of speech pattern masking sound, and training with speech in babble noise can help listeners to 'pick up' target sounds and 'tune out' particular sorts of background noise (Van Engen, 2012). Listeners in this study may make use of the speech cues (speech spectral components) in babble noise to pick up the most important speech information from babble background noise and tune out the irrelevant sound information. Previous studies from Loebach and Pisoni (2008) show that auditory perceptual training can influence the auditory performance on speech in noise tasks. It may affect the distribution of attention to speech stimuli by inhibiting the irrelevant sound cues (Melara *et al.*, 2002; Loebach and Pisoni, 2008; Tremblay *et al.,* 2009). Improvements from pre- to post-test were sustained over a period of several weeks for words presented in babble noise, but not for words presented in car or rain noise conditions (which performance returned to pre-test levels). The sustained improvement for speech identification amid babble noise was present for both training and control groups, suggesting that the sustained improvement may relate more to the nature of babble noise than to the benefits of exposure or training per se.

7.4.3 Background noise changed speech perception

According to results from this experiment, listeners' ability to identify speech sounds against babble noise still generalized to recognise speech stimuli against the car and rain noise. Loebach *et al*. (2008) demonstrate that training on speech may lead the listeners to make use of lexical judgments about stimuli, process sound information in a higher cognitive order, and reduce participants' attention to the lower order of focusing on acoustic features. However, earlier perceptual speech training studies suggest that auditory perceptual training may adjust human's auditory system listening ability by two actions: one is to increase the awareness of informative signal cues and the other one is to decrease the influence of less useful stimuli. These two speech processing actions may be combined together or just one of them may be involved, it depends on the signal and background noise used (Schwab *et al*, 1985; Francis *et al.,* 2007; Francis & Nusbaum, 2009).

Speech and environmental sounds are all complex meaningful real-world sounds, which may present information about objects, opinions, or events in a certain time or location. Neuroimaging studies demonstrate speech and environmental stimuli show overlapping patterns of activation (Lewis *et al.*, 2004; Loebach & Pisoni, 2008). Kidd *et al.* (2007) suggest that speech and environmental stimuli share the same auditory sound processing pathway to recognise familiar sounds. In order to store and locate an auditory stimulus, there is a common mechanism for the detection and recognition of both speech and environmental sounds: listeners' auditory attention may focus on the most important spectral and temporal information. Moreover, the signal processing mechanism for the auditory system to receive and identify useful information from speech sounds is as important as for recognising environmental stimuli. When they were doing speech-in-noise tests in this present study, the improvement might be due to either better processing of the speech or better filtering or 'tuning' out of the noise. Although the test background noise changed during the test periods, listeners' ability to recognise the useful speech information may not be significantly influenced.

7.4.4 Familiarity of test procedures and stimuli

The familiarity of the tested stimuli and test situation may lead to procedural learning in this study. According to the individual training performance, there was a greater overall improvement from pre-test session to day one performance, but no performance differences from their day three to post-test session. This is probably because once listeners were familiar with the test procedure and learned how to identify speech sounds from one kind of background noise, it would be possible for them to achieve their asymptotic performance. Even though the background noise was changed from one noise condition to another one during the pre-and post-test session, the ability to identify the speech sounds did generalize to other daily environmental noises. What is more, as the key words from test speech sentences are used in daily life and the background noise (babble, car, and rain noise) are all simulated from daily life, listeners may be familiar with these kinds of sounds. Therefore, the perceptual learning may be induced by the familiarity of the speech sounds and test procedure used in this study.

7.5 Conclusion

This study has shown that speech perception in noise is malleable with short-term training, the benefits of training (learning effect) are partly maintained over time, and that speech training with babble noise background noise is generalized to other environmental noise conditions, such as car and rain noise. This study highlights the need to consider both the target and background sounds when creating auditory training programmes. The outcomes provide important evidence for the use of background noise in perceptual auditory training programmes to improve people's listening ability in challenging environments. In the future, it will be important to consider how this research can be used to further investigate speech perception in older adults (e.g. Kim *et al*., 2006; Wong *et al.* 2010), provide better musical training (e.g. Parbery-Clark *et al.,* 2011) and explore more methods to train people who experience auditory perceptual difficulties. For example, Wong *et al.* (2010) suggested that there is a decline in the volume and cortical thinness of the prefrontal cortex (PFC) for older adults and due to this, older people's ability to perceive speech in noise is declined. They also found that a thicker PFC might be compensated for improving cognitive functions for older adults. Therefore, combining auditory and cognitive training strategies could be explored to improve older people's speech perception ability in noise.

The findings from this study could be used as a baseline for further training for related auditory plasticity research in hearing impaired people, such as effects of age and hearing loss level on speech perception in noise (Dubno *et al*., 1984; Helfer and Wilber, 1990), or children with learning problems, such as learning difficulties, Auditory Processing Disorders (APD) (Bradlow *et al.,* 2003; Ziegler *et al.,* 2005, 2009). Halliday *et al.* (2008) stated that the improvement of auditory learning ability is varied across different ages. Older people's frequency discrimination ability is not as good as younger adults who aged between 18-40 years. The majority of children's performances show a fluctuating pattern of learning performance and their auditory learning has a prolonged developmental time course. Therefore, in order to apply the findings from this current study for training people with age- related hearing loss or children with APD, a longer training duration design than that for normal hearing people need to be taken into consideration at first.

7.6 Summary

This study explored the effect of changing the background noise on auditory perceptual learning. Results showed that training with speech against babble noise generalised to speech against car and rain noise conditions. Part of the perceptual learning from speech against babble noise training was also sustained after several weeks' gap. Findings from this study suggest that people's listening performance can be improved with training of babble noise and that this generalises to other environmental sounds. As the control group received equal exposure to all the three noise types, the sustained learning with babble noise, but not other noises, implies that a structural feature of babble noise was conducive to the sustained improvement. These findings emphasise the importance of considering the background noise as well as the target stimuli in auditory perceptual learning studies.

Chapter 8: General discussion

8.1 Introduction

As the results from the individual studies of this thesis have already been discussed in previous relevant chapters, this chapter will present a broader discussion and reflection on the results from the previous chapters. Based on the research objectives, laid out in Chapter 2, a review of the results from the perceptual learning studies in this thesis will be provided in section 8.2. Following this review, the strengths and potential weaknesses of the auditory perceptual learning studies in this thesis will be summarized and described. Finally, the suggestions of experiments for further work in the auditory perceptual learning area will be considered.

8.2 A review of the results from previous chapters

The psychoacoustic perceptual learning study in Chapter 4 showed that training on the non-speech SAM detection task did not generalize to the SAM-rate discrimination task; this was true with all three different depths of SAM-rate discrimination stimuli (answer to objective 1 in section 2.4). The performance improvement (from pre- to post-test) of SAM-rate discrimination task was statistically indistinguishable among these three modulation depths (mean of SAM-rate threshold improvement for each modulation depth 100%: 5.66%; 70%: 7.27%; 40%: 7.58%). However, comparing the three different modulation depths (100%, 70%, and 40%), the most difficult condition (40% modulation depth) continued to have the highest discrimination threshold (i.e. worst performance) after training, compared to the other two conditions (100% and 70% modulation depths). Even though SAM detection and SAM rate discrimination tasks shared similar stimulus features, no generalization occurred from training with SAM detection task to SAM-rate discrimination task, regardless of the modulation depths. The results indicated that stimulus learning is not sufficient to improve perceptual learning between different SAM tasks. It was suggested that the outcomes from the psychoacoustic perceptual learning study in Chapter 4 might be due to the auditory system processing these two tasks (SAM detection and SAM-rate discrimination) separately. Another potential reason is overtraining occurring in the early stages of the experiment, resulting in asymptotic performance in many participants early on, and subsequently, could explain the poor generalizability from training task (SAM detection) over to the second task (SAM rate-discrimination).

The next steps of this project were studies of perceptual learning performance of nonsense stimuli (VCV) with fixed and random background noise in Chapters 5 and 6. Three experiments were carried out to explore the perceptual learning of nonsense stimuli identification in fixed and in random background noise. These studies were motivated by perceptual learning studies in the visual domain (Schubö *et al*., 2001). The detection performance of participants completing in the visual studies showed no improvement when target stimuli were masked with random visual noise, but participants could improve their detection performance by learning to ignore the fixed visual noise. Once participants had learned how to ignore the fixed visual noise and could successfully detect targets, this skill then generalised to random visual noise (Schubö *et al.,* 2001). As in daily life, people are more likely to meet noisy communication conditions with random background noise than fixed noise. Therefore, both the pre- and post-test sessions for all the test groups (fixed, random and control group) in Chapter 5 were using random babble noise as background noise. This experimental design was to find out which training method (fixed or random babble) was better to help people to be more sensitive to speech information in random background conditions.

Results from the experiments in Chapter 5 showed that participants' identification of auditory stimuli (nonsense syllables) improved in both fixed and random background noise across training sessions. In addition, nonsense syllable identification performance with random babble noise training was better than the performance with fixed noise training (with both pre- and post-test session using random noise; answer to objective 2 in section 2.4). However, the learning effect with nonsense syllables did not generalize from training normal hearing listeners to identify target sounds in fixed babble noise to identify sounds in random background noise.

Compared with the previous literature, these results were different from the perceptual learning study Schubö *et al.* (2001) carried out in the visual domain, which showed that perceptual visual learning could generalize from training with visual fixed noise to visual random noise. The results from Chapter 5 are also in contrast with the auditory perceptual study from Felty *et al.* (2009), which compared fixed with random noise in a single session (i.e they trained and tested participant performance with the same background noise) and demonstrated that listeners achieved better word recognition with fixed babble noise than with random babble noise. Details of the potential reasons for the different findings were summarized in the introduction section of Chapter 6 (see section 6.1).

A short single session nonsense syllables identification in noise experiment was subsequently carried out and reported in Chapter 6. This investigated the identification performance for nonsense speech sounds against fixed babble noise or random babble noise (between-subjects design) in a single session. The results showed that nonsense syllable identification in fixed babble noise was similar to that in random babble noise. These results confirmed that the test method differences between Chapter 6 and Chapter 5 led to nonsense speech sound identification in noise performance differences (answer to objective 3 in section 2.4). The Chapter 5 experiment used multiple sessions with pre, post-test (random background noise) and training session (random or fixed background noise). The Chapter 6 experiment used a single session, which trained and tested with the same type of background noise (random or fixed noise).

The final step of this project was conducted in Chapter 7 with more complex speech sounds combined with real life environment noise. It was intended to test whether the learning effect from training with speech in random babble noise generalized to car and rain noise. The results from this experiment revealed that people's speech recognition performance improved with the training of speech sounds with random babble noise. The perceptual learning effect from training with speech in babble noise also generalized to car and rain noise conditions (answered objective 4 in section 2.4) and part of the learning effect from speech in babble noise was sustained after several weeks. The sustained improvement for speech identification amid babble noise was present for both training and control groups, suggesting that the sustained improvement may relate more to the nature of babble noise than to the benefits of exposure or training per se. However, the training in babble produced substantially more improvement to performance in babble than to performance in other types of noise, suggesting that generalization was in fact rather limited. It maybe worth investigating whether using of mixed types of noises (i.e. another patterns of 'random noise') as training background noise can maintain the improvement performance in future auditory perceptual learning studies. These findings emphasise the importance of considering the background noise as well as the target stimuli in auditory perceptual learning studies.

There are several potential explanations for why performance improved or generalized following auditory training across the studies in this project, which will be assessed in the following sections.

8.2.1 The duration of training changes auditory perceptual performance

The duration of training plays an important part in the auditory perceptual learning process. Wright and Zhang (2009) stated that practice in perceptual training trials improved participants' perceptual learning performance on stimuli detection and discrimination. From perceptual learning in the visual domain, there is an effect named perceptual deterioration, which is caused by overtraining during perceptual learning. Overtraining in a perceptual task can generate an improvement in performance at the beginning, followed by gradual decline afterwards (Censor, Karni & Sigi, 2006; Mednick, Nakayama & Cantero, 2002; Ofen, Moran & Sagi, 2007). Perceptual deterioration occurs due to the limited capacity of early visual area, when the visual area becomes saturated with information during overtraining, it will be hard to consolidate newly acquired changes (Mednick, Nakyama & Stickgold, 2003; Ashley & Pearson, 2012). Wright and Sabin (2007) demonstrated that once perceptual training performance achieved a certain amount of learning, additional training had no further benefits for the listeners (no further improvement on the training task or no further generalization from training task to untrained task).

The training trials for perceptual training were critical for the whole perceptual training studies, especially for long-term training studies. The psychoacoustic study in Chapter 4 showed that the majority of listeners had already achieved their asymptotic performance for the SAM detection tasks after day one's training sessions (see results section 4.3.3). It indicated that overtraining took place through the SAM detection test sessions. In this case, overtraining was indicated by a slower perceptual learning gradient (day to day improvement rate) after they reached their asymptotic performance. Therefore, the influences of daily learning limits should also be taken into consideration in auditory training design for future study design. Wright and Sabin (2007) also demonstrated that if learning on two tasks had modified different circuitries at the same physiological level, training on one of the tasks would inherit some features which made that circuit less amenable to change. The overtraining on the trained task would then prevent the learning to be transferred to another task. Therefore, learning from outcomes of the psychoacoustic auditory training experiment in Chapter 4, it is suggested that the training length is vital for perceptual training design.

8.2.2 The fixed versus random background noise training changes auditory perceptual performance

The nonsense syllables training studies in Chapter 5 demonstrated that the random babble noise training method led to a greater improvement than the fixed babble noise trained method when tested with random babble noise. The results suggest that auditory training with nonsense syllable identification in random babble noise was better than training with nonsense syllables identification in fixed babble noise to improve people's speech understanding in random listening conditions. Based on my experiment design, there are two components of stimulus learning: one is the target (VCV) stimulus learning and the other is the babble background noise learning (Fixed or Random babble noise depends on attribution of the trained groups). Both the task learning (the identification of the nonsense syllables in babble noise task) and the procedural learning (familiarity with the lab environment, keyboard, PC screen, using the mouse and familiarity with the test requirement) were common across all groups of participants.

For the studies in Chapter 5, if a general improvement in listening skills was solely responsible for the observed improvements following training, one would expect similar improvements across both fixed and random background noise. In contrast, improvements in nonsense syllables perception in random babble noise were larger than improvements in fixed babble noise. The comparison of performance improvement differences for each group in terms of learning types in Chapter 5 demonstrated that the observed additional improvements in performance between random and fixed group appears to be due to learning of random noise rather than to task familiarity. Details of the description on performance improvement for test groups in terms of learning types can be seen in Table 5.5 in Chapter 5.

Felty *et al.* (2009) demonstrated that listeners showed greater improvement in word recognition performance when the same sample of background babble-noise was presented on each trial, compared with when different noise samples were presented on each trial. In contrast, the perceptual learning study with VCV stimuli in chapter 6 found that VCV identification improved by a similar amount when stimuli were presented against a random-noise background and against a fixed-noise background, indicating that the effects of background noise on perceptual learning differ with different types of target stimuli (very short VCV targets contrasting with longer word stimuli). More discussion can be seen in section 6.4 Chapter 6. Results in Chapter 6 also indicated that training and tested nonsense syllables with the same patterns (Fixed or Random) of babble noise can lead to performance improvement, but there's not a difference in correct responses between fixed and random training when they are tested with the same training babble noise. It is important to note that in Felty *et al.* (2009), participants only experienced fixed or random babble noise conditions. However, in Chapter 5 people were trained with either fixed babble or random babble, but always tested with random babble. Had the fixed training group been tested in fixed conditions, it is possible that they would also have shown more improvement than the random-trained group.

The performance in terms of how "Don't Know" responses affects both the changes in correct and incorrect response proportions should be further considered. The guess rate calculations for the VCV studies were based on the assumption that all the consonants were equally confusable with each other. Phatak and Allen (2007) indicated that the identification of consonants within speech noise are separated in to three sets: high score consonants: /t/, /s/, /z/, /ʃ/, /ʒ/; intermediate ones: /n/, /p/, /g/, /k/, /d/; low score ones: /f/, /θ/, /v/, /ð/, /b/, /m/). The consonants in this experiment belong to the intermediate and low score groups. Therefore, the confusion rate and identification rate for all the eight consonants may not be similar. This would lead to listeners to guess to make a decision rather than using "Don't know", especially in the situation when they heard the consonant but confused between pairs (i.e. /m/ and /n/. /b/ and /d/) to make a decision on which one it was. Alternatively, the participant might detect there was a signal, but they were not sure which one it was, then they might still choose the "Don't know" response. "Don't know" responses may reduce the number of correct responses and increase participants' decision bias in the learning process. Similar to the open set tasks in which participants report words or sentences, participants may end up of using don't know responses when they missed the words or part of the sentences. Without providing a "Don't know" option in open set tasks, some of the participants who are very confident may guess what the words are and some of them who are not confident may end up with no answer (a zero score). Therefore, guessing at random between the two options, compared to responding "Don't know" each time would lead to a reduction in "Don't know" responses being equally split between correct and incorrect responses. Further studies could investigate comparing forced choice tasks versus responses with a "Don't know" option to explore which one is more effective and accurate for performance learning improvement.

8.2.3 The influence of the similarity between target and interferer information

Background noise can interfere with speech understanding through two mechanisms: energetic masking and informational masking. Energetic masking occurs when the background noise has energy in the same frequency region as the speech signal, thus preventing the speech signal from being perceived. When the background noise fluctuates, as is likely with real-world environmental sounds and competing speech, the listener is afforded opportunities to ‘listen in the dips’, or ‘glimpse’ the speech signal (Howard-Jones & Rosen, 1993). Informational masking is a catch-all term for any masking that cannot be explained through energetic masking, and is likely to reflect difficulties with auditory scene analysis (Bregman, 1990), including failures of object formation and object selection (Shinn-Cunningham, 2008). Informational masking can be particularly problematic when the speech signal and background noise are similar, due to the difficulty of segregating the target speech sounds from the background masker (Brungart, 2001).

Due to the differing effects of energetic and informational masking, the amount and type of benefit that participants receive from perceptual training may differ depending on the type of background noise. Steady-state noise, such as speech-shaped noise, is likely to provide consistent energetic masking but little informational masking. On the other hand, the temporal variation in babble-noise will afford more opportunities for glimpsing, but increased informational masking if words are partially audible. Correspondingly, training strategies that improve glimpsing or segregation may be more useful for speech presented in babble than for speech presented in steady-state noise. Results from Chapter 5, 6 and 7 all showed that training nonsense syllables (VCV) and speech sound in babble noise improved the identification performance in noise and reflected that training improved listeners ability to glimpse signal stimuli in babble noise. Van Engen (2012) trained participants on English sentence recognition in three different background noise conditions: speech-shaped noise, Mandarin babble, and English babble. Van Engen (2012) found the post-performance for all the four groups was always better when tested with Mandarin babble than with English babble. However, considering the post-performance for the English babble test only, the English training group had better performance than those who had trained with Mandarin or SSN. The results suggest that it is better to train them with speech informative sounds as the masking background noise than with non-speech stimuli. Similarly, Green, Faulkner, and Rosen (2019) found that training with speech-in-babble-noise improved cochlear-implant users' perception of sentences in babble noise, but did not result in improved perception of phonemes in speech-shaped noise. These studies suggest that speech-like noise may enable listeners to develop strategies that allow them to 'listen in the dips', where energetic masking is reduced. This benefit of dip-listening appears to offset any costs associated with increased informational masking for babble noise relative to steady-state noise.

Felty *et al.* (2009) demonstrated that listeners showed greater improvement in word recognition performance when the same sample of background babble-noise was presented on each trial, compared with when different noise samples were presented on each trial. Similar results were found in a visual texture segmentation task in which a background mask was either consistent from trial-to-trial or varied on each trial (Schubö *et al.,* 2001). In contrast, the perceptual learning study with vowel-consonant-vowel (VCV) stimuli in Chapter 6 found that consonant identification improved similar when stimuli were presented against a random-noise background with against a fixed-noise background, indicating that the effects of background noise on perceptual learning differ with different types of target stimuli (very short VCV targets contrasting with longer word stimuli).

Results in Chapter 7 showed that participants in the control group had identical exposure to the different background noises and yet had better word identification accuracy for words in babble noise than car or rain noise at follow-up, several weeks after the initial study. Neuroimaging studies demonstrate that speech and environmental stimuli show overlapping patterns of activation (Lewis *et al*., 2004; Loebach & Pisoni, 2008), and share the same auditory sound processing pathway leading to sound recognition (Kidd *et al*., 2007). However, it remains unclear whether specific regions of the auditory cortex are selectively involved in processing speech. Overath and colleagues (2015) have argued there are structures in the auditory brain tuned for speech-specific spectro-temporal structure. One simple possibility for the different levels of sustained learning is that babble noise affords more opportunities than steady-state noise for ‘glimpsing’ (listening in the comodulated or uncomodulated dips; Rosen *et al*., 2013). While dips are present in the car and rain noise samples, they are less frequent and with reduced amplitude modulation (see Fig. 7.1). Potentially, through exposure and/or training, participants learned to utilize dips more effectively, and this specific learning was sustained over time, benefitting the babble noise condition but not the car and rain noise conditions. The benefits of training in babble noise generalized to car and rain background noise. However, the generalization was rather limited, it showed that the consideration of the noise type is important in training. As all three types of noise were broadly similar in their spectral and temporal profile. Future research could investigate whether learning benefits generalize to familiar noises with different spectro-temporal profiles (e.g., drumming), or whether generalization depends on perceptual similarity.

8.3 Strengths and weaknesses of the research approach

8.3.1 Modulation depths used for perceptual learning study

The modulation depth for stimuli indicates how much the modulated signal varies from the original level. It is one of the characteristics of the fluctuating signal. In the literature, there are many psychoacoustic studies which investigate the effect of modulation depths on AM stimuli thresholds for AM detection and discrimination tasks (Grant *et al.,* 1998; Edwards & Vienmeister, 1994, Fitzgerald & Wright, 2011), but no research was found to explore whether the characteristic of modulation depths for AM stimuli lead to AM sound perceptual learning performance differences. Most previous psychoacoustic perceptual learning studies focused mainly on task differences between different AM stimuli (Demany & Semal, 2002; Amitay *et al.,* 2005; Fitzgerald &Wright, 2011). Although Fitzgerald and Wright (2011) found that training with SAM detection task did not generalize to SAM- rate discrimination with full 100% modulation depths. This project started with a psychoacoustic experiment to explore whether training of SAM detection tasks make differences to SAM-rate discrimination tasks with lower modulation depths (mid: 70%; low: 40%), as well as 100% modulation depth. However, it was concluded that overtraining, a lack of mixed stimuli learning, and listeners' attention, were the potential reasons for no generalization from training with SAM detection to SAM rate discrimination task. These factors should be considered for any subsequent studies in this area.

8.3.2 Comparing fixed and random babble noise training methods in hearing domain

From perceptual learning in the visual domain, fixed and random visual noise training lead to different performance outcomes (Schubö *et al.*, 2001). People could improve their detection performance by learning to ignore fixed visual 'noise'. However, no performance improvement was found from training with random visual ‘noise’ (Schubö *et al.*, 2001). Once participants have learned to ignore the constant visual noise and can successfully detect targets, this skill was then transferred to new, random visual noise (Schubö *et al.*, 2001). This project investigated whether the same patterns from the visual domain could be found in auditory perceptual learning domain. Results showed that both fixed and random babble noise training methods lead to hearing performance improvements when tested with random babble noise. However, a better learning effect was found from random babble noise training compared to fixed babble noise training in this hearing research. Training with fixed babble noise did not generalize to random babble noise conditions. The findings emphasise the importance of considering the background noise in auditory perceptual learning studies. As ‘real-world’ background noise will be random, then training with this synthetic random babble noise could be used for training listening sensitivity in real world environments. Moreover, this project further showed that training on this random babble does have small transfer effects to other background noises (car, rain), but only in the short-term.

In addition, as normal hearing listeners participated in this experiment, it will also be worthwhile to identify whether this training method could be applied in the clinic to help hearing-impaired people. Burk *et al.* (2006) demonstrated that both older hearing impaired people and young normal hearing people were able to improve their speech in noise performance after training with repeated presentation of the test words. But older hearing people required a more advantageous SNR and more training time to improve their performance in the same way as that of the young normal hearing people. Therefore, the ability to understand speech under challenging listening conditions among hearing-impaired listeners may also benefit from training speech tasks with similar random noise conditions with higher SNR and longer training time. Further research can be conducted to explore whether the same perceptual learning patterns from fixed and random noise training in this thesis still can be obtained in perceptual learning with hearing-impaired people. However, Burk *et al.* (2006) also reported that although training on a set of monosyllabic words spoken by a single talker resulted in learning for recognition of those words in noise that was maintained over time and generalised to other talkers, improvements relative to untrained words were found only when the words were presented in isolation, and not when they were embedded in sentences. Therefore, the benefits from training were unlikely to be apparent in real world situations. The study in Chapter 7 trained with sentences in random babble noise; the results showed that benefits of training in babble noise generalized to car and rain background noise in the short-term. But the training in babble produced substantially more improvement on performance in babble than on performance in other types of noise, so that generalization was, in fact, rather limited. Further research should be conducted to explore ways on how to boost and sustain the perceptual learning benefits from speech in noise in real world communication environments.

8.3.3 The “Don’t know” responses used for perceptual learning

The “Don’t know” response was used in the nonsense syllable identification in noise experiment design, which is different from most of other auditory training research in the literature without this choice. For example, some psychoacoustic studies use closed set tasks by forced choices (Amitay *et al*., 2005; Fitzgerald & Wright, 2011), and some speech studies used open set sentences (Felty *et al*., 2009; Fu *et al*., 2005, Stacey & Summerfield, 2007). Similar to the VCV experiment “Don’t know” option, the open set sentence task includes a possible “Don’t know” response when people do not detect the single sound with no responses, but it will be scored as zero for the final performance outcome.

Comparing the training session for fixed and random training group in Chapter 5, the “Don’t know” response was similar at day1 for fixed and random group. But the “Don’t know” responses at the day 2 and day 3 showed that the fixed group had lower responses than the random group. Therefore, it demonstrated that participant felt that the fixed training is easier and had more confidence than the random training to get used to the background noise. However, comparing the correct responses across test sessions between fixed and random group, it showed that both group had similar correct response across day 1 to 3. Similar in Chapter 6, the “Don’t know” responses patterns differences between fixed and random groups indicated listeners’ confidence in what they heard and what responses they made. i.e. for fixed babble noise, participants rapidly became confident in what they thought of they'd heard and the "Don't know" responses dropped off after the first block. However, the "Don't know" responses in the random babble noise group showed that the confidence in what they thought of they'd heard increased much slower.

There are several advantages of including the "Don't Know" response in the experiment design. First, the "Don't Know" response gives another measurement of performance that is not necessarily linked to correct responses. Second, the "Don't know" choices added extra information to understanding perceptual learning process. The "Don't know" performance indicated that changing stimuli across test sessions affected listeners' reaction in responses. In addition, results from comparison between the guess rate (the amount of decrease in "Don't know" performance/8) and the improvements in the percentage of correct responses from both VCV studies confirmed that participant's performance improvement was not due to changes in response criteria, but due to listeners' perceptual learning. The third advantage for the "Don't know" responses was that the "Don't know" responses could give participants a choice for trials in which they did not detect the signal sound. As the listeners may feel the task VCV in noise task is hard to detect the VCV stimuli, it could potentially be more demotivating for them if they have to make random guesses. If participants are not motivated then they will not do the task to the best of their ability. Including "Don't know" choice in an experiment design can stop participants from being forced to guess when they cannot hear the signal. However, the only disadvantage of "Don't know" responses is that it may reduce the number of correct responses and increase participants' decision bias at initial stages of the learning process. The bias can be made between the eight consonant choices and the "Don't Know" response. That means participant may detect there is a signal, but they are not sure which one it is, then they may still choose the “Don’t know” response. More details of considerations of the “Don’t know” responses can be seen in section 5.5.2.

8.3.4 Environmental background noise used for perceptual learning

Previous studies in auditory perceptual learning showed a generalization effect from trained to untrained stimuli, but they mainly focused on changing the target stimuli using amplitude modulated sounds or speech stimuli (Clarke & Garrett, 2004; Hervais-Adelman *et al.,* 2011). Different studies used different background stimuli, with babble noise or speech-shaped noise commonly used. However, it was not clear whether training generalized to other types of noise, in particular real-world environmental noise such as car and rain noise. Even though previous studies have looked at training with environmental stimuli, this was conducted using environmental sounds as the target stimuli, such as footsteps, slamming door, air conditioner, dishwasher, etc. (Burkholder, 2005; Reed & Delhorne, 2005; Kidd *et al.*, 2007), not using environmental sounds as background noises. Research in this thesis used environmental sounds (car and rain noise) as background noise to explore the perceptual learning and generalization effects in hearing. The results showed that training with babble background noise transferred to environmental background noise conditions. However, the sustained learning with babble noise lasted over a period of several weeks, whereas the learning effects for other noise types was not sustained, implying that a structural feature of babble noise was conducive to the sustained improvement. It was also indicated that there was a short-term and sustained learning effect, as the follow-up evaluation showed no advantage to other noises. These findings emphasise the importance of considering the background noise as well as the target stimuli in auditory perceptual learning studies. This is a new finding in auditory perceptual learning research area.

There are various types of background noise that are present in people’s daily life. Here, three types of background noise were investigated in this thesis. Although they covered a common range of background noise noisy communication situations in real life, there could be other scenarios where acoustics differ. Further research can be carried out to explore perceptual learning using other of background noise, such as such train station, TV/radio sound mixed with babble noise, and music noise conditions in daily life.

8.4 Further work

8.4.1 Active control group

The learning effects that were found in this thesis might be due to participants learning the tasks and being more comfortable with the test environment, rather than learning to perceive the stimuli more effectively. Therefore, in future studies, it is suggested that another active training control group is added, which will be tested with similar tasks and procedures as the trained group, to confirm the perceptual learning was not mainly induced by the learning the tasks or stimuli and being more comfortable with the test environment in this study.

8.4.2 Participants selection

This project was carried out with people with normal hearing, the training performances of hearing aid and cochlear implant users are still unknown. In addition, the age range of participants in this thesis was from 18 to 40 years with a mean age of around 29 years. All the participants in this project were adults with normal hearing thresholds. Halliday *et al.* (2008) stated that the improvement of auditory learning ability is varied across different ages. After auditory frequency discrimination training, the mean frequency discrimination thresholds for the oldest people in their study were slightly lower (improved). However, older people's frequency discrimination ability is not as good as younger adults who are aged between 18-40 years. Some children's frequency discrimination thresholds can achieve an adult's level, but the majority of children's performances show a fluctuating pattern of learning performance. Auditory learning has a prolonged developmental time course. The results of studies in this project might not be applicable in a clinic with a hearing aid or cochlear implant user who aged less than 18 years or older than 40 years. The application of findings from this project for children and elderly people with hearing-impairments still needs to be tested. Further research can be carried out using children and/or simulated age related hearing loss people (simulated people with high frequencies hearing loss), noised induced hearing loss people (an audiogram with notch around 4 kHz) or cochlear implant listeners before the application of the findings from this project are put to clinical use.

8.4.3 Objective tests

The test methods used in this project were all based on participants' subjective responses. There are several objective measurements, which can be recorded without participants' responses, to monitor listeners' brain activity in research and clinic, such as electroencephalogram (EEG, i.e. Cortical auditory evoked potentials-

CAEPs, Middle latency response-MLR, mismatch negativity-MMN) magnetoencephalographic (MEG) and functional Magnetic Resonance Imaging (fMRI) tests. Jäncke *et al.* (2001) demonstrated that compared with fMRI, both EEG and MEG can record fast neurophysiological process in milliseconds with low anatomical precision, while fMRI can record slower neurophysiological process in seconds with high anatomical precision. Kelly *et al.* (2005) demonstrated that the auditory evoked potentials (CAEPs, MLR and MMN) are affected by auditory experience and correlated with speech perception with experienced CI users. Other research (Anderson, *et al.*, 2013; Fallon *et al.,* 2008; Purdy *et al.,* 2001) of CAEPs and auditory perceptual learning also showed that auditory training benefits for speech perception in older HAs and CI adult users. CAEPs can evaluate the auditory performance improvement with minimal influence of non-auditory factor (i.e. motivation) (Sharma, *et al.,* 2014). Therefore, CAEPs can be used as an objective measurement of auditory perceptual learning to evaluate auditory functions (Anderson, *et al.,* 2013; Kelly, *et al.,* 2005). CAEPs are useful for recording neurophysiologic changes associated with training, but they are not essential for observing the training effects during the training session (Bishop, 2013; Sharma, *et al.,* 2014). Barlow *et al.* (2016) showed that CAEPs changes did not relate to behaviour performance. They trained adult CI participants for 7 hours with psychophysical tasks (Gap-in noise detection, Frequency discrimination, Spectral rippled noise - SRN, Iterated rippled noise, Temporal modulation) and evaluated CI users' speech performance using words (Lexical neighbourhood test - LNT) in both quiet and eight babble noise speaker conditions. CAEPs were used to evaluate pre- and post-training performance in both quiet and noisy conditions using a speech stimuli /baba/ with varied syllable stress. Results showed that the SNR thresholds improved during the training period, but showed no change on the other psychophysical tasks. The LNT speech performance improved almost 11% after training. The reasons for no correlation between CAEPs and behaviour measures may be due to the CI participants having already reached their plateau for speech performance before they attended the study. However, it should be noted that there was no control group (without training) in this study. Based on the results, it is not possible to evaluate how effective the training was.

The objective monitoring approaches are particularly useful to track progress in the developing auditory brain, especially in participants who are unable to give responses (e.g. pre-lingual cochlear implant users). It would be therefore useful to further explore the relationship between this present project's subjective test results and other objective tests to establish clinical standards to test hearing-impaired people who cannot provide responses themselves (i.e. infants).

8.4.4 Background noise

The background noises used in this PhD work were limited to babble noise, car and rain noise. In people's daily life, there are a variety of noise types. This study used only three types of background noise, which were too limited to represent the communication situations in real life. In addition, the general signal in this study was 65 dB SPL, which could only represent the normal speaker's speech levels in daily life. However, people have to listen to numerous levels of sound in their real life. Therefore, although this study was limited to the use of background noises, the same experimental design has the potential to assess multi-level and other types of noise situations. In addition, the study in Chapter 7 showed that benefits of training in babble noise generalized to car and rain background noise. But the training in babble produced substantially more improvement on performance in babble than on performance in other types of noise, so that generalization was, in fact, rather limited. Similarly, Green, Faulkner, and Rosen (2019) found that training with speech-in-babble-noise improved cochlear-implant users' perception of sentences in babble noise, but did not result in improved perception of phonemes in speech-shaped noise. These findings raised important questions for future research into the role of the background noise and target stimulus. It is suggested that future auditory perceptual training studies can make use of background noise to investigate other different types and levels of noise conditions for speech in noise tests.

8.4.5 Other languages

Except for the psychoacoustic AM sound perceptual study in Chapter 4, the target sounds used across the rest of the experiments in this thesis were carried out using the English language. However, different languages have different rhythms, stresses, and cadences, so the application of the findings from Chapters 5, 6, and 7 are limited to English native speakers and English. It is unknown whether the same perceptual learning patterns from English language would also be observed in other languages such as Chinese language. The speech spectrum of Mandarin is 3-4 dB higher than English from 2500 to 4000 Hz (Byrne *et al.,* 1994). From the frequency-importance function term, the spectrum regions with 2000-4000 Hz are more important for Chinese language than for English language (Chen *et al.,* 2008). From the speech intelligibility aspect (this relates to hearing aid fitting), research from Education University of Hongkong (2014) demonstrated that there are phonetics differences between Mandarin and English. For example, 12 consonants in English not found in the Mandarin initials (/v/, /θ/, /ð/, /z/, /s/, /ʃ/, /ʒ/, /h/, /tʃ/, /dʒ/, /r/, /j/). Although both consonants /b/ and /d/ are voiced in English, they are voiceless initials in Mandarin. This leads to Mandarin speakers with weaker pronunciation for voiced English consonants. Therefore, it is suggested that further research should be carried out using a different language (i.e. tonal language).

Chapter 9 Conclusion

This thesis focused on learning from auditory training studies to investigate ways of improving identification performance of target sounds in masking noise information in people with normal hearing. Psychoacoustic (AM sound) and speech stimuli (VCV nonsense stimuli and BKB sentences) were used as the test target sounds. The background noises used in this research were babble noise (fixed and random patterns) and environment sounds (car and rain). The studies in this thesis were carried out using people with normal hearing. A better understanding of auditory perceptual training for people with normal hearing could help devise better training for people with a hearing impairment. The learning outcomes from the studies reported in this thesis led to two suggested guidelines at the end of this chapter to build up clinical tools for training of hearing impaired persons to improve their hearing ability in everyday noisy environments.

This thesis included studies relating to both psychoacoustic and speech perceptual training approaches. For the psychoacoustic approach, AM stimuli were used to investigate whether SAM detection tasks training generalized to SAM-rate discrimination tasks with different modulation depths. Based on the test experiences and results from the preliminary psychoacoustic study, four perceptual learning experiments were conducted utilising the speech perceptual learning approach. It is known that the ability to detect speech signals in a noisy environment is critical in people's daily communication. Auditory perceptual learning studies are mainly focused on changing the target stimuli using amplitude modulated sounds or speech stimuli. Different studies use different background stimuli, babble noise or speech-shaped noise is commonly used. However, not many researchers have explored the auditory learning and generalization effect by changing background noise. The speech perceptual learning approach in this thesis was used to explore human perceptual learning performances by changing the test background noise. Three training experiments were conducted to test the identification performance of nonsense stimuli with fixed and random babble noises. Following the results, the last experiment investigated whether training with BKB speech sentences in random babble noise generalized to other environmental background noise conditions. The auditory perceptual learning experiments in this thesis support the following findings:

1. Even though SAM detection and SAM rate discrimination tasks shared similar stimuli features, there was no generalization from training with SAM detection tasks to SAM-rate discrimination tasks, regardless of the modulation depths. The results indicated that stimulus learning is not sufficient to improve perceptual learning between different SAM tasks (SAM detection and SAM rate discrimination task) (Chapter 4).

2. Auditory training with random babble noise produced better identification performance (for nonsense stimuli) than with fixed babble noise (learning effect). The perceptual learning effect from training with nonsense stimuli against fixed babble background noise did not generalize to nonsense stimuli against random babble background environmental noise condition (no generalization effect). The results suggested that perceptual learning of nonsense stimulus identification in noise is improved by random babble

noise training. In addition, training nonsense syllables in fixed noise has no learning effect when testing in random noise (Chapter 5).

3. Single session nonsense speech sounds recognition with fixed babble noise was similar to the condition with random babble noise. It confirmed that test method differences (multi-sessions vs single session) can lead to perceptual learning differences. Results also indicated that training and testing nonsense syllables with the same patterns (Fixed or Random) of babble noise can lead to performance improvement, but the improvements will not cause perceptual learning of nonsense stimulus identification differences. This similarity in result between noise conditions is despite participants appearing to have more confidence in the fixed condition due to the reduction of "Don't know" responses (Chapter 6).

4. Auditory learning effect from training speech in babble noise generalized to speech in car and rain background noise conditions. Both groups sustained their learning over a period of several weeks for speech-in-babble noise. As the control group received equal exposure to all three noise types, the sustained learning with babble noise, but not other noises, implies that a structural feature of babble noise was conducive to the sustained improvement (Chapter 7).

Understanding of speech in noise is important for hearing impaired people. In terms of providing guidelines on the learning from the findings of the reported studies which were mainly focused on non-speech and speech in noise experiments (findings 2-4 above). The following two guidelines can be used as baselines to develop an effective auditory training tool to help with training people with hearing impairment in challenging conditions.

The first guideline proposes that better auditory perceptual learning outcomes can be achieved by using random patterns of background noise from trial to trial, when training people to improve their hearing ability in noisy environments, rather than using a fixed pattern of background noise. While single session studies show improved performance on a fixed noise task (e.g. Felty *et al.*, 2009), it is the training on random noise that actually improves performance to stimuli detection in the more real life like random noise situations. It is recommended therefore that future studies focus on use of a random pattern of training background noise.

The second guideline proposes that given the limited generalization in Chapter 7, I suggested that using mixed types of noises would be advantages for perceptual learning and may boost the sustained learning from training to other background noises that is possible. Random babble noise is a good background noise to train with, due to some generalisation to other noise types (car, rain) – although only in the short term. Transfer learning could potentially be enhanced using mixed noise types but this would have to be further investigated. However, in order to apply the guideline for hearing impaired people in real life, it will be worthy to find out apart from babble noise, what other types of noise they are usually exposed to in their daily life (i.e TV, radio, music etc) and use those noises types for training.

As both normal hearing and hearing impaired people can improve their speech recognition performance in background noise via auditory perceptual training, it is likely that training people with hearing impairments in speech tasks with random noise conditions or changing background noise types would also be beneficial. In order to apply these learning outcomes to training with hearing-impaired people in clinic, further auditory training investigations would need to be conducted with hearing-impaired people. In clinic, hearing aid users mainly struggle with conversations in noisy environments and felt embarrassed to talk with people in these situations. If they received speech in noise training after they received their hearing aids, they may get more confidence in communication with others in noisy environments (Henshaw & Ferguson, 2013). As hearing impaired people require higher SNR levels and longer training duration to improve their performance in the same way as that of normal hearing people, these two parameters would have to be taken into consideration for the experimental design for auditory training with hearing-impaired people. Once the optimal SNR level and training duration are identified for training with hearing-impaired people, additional research can be conducted to investigate training in speech tasks with random noise conditions, or changing background noise. It would also better to think about training hearing impaired people at their individual level of difficulty, i.e. maintaining the speech in noise task above chance level of accuracy using speech reception threshold (SRT).

References

Action on Hearing Loss. (2014). *Hearing Matters.* London: Action on Hearing Loss.

Adank, P. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology Human Perception and Performance, 35(2)*, 520-529.

Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66, 610–618.

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: specificities to stimulus characteristics. *Vision Research, 36(21)*, 3487–3500.

Ahissar M., & Hochstein S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*, 457–464.

Ahissar, M., Laiwand, R., & Hochstein, S. (2001). Attentional demands following perceptual skill training. *Psychological Sci*ence*, 12*, 56–62.

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 285–299.

Amitay, S., Hawkey, D. J., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception and Psychophysics, 67*, 691–698.

Amitay, S., Irwin, A., & Moore, D. (2006). Discrimination learning induced by training with identical stimuli, *Nature Neuroscience, 9*, 446–1448.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42*, 529–555.

Anderson, S., Kraus, N. (2013). Auditory training: evidence for neural plasticity in older adults. *Perspectives on Hearing and Hearing Disorders: Research and Diagnostics,* 17(1), 37–57.

Atienza, M., Cantero, J.L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning Memory, 9*,138–150.

Attneave, F., & Olson, R. K. (1971). Pitch as medium: A new approach to psychophysical scaling. *American journal of Psychology, 84*, 147-166.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*, 233–250.

Bacon, S.P., Opie, J.M., & Montoya, D.Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech Language and Hearing Research, 41*, 549–563.

Baese-berk, M.M., Bradlow, A.R., & Wright, B.A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of American Express Letters, 133(3)*, 174-180.

Barlow, N., Purdy, S. C., Sharma, M. Giles, E., Narne, V. (2016) The Effect of Short-Term Auditory Training on Speech in Noise Perception and Cortical Auditory Evoked Potentials in Adults with Cochlear Implants. *Seminar Hearing*, 37(1), 84-98.

Bao, S., Chang, E.F., Woods, J., & Merzenich, M.M. (2004). Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. *Nature Neuroscience, 7*, 974–981.

Bench J, Kowal A,& Bamford J. (1979). The BKB (Bamford-Kowal-Bench) sentences lists for partially hearing children. *British Journal of Audiology*. 13, 108–112.

Ben-David, B.M., Campeanu, S., Tremblay, K.L., & Alain, C. (2010). Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiolgical, 48(6)*, 797–807.

Bernstein, J.G.W., & Grant, K.W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normalhearing and hearing-impaired listeners. *Journal of Acoustical Society of America,* 125, 3358-3372.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., & Ludvigsen, C. (1994). An international com- parison of long-term average speech, *Journal of Acoustical Society of America*, 96, 2108– 2126.

Bishop, D.V. (2013). Research Review: Emanuel Miller Memorial Lecture 2012—neuroscientific studies of intervention for language impairment in children:

interpretive and methodological problems. *Journal of Child Psychology and Psychiatry, and Allied Disciplines,* 54(3), 247–259.

Borrie S. A., McAuliffe M. J., & Liss J. M., (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech Language and Hearing Research, 55*, 290–305.

Bradlow, A. R., & Bent, T. (2003). Listener adaptation to foreign accented English. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, 2881–2884

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech, *Cognition, 106(2),* 707–729.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Bradlow, A.R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: sentence perception in noise. *Journal of Speech Language and Hearing Research, 46*, 80–97.

Brungart, D.S., Simpson, B.D., Ericson, M.A., & Scott, K.R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America, 110*, 2527-2538.

BSA. (2011). *Recommended Procedure: Pure Tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels.* The British Society of Audiology.

Burk, M.H., Humes, L.E., Amos, N.E., & Strauser, L.E. (2006). Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear Hearing, 27*, 263–278.

Burkholder, R. A. (2005). *Perceptual learning of speech processed through an acoustic simulation of a cochlear implant (Ph.D. thesis).* Indiana University, Indiana

Burns, OM., & Rajan, R. (2008). Learning in a task of complex auditory streaming and identification. *Neurobiology of Learning and Memory, 89(4)*, 448–461.

Buss, E. (2008). Across-channel interference in intensity discrimination: the role of practice and listening strategy. *Journal of Acoustical Sococity of Ammerica, 123,* 265–272.

Censor, N., Karni, A. & Sagi, D. (2006). A link between perceptual learning, adaptation and sleep. *Vision Research, 46 (23)*, 4071–4074.

Chen, J., Qu, T., Wu, X., Huang, Q., Huang, Y., Li, L., Chi, H. (2008). Frequency importance function of Mandarin Chinese speech. The Journal of the Acoustical Society of America, 123(5) 3323.

Clarke, C. M. (2000). Perceptual adjustment to foreign-accented English. *Journal of Acoustical Society America, 107*, 2856.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign accented English. *Journal of the Acoustical Society of America. 116(6)*, 3647–3658.

Cooke, M. A., (2006). Glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America, 119*, 1562–1573.

Corey, D. P., & Hudspeth, A. J., (1983). Kinetics of the receptor current in bullfrog saccular hair cells. *Journal of Neuroscience, 3*, 962–976.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269–284.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of Acoustical Society America, 116*, 3668–3678.

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 704–728.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettingan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise vocoded sentences. *Journal of Experimental Psychology: General, 134*, 222–241.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*, 132–147.

Delay, E. (2001). Cross-modal transfer effects on visual discrimination depends on lesion location in the rat visual system. *Physiology and Behaviour, 73(4)*, 609–620.

Demany, L. (1985). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America, 78*, 1118–1120.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America, 111*, 1377–1388.

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

De Vries, H. L. (1948). Brownian movement and hearing. *Physica, 14 (1)*, 48–60.

Dubno, J.R., Dirks, D.D., & Morgan, D.E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of Acoustical Society America, 76*, 87–96.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons. *Journal of the Acoustical Society of America, 74*, 739–743.

Dorman, M. F., & Wilson, B.S. (2004). The design and function of cochlear implants. *American Scientist*, *92*, 436-445.

Dupoux, E., & Green, K. P. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 914–927.

Edwards, B. W., & Viemeister, N. F. (1994). Frequency modulation versus amplitude modulation discrimination: Evidence for a second frequency modulation encoding mechanism. *Journal of the Acoustical Society of America, 96*, 733–740.

Education University of Hong Kong. (2014). Comparison of English and Mandarin (Segmentals), *http://corpus.eduhk.hk/English_Pronunciation/?page_id=328,* Accessed 20 December 2018.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, *67(2)*, 224–238.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability overtime. *Journal of the Acoustical Society of America, 119(4)*, 1950–1953.

Engen, K. J. V., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of Acoustical Society America, 121*, 519–526.

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

Fallon, J. B., Irvine, D. R., Shepherd, R. K. (2008). Cochlear implants and brain plasticity. *Hearing Research,* 238, 110–117.

Feddersen, W., Sandel, T., Teas, D., & Jeffress, L. A. (1957). Localization of high-frequency tones. *Journal of the Acoustical Society of America, 29*, 988–991.

Fenn, K.M., Nusbaum, H.C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425(6958),* 614–6.

Felty, R. A, Buchwald, A., & Pisoni, D. B., (2009). Adaptation to frozen babble in spoken word recognition. *Journal of Acoustical Society America,125(3)*, EL93–EL97.

Fitzgerald, M. B. & Wright, B.A. (2011). Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task. *Journal of the Acoustical Society of America, 129(2)*, 898–906.

Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics, 16*, 425–442.

Francis, A. L., Fenn, K., & Nusbaum, H. C., (2007). Effects of training on the acoustic phonetic representation of synthetic speech. *Journal of Speech Language & Hearing Research, 50*, 1445–1465.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology Human Perception & Performance, 28*, 349–366.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention Perception & Psychophysics, 71*, 1360–1374.

Fu, Q. J. (2002). Temporal processing and speech recognition in cochlear implant users, *Neuroreport, 13*, 1635–1639.

Fu, Q. J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*, 19–27.

Fu, Q. J., Nogaki, G., & Galvin, J. J. III (2005). Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation, *Journal of the Association for Research in Otolaryngology, 6*, 180–189.

Gaab, N., Paetzold, M., Becker, M., Walker, M.P., & Schlaug, G. (2004). The influence of sleep on auditory learning – A behavioral study. *Neuroreport, 15*, 731–734.

Gass, S., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65–89.

Gökaydin, D., Ma-Wyatt, A., Navarro, D., & Perfors, A. (2011). Humans use different statistics for sequence analysis depending on the task. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Austin*, 543–548

Goldstone, R. L. (1998). Perceptual learning, *Annual Review of Psychology, 49*, 585–612.

Graddol, D. (1997). *The Future of English?* The British Council, London.

Grose, J. H., Hall, J. W. (1993). Comodulation masking release: is comodulation sufficient? *Journal of The Acoustical Society of America, 93*, 2896–2902.

Grant, K. W., Summers, V., & Leek, M. R. (1998). Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners. *Journal of Acoustical Society America, 104*, 1051–1060.

Green, T., Faulkner, A., & Rosen, S. (2019). Computer-based connected-text training of speech-in-noise perception for cochlear implant users. *Trends in Hearing, 23*, 1-11.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 421–433.

Gilbert, C.D. (1994). Early perceptual learning. *Proceedings of the National Academy of Sciences, 91*, 1195–1197.

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. *Hearing Research, 184*, 41–50.

Gustafsson, H. A., Arlinger, S. D. (1994). Masking of speech by amplitude modulated-noise. *Journal of the Acoustical Society of America, 95*, 518–529.

Gutnick, H. N. (1982). Consonant-feature transmission as a function of presentation level in hearing-impaired listeners. *Journal of Acoustical Society America, 72(4)*, 1124–1130.

Habib, M., Davaure, V., Camps, R., Espesser, R., & Joly-Pottuz, B. (2002). Phonological training in children with dyslexia using temporally modified speech: A three-step pilot investigation. *International Journal of Language & Communication Disorders, 37*, 289–308.

Hall, J. W. III, and Grose, J. H. (1994). Development of temporal resolution in children as measured by the temporal modulation transfer function. *Journal of the Acoustical Society of America, 96*, 150–154.

Hallé, P.A., Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behaviour and Development, 17*, 119–129.

Halliday, L. F., Taylor, J. L., Edmondson-Jones, A. M., & Moore, D. R. (2008). Frequency discrimination learning in children. *The Journal of the Acoustical Society of America,* 123, 4393–4402.

Halliday, L. F., Taylor, J. L., Millward, K. E. & Moore, D. R. (2012). Lack of Generalization of Auditory Learning in Typically Developing Children. *Journal of Speech Language and Hearing Research, 55(1)*, 168-181.

Harris, G. G. (1968). Brownian motion in the cochlear partition. *Journal of Acoustical Society America, 44*, 176–186.

Hawkey, D. J., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7,* 1055–1056.

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

Helfer, K.S., Wilber, L.A. (1990). Hearing-loss, aging, and speech-perception in reverberation and noise. *Journal of Speech and Hearing Research, 33*, 149–155.

Henshaw, H., & Ferguson, M.A., (2013). Efficacy of individual computer-based auditory training for people with hearing loss: asystematic review of the evidence. *PLoS One, 8(5)*, e62836.

Hervais-Adelman, A., Davis, M. H., Taylor. K., Johnsrude, I.S., & Carlyon, R.P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 283–295.

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research, 37*, 2133–2141.

Howard-Jones, P.A., & Rosen, S. (1993). Uncomodulated glimpsing in 'checkerboard' noise. *Journal of the Acoustical Society of America, 93*, 2915-2922.

Irvine, D. R. F., Martin, R. L., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *Journal of the Acoustical Society of America, 108*, 2964–2968.

Jäncke, L, Gaab, N., Wüstenberg, T., Scheich, H., Heinze, H., J. (2001). Short-term functional plasticity in the human auditory cortex: an fMRI study. *Cognitive Brain Research,* 12, 479-485.

Jennings, A. R. (2005). *On mechanisms for the analysis of spectral and temporal envelope shape in the human auditory system (Ph.D. thesis).* Newcastle University, Newcastle.

Jenkins, J. (2000). T*he Phonology of English as an International Language.* Oxford: Oxford University Press.

Karmarkar, U., & Buonomano, D.V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning Memory, 10*, 141–147

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard P., & Adams M. M. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceeding of the National Academy of Sciences of the United States of America, 95*, 861–868

Karni. A, & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences, 88*, 4966–4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 365(6443)*, 250–252.

Kates, J. M., & Weiss, M. R. (1996). A comparison of hearing-aid array-processing techniques, *Journal of Acoustical Society America,* 99, 3138-3148.

Kelly, A. S., Purdy, S.C., Thorne, P, R. (2005). Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clinical Neurophysiology,* 116,1235–1124.

Kidd, G. R., Watson, C. S., & Gygi, B., (2007). Individual differences in auditory abilities. *Journal of Acoustical Society America, 122*, 418–435.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Foster, B. B., & Liddle, P. F. (2001). An event-related fMRI study of visual and auditory oddball tasks. *Journal of Psychophysiology, 15*, 221–240.

Killion, M. C., Niquette, P. A., & Gudmundsen, G. I., (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of Acoustical Society America, 116*, 2395–2405.

Kim, S. H., Frisina, R. D., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2006). Effect of age on binaural speech intelligibility in normal hearing adults. *Speech Communication, 48*, 591-597.

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science, 16,* 617–625.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*, 141–178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Cognitive Psychonomic Bulletin & Review, 13(2)*, 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*, 1–15.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107*, 54–81.

Langhans, A., & Kohlrausch, A., (1992). Differences in auditory performance between monaural and diotic conditions. I. Masked thresholds in frozen noise. *Journal of Acoustical Society America, 91*, 3456–3470.

Leek, M.R. (2001). Adaptive procedures in psychophysical research. *Perception and Psychophysics, 63(8)*, 1279–1292.

Levitt, H. (2001). Noise reduction in hearing aids: A review. *Journal of rehabilitation research and development, 38*, 111–121.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental stimuli. *Cerebral Cortex, 14*, 1008–1021.

Linkenhoker, B.A. & Knudsen, E.I. (2002). Incremental training increases the plasticity of the auditory space map in adult barn owls. Nature 419, 293–296.

Loebach, J. L., Bent, T., & Pisoni, D. B. (2008). Multiple routes to the perceptual learning of speech. *Journal of the Acoustical Society of America, 124(1)*, 552–561.

Loebach, J. L., & Pisoni, D. B., (2008). Perceptual learning of spectrally-degraded speech and environmental sounds. *Journal of the Acoustical Society of America, 123*, 1126–1139.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2009). Transfer of auditory perceptual learning with spectrally reduced speech to speech and nonspeech tasks: implications for cochlear implants. *Ear Hear, 30*, 662–674.

Loizou, P. C. (1999). Introduction to cochlear implants. *IEEE Engineering in Medicine and Biology, 18(1),* 32-42.

Lorenzi, C., Dumont, A., & Fullgrabe, C. (2000). Use of temporal envelope cues by children with developmental dyslexia, J*ournal of Speech Language and Hearing Research, 43*, 1367–1379.

Mathers, C., Smith, A., & Concha, M. (2000). *Global burden of hearing loss in the year 2000.* Geneva: WHO.

Maye, J., Aslin, R., & Tanenhaus, M. (2003). In search of the weckud wetch: Online adaptation to speaker accent. *In Proceedings of the 16th Annual CUNY Conference on Human Sentence Processing*, Cambridge, MA.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32*, 543–562.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research, 26*, 451–458.

McQueen, J. M., Cutler, A., Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*, 1113–1126.

Mednick, S. C., Arman, A. C., & Boynton G. M. (2005). The time course and specificity of perceptual deterioration. *Proceedings of the National Academy of Sciences of the United States of America, 102(10)*, 3881–3885.

Menning, L., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport, 11*, 817–822.

Melara, R. D., Rao, A., & Tong, Y. (2002). The duality of selection: Excitatory and inhibitory processes in auditory selective attention. *Journal of Experimental Psychology: Human Perception & Performance, 28*, 279-306.

Micheyl, C., Delhommeau, K., Perrot, X. & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*, 36–47.

Millward, K. E., Hall, R. L., Ferguson, M. A., & Moore, D. R. (2011). Training speech-in-noise perception in mainstream school children, *International Journal of Pediatric Otorhinolaryngology, 75(11)*, 1408-1417.

Mollon, J. D., & Danilova, M. V. (1999). Three remarks on perceptual learning. *Spatial Vision*, *10(1):5*

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of Acoustical Society America, 54*, 610–619.

Moore, B. C. J. (2004). *An Introduction to the Psychology of Hearing.* London: Academic Press, 5th edition.

Moore, D. R., Amitay, S., & Hawkey, D. (2003) Auditory perceptual learning. *Learn Mem, 10*, 83–85.

Moore, D. R., Ferguson, M. A., Halliday, L. F., & Riley, A. (2008). Frequency discrimination in children: Perception, learning and attention. *Hearing Research, 238*, 147–154.

Moore, B. C. J., & Glasberg, B. R. (1989). Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. *Journal of Acoustical Society America, 86*, 1722–1732.

Moore, D. R., Rosenburg, J., & Coleman, J. (2005). Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children, *Brain & Language, 94,* 72–85.

Moore, D., & Shannon, R. (2009). Beyond cochlear implants: awakening the deafened brain. *Nature Neuroscience, 12*, 686–691.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behaviour, 16*, 519–533.

Müller, G., & Pilzecker, A. (1900). Experimental contributions to the theory of memory. *Zeitschrift für Psychologie Eganzungsband, 1*, 1–288.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*, 289–306.

Näätanen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13*, 201–288.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perception learning in speech. *Cognitive Psychology, 47*, 204–238.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60(3)*, 355–376.

Ortiz, J. A. & Wright, B. A. (2003). Different effects of overnight consolidation on three types of learning in interaural-time-difference discrimination. *Association for Research in Otolaryngology Abstracts, 26*, 76.

Ortiz, J. A. & Wright, B. A. (2005). Effects of different amounts of brief training and rest on the generalization of learning from interaural level - difference to interaural-time-difference discrimination. *Journal of Acoustical Society America, 117*, 2561.

Ortiz, J. A. & Wright, B. A. (2009). Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *Journal of experimental psychology human perception and performance, 35(1)*, 188-194.

Overath, T., McDermott, J.H., Zarate, J.M. and Poeppel, D. (2015). The cortical analysis of speech-specific temporal struature revealed by response to sound quilts. *Nature Neuroscience, 18*, 903-911.

Patterson, R. D., Johnson-Davies, D., & Milroy, R. (1978). Amplitude modulated noise: The detection of modulation versus the detection of modulation rate, *Journal of the Acoustical Society of America, 63*, 1904–1911.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS One,* 6:e18082.

Paffen, C.L., Verstraten, F.A., Vidnyánszky, Z. (2008). Attention-based perceptual learning increases binocular rivalry suppression of irrelevant visual features. *Journal of Vision, 8(4)*, 1–11.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology Human Perception & Performance, 31*, 1315–1330.

Petkov, C., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Loods, D. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience, 7(6)*, 658–663.

Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *Journal of Acoustical Society America, 121*, 2312–2316.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing.* London: Academic Press.

Plack, C. J., & Carlyon, R.P. (1995). *Loudness perception and intensity coding.* Handbook of perception and Cognition, Orlando: Academic Press.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of Acoustical Society America, 63*, 533–549.

Plomp, R. (1983). The role of modulation in hearing. In R. Klinke & R. Hartman (Eds), *Hearing-Physiological Bases and Psychophysics* (pp. 270-276). Berlin: Springer-Verlag.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Reorganization through top-down influences. *Journal of Neuroscience, 26*, 4970–4982.

Purdy, S. C., Kelly, A. S., Thorne, P. R. (2001) Auditory evoked potentials as measures of plasticity in humans. *Audiology Neurootology,* 6(4), 211–215.

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine, 13*, 214–232.

Reed, C. M., Delhorne, L. A., (2005). Reception of environmental sounds through cochlear implants. *Ear and Hearing, 26(1),* 48–61.

Rhebergen K. S., & Versfeld, N. J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *Journal of Acoustical Society America, 117*, 2181–2192.

Ricketts, T. A., & Hornsby, B. W., (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction. *Journal of the American Academy of Audiology, 16*, 270-277.

Robinson, K., Summerfield, A. Q. (1996), Adult auditory learning and training, *Ear Hear, 17*, 51–65.

Rocheron, I., Lorenzi, C., Fullgrabe, C., & Dumont, A. (2002). Temporal envelope perception in dyslexic children, *NeuroReport, 63*, 1904–1911.

Rochet, B.L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.Timonium* (pp. 379-410). MD: York.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. *In Varieties of Memory and Consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory, and linguistic aspects. *Philosophical transactions of the royal society of London series a mathematical physical and engineering sciences, 336(1278)*, 367-373

Roth, D. A., Amir, O., Alaluf, L., Buchsenspanner, S. & Kishon-Rabin, L. (2003). The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear. *Journal of Basic and Clinical Physiology and Pharmacology, 14*, 137–150.

Roth, D. A., Kishon-Rabin, L., Hildesheimer, M. & Karni, A. (2005). A latent consolidation phase in auditory identification learning: Time in the awake state is sufficient. *Learn Memory, 12*, 159–164

Rowan, D. & Lutman, M. (2005). Generalisation of learning with ITD discrimination across frequency and type of cue. *Association for Research in Otolaryngology Abstracts, 28*, 257.

Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology, 7(7)*, 461–467.

Saberi, K. & Perrott, D.R. (1990). Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. *Journal of Acoustical Society America, 87*, 1732–1737.

Saberi, K., & Green, D. M. (1997). Evaluation of maximum-likelihood estimators in nonintensive auditory psychophysics. *Perception & Psychophysics, 59*, 867-876.

Sarro, E. C. & Sanes, D. H. (2009). Prolonged maturation of auditory perception and learning in gerbils. *Developmental Neurobiology, 70*, 636–648.

Savion-Lemieux, T., & Penhune, V. B. (2005). The effects of practice and delay on motor skill learning and retention. *Experimental Brain Research, 161*, 423–431.

Schäffler, T., Sonntag, J., Hartnegg, K., & Fischer, B. (2004). The effect of practice on low-level auditory discrimination, phonological skills, and spelling in dyslexia. *Dyslexia, 10*, 119–130.

Schlauch, R. S., & Rose, R. M. (1990). Two-, three-, and four-interval forced-choice staircase procedures: Estimator bias and efficiency. *Journal of the Acoustical Society of America, 88*, 732-740.

Schow R., Nerbonne M., (2006). *Introduction to Audiologic Rehabilitation*. Boston, MA: Pearson Education.

Schubö, A., Schlaghecken, F., & Meinecke, C. (2001). Learning to ignore the mask in texture segmentation tasks. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 919–931.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395–408.

Semal, C., & Demany, L. (1990). The upper limit of "musical" pitch. *Music Perception, 8*, 165-176.

Seitz, A. R., Nanez, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision, 6*, 966–973.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303–304.

Shafiro, V., Sheft, S., Gygi, B., Ho, K.T. (2012). The influence of environmental sound training on the perception of spectrally degraded speech and environmental sounds. *Trends in Amplification, 16*, 83–101.

Sharma M, Purdy S C, Kelly A S. (2014). The contribution of speech-evoked cortical auditory evoked potentials to the diagnosis and measurement of intervention outcomes in children with auditory processing disorder. *Seminar Hearing,* 35(1), 51–64.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243–261.

Shinn-Cunningham, B.G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*, 182-186.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. P*erception & Psychophysics, 52(5),* 582–588.

Shelton, B. R., & Scarrow, I. (1984). Two-alternative versus three alternative procedures for threshold estimation. *Perception & Psychophysics, 35*, 385-392

Shofner, W. P., & Niemiec, A. J. (2010). Comparative psychoacoustics. *Oxford Handbook of Auditory Science: Hearing, 3*, 145.

Sidaras Sabrina, K., Alexander Jessica, E. D., & Nygaard Lynne, C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America, 125*, 3306–3316.

Song, J.H., Skoe, E., Banai, K., & Kraus, N. (2011). Training to improve hearing speech in noise: biological mechanisms. *Cereb.Cortex, 22*, 1180–1190.

Stacey, P. C., & Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *Journal of the Acoustical Society of America, 121*, 2923–2935.

Stacey, P. C., & Summerfield, A. Q. (2008). Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *Journal of Speech, Language and Hearing Research, 51*, 526-538.

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research.* Baltimore: York Press.

Sweetow, R. W., & Sabes, J. H. (2006). The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program. *Journal of the American Academy of Audiology, 17*, 538-558.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science, 8*, 432–443.

Szpiro, S. F., Wright, B. A., & Carrasco, M. (2014). Learning one task by interleaving practice with another task. *Vision Research, 101*, 118–124.

Takagi, N., & Mann, V. A. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics, 16*, 379–405.

Tremblay, K. L., (2007). Training-related changes in the brain: evidence from human auditory-evoked potentials. *Semin Hear, 28*, 120–132.

Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech Language and Hearing Research, 45*, 564–572.

Tremblay, K. L., & Kraus, N. & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *NeuroReport, 9*, 3557–3560.

Tremblay, K. L., Shahin, A. J., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology, 120*, 128–135.

Van Engen, K. J. (2012). Speech-in-speech recognition: a training study. *Language and Cognitive Processes, 27(7-8)*, 1089-1107.

Van Wassenhove, V., & Nagarajan, S.S. (2007). Auditory cortical plasticity in learning to discriminate modulation rate. *Journal of Neuroscience, 7*, 2663–2672.

Van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication, 35*, 103–113.

Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *Journal of the Acoustical Society of America, 130*, 429–439.

Vogels, R., & Orban, G. A. (1985). The effect of practice on the oblique effect in line orientation judgments. *Vision Research, 25(11)*, 1679–1687.

Weil, S. A. (2001). Foreign accented speech: Encoding and generalization. *Journal of Acoustical Society America, 109*, 2473(A).

Wiesenfeld, K. & Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and squids. *Nature, 373*, 33–36.

Wilson, B. & Dorman, M. (2008). Cochlear implants: A remarkable past and brilliant future. *Hearing Research*, 242:3-21.

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology, 14*, 453–470.

Wingstedt, M., & Schulman, R. (1984). Comprehension of foreign accents. In W. Dressler, H. Luschutzky, O.Pfeiffer, & J. Rennison (Eds.). *Phonologica* (pp. 339-345), Cambridge: Cambridge University Press.

Witton, C., Stein, J. F., Stoodley, C. J., Rosner, B. S., & Talcott, J. B. (2002). Separate influences of acoustic AM and FM sensitivity on the phonological decoding skills of impaired and normal readers*, Journal of Cognitive Neuroscience, 14*, 866–874.

Wong, P., Ettlinger, M., Sheppard, J., Gunasekera, G., & Dhar, S. (2010). Neuroanatomical characteristics and speech perception in noise in older adults. *Ear Hear, 31*, 471–479.

Wong, P. C. M., & Perrachione, T. K., (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistic, 28*, 565-585.

WHO (2004). *The global burden of disease: 2004 update.* Geneva: WHO.

Wright, B. A. (2001). Why and how we study human learning on basic auditory tasks. *Audiology and Neuro-Otology, 6(4)*, 207–210

Wright, B. A., Buonomano, D. V., Mahncke, H. W., & Merzenich, M. M. (1997). Learning and generalization of auditory temporal-interval discrimination in humans. *Journal of Neuroscience, 17(10)*, 3956–3963.

Wright, B. A. & Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences, 98*, 12307–12312

Wright, B. A., Sabin, A. T. (2007). Perceptual learning: how much daily training is enough? *Experimental Brain Research, 180(4)*, 727–736.

Wright, B. A., Sabin, A. T., Zhang, Y., Marrone, N., & Fitzgerald, M. B. (2010). Enhancing perceptual learning by combining practice with periods of additional sensory stimulation. *Journal of Neuroscience*, 30, 12868–12877.

Wright, B. A., Wilson, R. M., & Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task, *Journal of Neuroscience, 30*, 11635–11639.

Wright, B. A. & Zhang, Y. (2006). A review of learning with normal and altered sound-localization cues in human adults. *International Journal of Audiology, 45*, 92–98.

Wright, B. A. & Zhang Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society of London Series a Mathematical and Physical Sciences, 364(1515)*, 301–311.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistic, 28*, 565-585.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., Clarke, W. D. III (1994). The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. P*roceedings of the International Conference of Spoken Language Processing*, Yokohama: Acoustical Society of Japan, 2023–2026.

Yost, W. A. (2007). *Fundamentals of Hearing: 5th edition*. New York: Academic.

Yotsumoto, Y., & Watanabe, T. (2008). Defining a link between perceptual learning and attention. *PLoS Biol, 6(8)*, 1623–1625.

Yund, E. W., Woods, D. L., (2010). Content and procedural learning in repeated sentence tests of speech perception. *Ear Hear, 31*, 769–778.

Zhang, Y. & Wright, B. A. (2009). An influence of amplitude modulation on interaural level difference processing suggested by learning patterns of human adults. *Journal of Acoustical Society America, 126*, 1349–1358.

Zhang, J., Zhang, G., Xiao, L., Klein, S., Levi, D., & Yu, C., (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience, 30*, 12323–12328.

Zeng, F., Fu, Q., & Morse, R. (2000). Human hearing enhanced by noise. *Brain Research, 869*, 251–255.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences, 102*, 14110–14115.

Ziegler, J.C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech perception- in-noise deficits in dyslexia. *Developmental Science, 12*, 732–745.

Appendix 1 Consent form

THE UNIVERSITY OF WARWICK

CONSENT FORM

(Biomedical and Scientific Research Ethics Committee) Study Number:

Patient Identification Number for this study:

Title of Project:

Name of Researcher(s): Miss Liping Zhang & Professor Paul Jennings

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated **04/02/2013** (version 2012-13.02) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my legal rights being affected. ☐
3. I understand that the data collected during the study, may be looked at by individuals from The University of Warwick, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐
4. I agree to take part in the above study. ☐

Name of Participant	Date	Signature
Name of Person taking consent	Date	Signature

BSREC Consent form template; version number: 2012-13.01; Version date: 01Oct12
Investigator's consent form date of issue: **03/03/13**
Investigator's consent form version number: **1.0**

Appendix 2 Particianpt information leaflet

THE UNIVERSITY OF
WARWICK

PARTICIPANT INFORMATION LEAFLET

Study Title:	**Generalization resulting from training on SAM-detection task to SAM-rate Discrimination task with different depths**
Investigator(s):	**Liping Zhang (PhD Student in Engineering, WMG)** **Dr James Harte (Assistant Prof. of Biomedical Engineering, WMG)**

Introduction

You are invited to take part in a Research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take the time to read the following information carefully. Talk to others about the study if you wish.

(Part 1 tells you the purpose of the study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study)

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

PART 1

What is the study about?

Speech is one of the most important sounds that our auditory system needs to process; it carries information in a robust and redundant way, making it a reliable means for communication when significant distortion removes parts of, or background noise masks it. A person with normal hearing can make use of the context, rhythm, stress and intonation in speech to work out what the missing components are and make sense of it. From a purely physical acoustics point of view, much of the information in speech is carried in the rapid fluctuations in sound pressure amplitude and frequency over time - which we term as amplitude and frequency modulation respectively. This study aims to investigate if it is possible to improve our brains ability to process auditory stimuli by training a listener to identify smaller and subtler changes in amplitude across time. We aim to see if practise in normal hearing listeners can lead to better performance to detect the changes in amplitude and frequency modulated stimuli. It is envisaged that the results of this study will contribute to new methods of auditory training for hearing impaired individuals in the future.

Do I have to take part?

It is entirely up to you to decide. We will describe the study and go through this information sheet, which we will give you to keep. If you choose to participate, we will ask you to sign a consent form to confirm that you have agreed to take part (if part of this study is an online or postal questionnaire/survey, by returning a completed questionnaire/survey, you are giving your consent for the information that you have supplied to be used in this study and formal signed consent will not be collected where postal or online questionnaires/surveys are concerned). You will be free to withdraw at any time, without giving a reason and this will not affect you or your circumstances in any way.

What will happen to me if I take part?

You will be asked to attend between 2 and 9 experimental sessions, depending on whether you assigned as part of a control or test group. This choice is predominantly on a random basis.

The first session will last up to 2 hours maximum. The first session is both a screening session to ensure participants meet the study inclusion criteria and an experimental session of this study as well. Durning this session, the status of your hearing will be determined by an audiometry - where you will listen to a series of pure tones whose level will be reduced until you can only just hear them. Your pure tome audiogram threshold will be obtained to ensure your hearing threshold is within normal levels (Pure tone audiometry threshold ≤20 dB HL). Your ear canal will also be examined via an otoscope, to make sure you do not have any infections, wax or drum perforation etc. If we notice your hearing levels are above 20 dB HL or your ear canal has large wax buildups, we will not continue the experiment and you may consider seeing your GP to discuss further clinic tests. We will provide you with publicly available information on hearing loss from the UK charity action on hearing loss. It will help you to decide what to do next. We will try our best to answer any question you may have, but cannot provide clinic advice as neither of the study investigators are clinically registers. It should be noted that any hearing loss detected is likely to be small otherwise it would be already be known to you as it would significantly affect your daily life.

After this you can begin the experiment, where you will be seated on a comfortable chair in a sound-proofed room. we will place a set of headphones on your head, and you will be asked to sit still and concentrated to listen to a series of three different sounds(one target and two standard sounds), after that, you will be required to make a choice(Decide which one is the target sound) from the three. If at any time you feel uncomfortable, and want to stop the experiment, or simply require a comfort break, you can press a handheld button which will inform us of your desire to suspend/stop the experiment. The purpose of this session is to determine your baseline abilities in performing these listening experiments. We expect a large variation across subjects, and you are not judged on your performance - our interest is in how you improve over time with practice.

For those in the 'test group', we will then arrange a series of 7 x 1 hour training sessions on consecutive days when it is convenient for you.

All subjects will be invited to the last experimental session, that repeats the measurements from the first session. Thus we will compare the pre- and post-training performance in the test group, and have a baseline comparision in the control group.

What are the possible disadvantages, side effects, risks, and/or discomforts of taking part in this study?

The volume of the sounds you will hear will be limited to a comfortable and completely safe level. Our experiment is governed by the 2005 Control of Noise at Work regulations. Essentially, you will be exposed to noise/sound at well below acceptable daily exposure levels, known not to lead to hearing losses or tinnitus.

The experiment will be conducted in a large sound proof room, with the experimenter sitting outside. The lights will be on, and the room is very large so it is unlikely that you should feel discomfort. However, if at any time you feel uncomfortable and want to stop the experiment, or simply require a comfort break, you can press a handheld button which will inform us of your desire to suspend/stop the experiment.

What are the possible benefits of taking part in this study?

You will be paid £5/hr for your time and you will be making a contribution to basic hearing research.

Expenses and payment

You will be paid £5/hr for your participation in this experiment to cover the cost of your time.

What will happen when the study ends?

The results of this research study will be submitted to a peer reviewed scientific journals or conferences. You personal informantion will not be identified in any publication. You will have access to such a publication if required.

Will my taking part be kept confidential?

Yes. We will follow strict ethical and legal practice and all information about you will be handled in confidence. Further details are included in Part 2.

What if there is a problem?

Any complaint about the way you have been dealt with during the study or any possible harm that you might suffer will be addressed. Detailed information is given in Part 2

This concludes Part 1.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

PART2

Who is organising and funding the study?

University of Warwick

What will happen if I don't want to carry on being part of the study?

Participation in this study is entirely voluntary. Refusal to participate will not affect you in any way. If you decide to take part in the study, you will need to sign a consent form, which states that you have given your consent to participate.

If you agree to participate, you may nevertheless withdraw from the study at any time without affecting you in any way.

You have the right to withdraw from the study completely and decline any further contact by study staff after you withdraw.

What if there is a problem?

This study is covered by the University of Warwick's insurance and indemnity cover. If you have an issue, please contact Jo Horsburgh (details below).

Who should I contact if I wish to make a complaint?

Any complaint about the way you have been dealt with during the study or any possible harm you might have suffered will be addressed. Please address your complaint to the person below, who is a Senior University of Warwick official entirely independent of this study:
Jo Horsburgh
Deputy Registrar
Deputy Registrar's Office
University of Warwick
Coventry, UK, CV4 8UW.
T: +00 44 (0) 2476 522 713 E: J.Horsburgh@warwick.ac.uk

Will my taking part be kept confidential?

Yes. All information which is collected about you during the course of the research will be kept strictly confidential. You will not be identified in any report/publication.

What will happen to the results of the study?

The results will be published in conference or an international peer reviewed journals.
Who has reviewed the study?
This study has been reviewed and given favourable opinion by the University of Warwick’s Biomedical and Scientific Research Ethics Committee (BSREC) .

What if I want more information about the study?

If you have any questions about any aspect of the study or your participation in it not answered by this participant information leaflet, please contact:

Liping Zhang,
PhD student
Institute of Digital Healthcare, WMG,
University of Warwick

Email: Zhang_l@wmg.warwick.ac.uk
Phone: 02476173764

Thank you for taking the time to read this participant information leaflet.

Appendix 3 Ethical protocol

THE UNIVERSITY OF WARWICK

BIOMEDICAL & SCIENTIFIC RESEARCH ETHICS COMMITTEE

(BSREC)

PROTOCOL GUIDANCE

Title:	**Generalization resulting from training on SAM-detection task to SAM-rate Discrimination task with different depths**
Abstract:	Speech is one of the most important sounds that our auditory system needs to process; it carries information in a robust and redundant way, making it a reliable means for communication when significant distortion removes parts of, or background noise masks it. A normal hearing person can make use of the context, rhythm, stress and intonation in speech to work out what the missing components are and make sense of it. From a purely physical acoustics point of view, much of the information in speech is carried in the rapid fluctuations in pressure amplitude and frequency over time - which we term as amplitude and frequency modulation respectively. This study aims to investigate if it is possible to improve our brains ability to process auditory stimuli by training a listener to identify smaller and subtler changes in amplitude across time. We aim to see if practise in normal hearing listeners can lead to better performance to detect the changes in amplitude and frequency modulated stimuli. It is envisaged that the results of this study will contribute to new methods of auditory training for hearing impaired individuals in the future.
Contact details	
Chief Investigator: *NB: If this study is below PhD level, the CI will be your Academic Supervisor*	Dr James Harte
Principal Investigator(s):	Miss Liping Zhang
Background – why are you researching this area? What does the previous evidence say? What are the gaps in knowledge?	
The cues of fluctuation or modulation in sounds are important to obtain	

critical information from sounds or speech, so the enhancement of this ability can improve people's hearing performance (Plomp, 1983; Rosen, 1992). It is assumed that practise can lead to better performance to detect the changes in amplitude modulation stimulus, especially for people with problems in detecting amplitude modulated sounds. Historical research has also indicated that human's perceptual skills to detect and discriminate sounds can be improved after certain amounts of auditory training (Hall and Grose, 1994; Hawkey, Amitay and Moore, 2004).

In theory, sinusoidal amplitude modulation (SAM)-Detection and SAM-rate-Discrimination tests have different cues for neural substrates to process during decision making (Fitzgerald and Wright, 2011). The SAM-Detection test mainly focuses on the differences of depths from the target to standard stimulus. While the modulation rate difference between the target stimulus and the standard one is the critical cue for SAM-rate-Discrimination condition.

Wright and Zhang (2009) showed that auditory learning ability generalize across frequency, ear, stimulus duration, different presentation style etc. However, Fitzgerald and wright (2011) argued that the cross-learning effect could not generalize from SAM-Detection to SAM-rate-discrimination. Fitzgerald and wright (2011) used a 100% modulation depth for the SAM-rate–discrimination tasks in their study. Patterson et al (1978) indicated that 100% modulation depth for a discrimination test is too high to get the optimal rate-Discrimination threshold. Therefore, this study hypothesises that the generalization effect may occur from SAM-Detection to SAM-rate-discrimination, if significantly lower modulation depths are used for the SAM-rate-Discrimination tasks.

Aims/Objectives and Purpose of the study

This project aims to see whether there will be a generalization effect from training on SAM-detection test to SAM-rate Discrimination test with three different fixed modulation depths.

Objectives

1. Compare the SAM-detection thresholds (see note 1*) from pre and post SAM-detection test results to find whether there is an improvement after the training session.

2. Compare the SAM-rate-Discrimination thresholds (see note 2*) from the pre and post-test of SAM-rate-Discrimination tests to see whether there is generation effect (see note 3*) from SAM-detection training to SAM-rate-Discrimination with different fixed modulation depths.

3. Compare the SAM-rate-Discrimination thresholds from the three different modulation depths of SAM-rate-Discrimination tests to investigate which SAM-rate-Discrimination modulation depths will obtain the largest improvement after the SAM-Detection training.

Notes:

1* SAM-detection threshold is the minimum difference in the SAM depth of the target sound that needs to be detected from the standard SAM sound. During the test, it is usually measured in logarithmic scale (in dB) (see the details in the procedure part).

2* SAM-rate-Discrimination threshold is the minimum difference in the SAM rate that requires to be discriminated between a faster SAM rate (target sound in test) and the standard SAM rate. It is measured as a function of modulation rate and unit in Hz.

3* Generation effect: it standards for the crossing leaning effect from one task to another different task.

Design/Methodology – please include information about whether your study is qualitative/quantitative, retrospective/prospective; what interventions you will use (e.g. describe all surveys, tests, observations); what sample size you will use; how you reached that sample size and how you will analyse the data; how you will recruit/select your participants/subjects; what allowances will you put in place so that participants/subjects can withdraw from the study at any time; and, what your process is for gathering informed consent:

This is a quantitative study generating primary data.

Test subjects will be recruited from the student and staff population of the University of Warwick, this will be carried out by word of mouth and by simple advertisement on WMG notice boards. The PIS and consent form will be given before they attend this study to let the potential participates have enough time to consider. Participation in this study is completely voluntary and no pressure will be exerted on potential participants to take part. Up 30 subjects with normal hearing are to be selected from the following inclusion standards:

- Adult subjects aged between 18 years to 40 years, who are willing to participate in this study.
- Normal hearing subjects with pure thresholds (Pure tone audiometry threshold ≤20 dB HL).
- Have normal middle ear and external ear. Have no current ear problems (e.g. pain, ear infection, medication for ear problems et al).
- No complaints of suffering from tinnitus or sensitive to loud sounds.
- Have not been exposed to loud noise in the last 24 hours.
- Not regularly using known ototoxic drugs (e.g. aspirin, gentamicin, tobramycin, cisplatin and carboplatin et al).

All subjects will be recruited from staff and students of the University of Warwick, and each will attend a first pre-test session lasting approximately 2 hours. During this session, each subject will be given the consent form and have ample time to read it again before they sign it and feel free to ask questions about the nature of the study. A pure tone audiogram test will be carried out (by a trained audiologist – Miss Liping Zhang) to make sure the subject qualifies to participate in this study. After that, the instructions for the experimental tests to be carried out will be given to the subject to read and ensure they understand during the experiment. At this stage subjects will be put into either a control or test group. All subjects will be informed that they can withdraw from the study at any time.

Control group subjects will undergo a 'pre-test' experiment during this first session, and then a follow-up 'post-test' experiment around a week later, lasting approximately 2 hours. The Test group subjects will undergo both the pre- and post-test experiments, as well as 7 daily training sessions in between the two, each lasting 1 hour. All sessions will be carried out within a single-walled sound proofed room in the International Manufacturing Centre, WMG, University of Warwick.

Pre-test experimental session:
Each subject will be presented with a series of band-limited noises (at comfortable levels – see below) and asked in the SAM detection task to identify the sound which is amplitude modulated; and to detect the smallest change in modulation frequency for the SAM discrimination task. This will be repeated five times for each of the four conditions - one SAM-detection condition and three SAM-rate- discrimination conditions.

Training sessions:
During this project, a training session will be taken for SAM-detection condition between the pre- and post-session. It consists of 7 daily one hour sessions on consecutive days (except weekends). The listeners are required to complete 12 thresholds in each session. Below are details about the test procedures. Figure1 shows the flowchart for SAM-detection and SAM-rate – discrimination tests.

Post-test experimental session:
Repeat of the pre-test experimental session, with one SAM-detection condition and three SAM-rate- discrimination conditions being tested. Five thresholds will be obtained from each condition.

Experimental methods and calibration:
The experimental apparatus will be calibrated before each subject takes the experiment to make sure the sounds are less than the spectrum level (see note 4*) of 40dB SPL.

A three-interval/ alternative forced choice procedure (3IFC/3AFC) is used to determine the thresholds for SAM-detection and SAM-rate- discrimination conditions. The modulation depths and rate are varied and targeting 79.4% correct performance (Levitt, 1971).

For the SAM-detection test, the standard sound is un-modulated noise and the target sound is a 3-4 kHz band-pass carrier modulated at 80 Hz. In this test condition, modulation depth of standard sound is measured to determine the modulation detection threshold with an adaptive tracking procedure. There will be three intervals, which include two standard signals and one target sound, processed during the test. Noticeably, the target signal is presented randomly during the test procedures. The listener is instructed to decide which interval contains the target stimuli. The starting modulation depth (m) is 100% modulation and the modulation index in decibels is $20\text{Log}_{10}(m)$. The initial step size is 4dB and then reduces to 2dB after three test reversals. The mean of the last 10 reversals in the adaptive track will be calculated as the SAM-detection threshold.

For the SAM-rate-discrimination conditions test, a 3-4 kHz band-pass carrier modulated at 80 Hz with three depths (high-100%, mid-70% and low-40%) used as the standard sound and the target sound is the same carrier with a higher modulation rate. During this test, the modulation rate of target sound is measured to determine the modulation detection threshold by the 3IFC adaptive tracking procedure. Subjects will give a response about which interval is different from the other two. The initial rate difference between the standard and target stimulus is 15 Hz, then decreases to 3 Hz after the third interval and 1 Hz thereafter.

Note:

4* Spectrum level: The level of the part of a specified signal at a specified frequency that is contained within a specified frequency bandwidth, centered at the particular frequency. Noise over-exposure is the predominant risk in this study, and will be controlled exactly via software and hardware. The maximum stimulus levels used in this study will be governed by the Control of Noise at Work Regulations 2005 (http://www.hse.gov.uk/noise/regulations.htm) – which came into effect for all industry sectors in the UK on 6 April 2006. The aim of the Noise regulations is to ensure that workers' hearing is protected from excessive noise at their place of work, to ensure their hearing is not damaged either by loss of sensitivity or lead to tinnitus. In this study, we will ensure that each experiment session (only one per day maximum) is below the **lower exposure action value** as stipulated by the Noise regulations – i.e. limit daily personal noise exposure to below 80 dB (A-weighted) and ensure that no peak sound pressure should be above 135 dB (C-weighted). The Noise regulations actually allow exposure up to 87 dB (A-weighted) and a peak sound pressure of 140 dB (C-weighted). Therefore, by ensuring the experiments are below the much lower (recall that dB is a logarithmic scale) exposure action value ensures there is no chance of hearing damage. The stimulus levels will be calibrated using the industrial standard IEC 711 acoustic coupler and a precision microphone. Maximum sound pressure levels are then completely controlled via software (custom written for MATLAB) prior to presentation to the subjects.

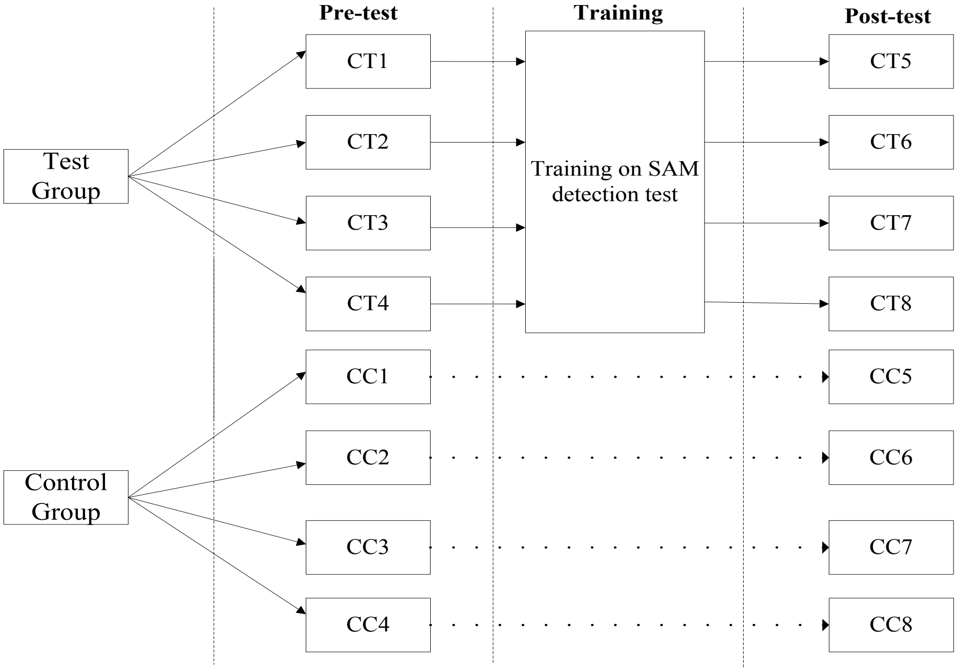


Fig. 1, The flow chart for the SAM-detection and SAM-rate – discrimination tests

CT1= Pre-test for SAM-Detection (Test group): standard sound is unmodulated noise and the target sound is a 3-4 kHz bandpass carrier modulated at 80 Hz

CT2= Pre-test for SAM-rate- Discrimination-1(test group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **100%** modulation depth

CT3= Pre-test for SAM-rate- Discrimination-2(Test group) standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **70%** modulation depth

CT4= Pre-test for SAM-rate- Discrimination-3(Test group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **40%** modulation depth

CT5= Post-test for SAM-Detection (Test Group): standard sound is unmodulated noise and the target sound is a 3-4 kHz bandpass carrier modulated at 80 Hz

CT6= Post-test for SAM-rate- Discrimination-1(Test Group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **100%** modulation depth
CT7= Post-test for SAM-rate- Discrimination-2(Test Group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **70%** modulation depth
CT8= Post-test for SAM-rate- Discrimination-3(Test Group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **40%** modulation depth
CC1= Pre-test for SAM-Detection (Control group): standard sound is unmodulated noise and the target sound is a 3-4 kHz bandpass carrier modulated at 80 Hz
CC2= Pre-test for SAM-rate- Discrimination-1(Control group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **100%** modulation depth
CC3= Pre-test for SAM-rate- Discrimination-2(Control group) standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **70%** modulation depth
CC4= Pre-test for SAM-rate- Discrimination-3(Control group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **40%** modulation depth
CC5= Post-test for SAM-Detection (Control Group): standard sound is unmodulated noise and the target sound is a 3-4 kHz bandpass carrier modulated at 80 Hz
CC6= Post-test for SAM-rate- Discrimination-1(Control Group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **100%** modulation depth
CC7= Post-test for SAM-rate- Discrimination-2(Control Group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **70%** modulation depth
CC8= Post-test for SAM-rate- Discrimination-3(Control Group): standard sound is a 3-4 kHz bandpass carrier that SAM at 80 Hz with **40%** modulation depth

Ethical considerations – include details of the Research Ethics Committee (REC) which will review the study/project and see end of the document for further guidance

Taking part in this study is entirely voluntary. It is up to the subject to decide whether or not to take part in this study. Before we conduct the main test, the participant will be trained to be familiar with the test procedures. We will make sure that they could understand the purposes of this study. The main test will only be carried out, once the subjects have given the completed consent form.

All information collected during the research will be kept strictly confidential. The subject's personal information will not be identified in any report/publication. The un-anonymised data will be stored only on the Chief and Principle investigator’s personal computers for 10 years. Only the primary research data, which cannot identify for individuals, will be published in research journals or conferences.

Participants are free to withdraw at any time they want even after they took part in the study. A decision to withdraw at any time, or a decision not to take part in this study, will not affect the standard of care they could receive.
Damage of subjects’ hearing will be extremely unlikely because the volume of the presented sound will be kept below participant’s maximum uncomfortable level. The risk of damage to hearing is also minimised according to the Control of Noise at Work Regulations 2005.

Participants will be paid £5 per hour for their time. The data collected from this experiment will make contribution to research.

Financing

This project is funded by the University of Warwick Research Development Fund (RDF1073).

Dissemination and Implementation

The results of this research study will be submitted to scientific journals/conferences. Participants will have access to such a publication if required

References

Fitzgerald, M. B. and Wright, B.A. (2010) Perceptual learning and generalization resulting from training on an auditory amplitude-modulation detection task. *Journal of the acoustical society of America.* **Vol.129**(2) p. 898 - 906

Hall, J. W. III, and Grose, J. H. (1994) Development of temporal resolution in children as measured by the temporal modulation transfer function, , *Journal of the Acoustical Society of America*. **Vol.96**, p.150–154.

Hawkey, D. J., Amitay, S., and Moore, D. R. (2004) Early and rapid perceptual learning, *nature neuroscience*. **Vol.7** (10), p.1055–1056.

Levitt, H. (1971). Transformed up-down procedures in psychoacoustics. *Journal of the Acoustical Society* of America, **vol.** 49,p. 467–477.

Plomp, R. (1983) The role of modulation in hearing. *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman (Springer-

Verlag, Berlin), p. 270–276.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory, and linguistic aspects, *Philosophical transactions of the royal society of London series a mathematical physical and engineering sciences*. **vol. 336**. no. 1278. p. 367-373.

Patterson, R. D., Johnson-Davies, D., and Milroy, R. (1978) Amplitude modulated noise: The detection of modulation versus the detection of modulation rate, *Journal of the Acoustical Society of America*. **Vol. 63**,p.1904–1911.

Wright, B. A., and Zhang, Y. (2009) A review of the generalization of auditory learning, *philosophical transactions of the royal society of London series a mathematical physical and engineering sciences*. **vol. 364**(1515) p. 301-311

Appendices (e.g. questionnaire(s), patient information leaflet(s), consent form(s), interview schedule(s), interview topic guide(s))

patient information leaflet, consent from, instructions

Ethical Considerations

Ethical Considerations can include any or all of the points described below, along with others. An ethics committee will expect to see evidence in the protocol that the applicant has given consideration to these issues, and designed the study so as to address these. These points should be addressed specifically in the '**Ethical Considerations'** section of the protocol, and in other relevant sections as appropriate, e.g. the Method section may also include a description of the informed consent process.

As a *minimum*, the section on Ethical Considerations should contain sub-sections examining **Informed Consent**, and **Participant Confidentiality and Data Security**.

Informed Consent

Describe the process you will use to ensure your participants are freely giving fully informed consent to participate. This will usually include the provision of an information sheet, and will normally require the completion of a consent form, unless it is a self-completion questionnaire based study, or there is justification for not doing (which must be clearly detailed).

Participant Confidentiality and Data Security

Provide details of the degree of anonymity of the data you will have access to. If the data you will access contains identifiable data, state what this data will be. If the data you will access has been anonymised, clarify how this has been done (bear in mind that combinations of demographic data can still identify individual participants from the original dataset, particularly for small sample sizes).

State how long study information (including research data, consent forms and administrative records) will be retained for. Also, state in what format(s) the information will be retained (for example, as physical and/or electronic copies), and state the specific physical location where the data will be stored (for example, where within the University of Warwick). Detail the security arrangements for the stored data, e.g. passwords on files and computers, and locked cabinets and offices for paper records.

Right of Withdrawal

Participants should be able to withdraw from the research process at any time. Participants also should be able to withdraw their data if it is identifiable as theirs, and should be told when this will no longer be possible (e.g. once it has been included in a final report or publication). Describe the exact arrangements for withdrawal from participation and withdrawal of data depending on your study design

Process for dealing with sensitive disclosures

If it is possible that criminal or other disclosures requiring action (e.g. evidence of professional misconduct) could be made during the study, the procedures that will be put in place to deal with these issues should be detailed. In certain circumstances there may be a need for disclosures to be communicated outside of the research team. The limits to confidentiality must be made clear to participants at the outset. The Participant Information Sheet should make it clear to potential participants under what circumstances action may be taken and what that may be.

Benefits and risks

Describe any expected benefits to the research participant, e.g. will participants receive a copy of the final report. Also, describe any possible risks to the research participant, e.g. what is the potential for adverse effects resulting from study participation. The potential for each of these should be identified and the protocol should state how you will minimise these risks and deal with any untoward incidents and adverse reactions.

Other Issues

Provide details of any other ethical issues or risks that may arise as a result of the dissemination of the research findings. For example, provide details if there are any anticipated limitations or restrictions on how the research findings might be disseminated or published (perhaps imposed by research funders, sponsors, or collaborating bodies). Outline the risks and how they will be minimised, if the dissemination of findings might present risks to the participants.

Further reading:

British Psychological Association guidance. Accessible at:
http://www.bris.ac.uk/Depts/DeafStudiesTeaching/dissert/BPS%20Ethical%20Guidelines.htm

ESRC Framework for Research Ethics (2010):
http://www.esrc.ac.uk/_images/Framework-for-Research-Ethics_tcm8-4586.pdf .

MRC Good research practice: Principles and guidelines (2012):
http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc002415.pdf

Appendix 4 Ethical approval letter

Tuesday 5th March 2013

Warwick
Medical School

PRIVATE
Liping Zhang, c/o James Harte
WMG
International Digital Lab
University of Warwick
Coventry
CV4 7AL

Dear Liping,

Study Title and BSREC Reference: *Generalization resulting from training on SAM-detection task to SAM-rate Discrimination task with different depths – REGO-2013-065*

Thank you for submitting your revisions to the above-named project to the University of Warwick Biomedical and Scientific Research Ethics Sub-Committee for Chair's Approval.

I am pleased to confirm that I am satisfied that you have met all of the conditions and your application meets the required standard, which means that full approval is granted and your study may commence.

I take this opportunity to wish you success with the study and to remind you any substantial amendments require approval from the committee before they can be made. Please keep a copy of the signed version of this letter with your study documentation.

Yours sincerely,

David Davies
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific
Research Ethics Subcommittee
Enquiries: Amy Ismay
B032 Medical School Building
Warwick Medical School,
Coventry, CV4 7AL.
Tel: 02476-151875
Email: A.C.Ismay@warwick.ac.uk

Appendix 5 VCV pilot studies

Both of the pilot studies included two test session, one is the practice session and the other is the training session. For the practice session, participants were required to do two blocks (one male voice block and one female voice block) of VCV stimuli without background noise test. Each VCV stimuli block contained 64 trials and 8 consonants. The eight constants were presented randomly for four times in each block in both the practice and training sessions. Followed up by the practice session, the training session conducted in three consecutive days. In this session, 10 blocks of VCV stimuli (five blocks of male voice and five female voice) were displayed and it token around 30 mins with 640 test trials per day. The VCV stimuli combined with same section fixed background noise on every single trial was tested per day for each person, but different sections of fixed babble noise were displayed for different listeners. Details about the parameters for the test stimuli were described in the coming part for each pilot study. The following Fig.1 showed the test flowchart for the VCV pilot study.

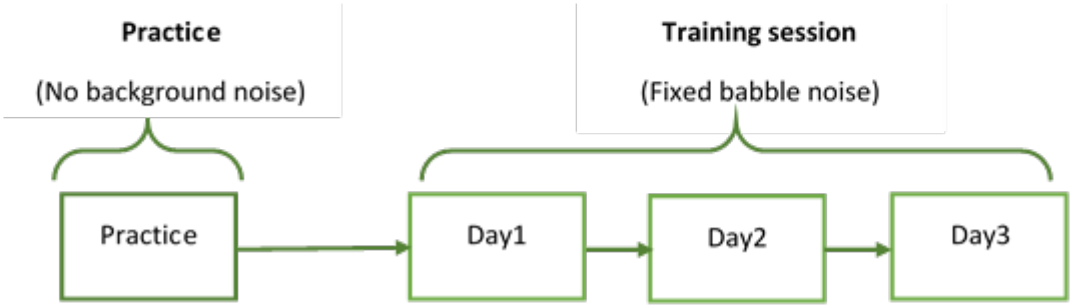


Fig.1 Flowchart for the VCV pilot study

1. Pilot study one

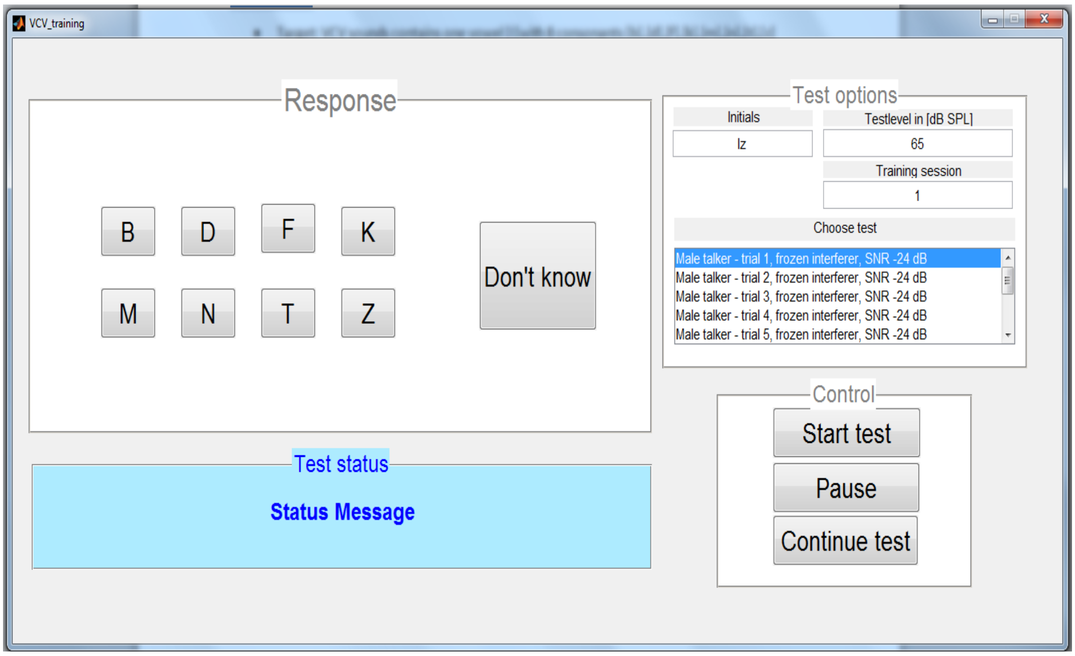
1.1 Parameters

1. **Signal to noise ratio:** SNR=-24dB
2. **Vowel:** /I/

3. **Consonants**: /b/,/d/,/f/,/k/,/m/,/n/,/t/,/z/
4. **Noise:** Speech shaped noise
5. **Sound out level:** 65dB SPL
6. **Participants:** Three people within normal hearing limits

1.2 Test interface

The following Fig.2 showed the test interface for the VCV pilot study one. As showed in Fig.2, there are nine choices on the figure. Participants were instructed to click one of the eight consonants that they heard during the test. For example, if they heard /IBI/, then they need to click the /B/choice on the screen. If they were struggling to detect the target stimuli form the background noise, then click the "Don't Know" choice. The "test status" showed how many trials left for each test block. The "test option" panel were mainly for entering the test parameters, listener's ID information and choosing test blocks (male and female voice blocks displayed here). The buttons on the control session "Pause" or "Continue Test" were used to let the listeners to help themselves to have a rest if they felt tired during the test.



Test options

Initials: lz

Testlevel in [dB SPL]: 65

Training session: 1

Choose test:
Male talker - trial 1, frozen interferer, SNR -24 dB
Male talker - trial 2, frozen interferer, SNR -24 dB
Male talker - trial 3, frozen interferer, SNR -24 dB
Male talker - trial 4, frozen interferer, SNR -24 dB
Male talker - trial 5, frozen interferer, SNR -24 dB

Test status: Status Message

Fig. 2 Test interface for the VCV pilot study one

1.3 Results calculation

Accuracy of VCV stimuli with speech shaped background noise was recorded by the correct responses (green colour) for each eight consonant. Table 1 showed the raw correct responses (yellow colour) data from participant 1 with one of the male voice block and Table 2 revealed the final percentage of correction across all the eight consonants /b/,/d/,/f/,/k/,/m/,/n/,/t/,/z/. As you can see, the horizontal line of the table showed the nine response choices, while the vertical line showed the eight consonants. The more correct responses from the participants, the larger of the data will be in the diagonal of these tables.

Table 1. The matrix of responses results for the test

Reponses Stimuli	b	d	f	k	m	n	t	z	No idea
b	0	0	0	0	0	0	0	0	8
d	0	5	0	0	0	0	3	0	0
f	0	0	2	0	0	0	0	0	6
k	0	0	0	0	0	0	1	1	6
m	0	0	0	0	0	0	0	0	8
n	0	0	0	0	0	0	0	0	8
t	0	0	0	0	0	0	8	0	0
z	0	0	0	0	0	0	0	8	0

Table 2. The final correction for each of the consonant

Reponses Stimuli	b	d	f	k	m	n	t	z	No idea
b	0	0	0	0	0	0	0	0	100
d	0	62.5	0	0	0	0	37.5	0	0
f	0	0	25	0	0	0	0	0	75
k	0	0	0	0	0	0	12.5	12.5	75
m	0	0	0	0	0	0	0	0	100
n	0	0	0	0	0	0	0	0	100
t	0	0	0	0	0	0	100	0	0
z	0	0	0	0	0	0	0	100	0

1.4 Results

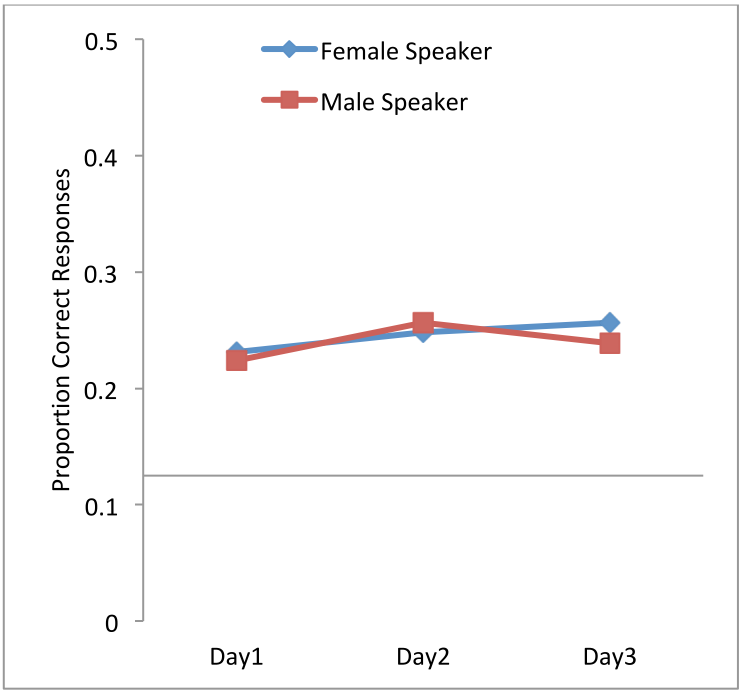


Fig. 3. Proportion of correct responses as a function of training (day 1 to 3), averaged across stimulus types and plotted separately for stimuli produced by a female (blue line) and a male (red line) speaker. X-axis crosses at chance performance level (1/8 = 0.125).

As shown in Fig. 3, based from the pilot study 1, overall it did not seem to make much difference if the talker was male or female. Listeners' average score (across three days, three participants and eight consonants) for male talker were 23.96%, and for female talker was 24.51%. It showed that there was not much difference from the female and male speaker. Regarding the overall performance from day1 to day 3, there was little improvement observed (Male voice improvement: 1.46%; Female voice improvement: 2.5%). As the listeners were not all native speakers, it might affect the day-to-day learning performance. In addition, three participants (at least one of whom is not a native English speaker) are not enough to discard the present paradigm. So probably, the results would be better if native English speakers were chosen for this study. Details about listeners' VCV identification performance in the training session can be seen in Table 3 (female talker) and Table 4 (male talker).

Table3: Proportion of correct responses (in percentage: %) as a function of training (day 1 to 3) for each target consonant (female speakers).

Female	b	d	f	k	m	n	t	z	AVE
Day1	1.67	25.83	6.67	0.00	0.00	8.33	75.00	67.50	23.13
Day2	2.50	31.67	8.33	2.50	6.67	12.50	66.67	67.50	24.79
Day3	0.83	50.00	6.67	2.50	3.33	4.17	66.67	70.83	25.63

Table 4: Proportion of correct responses (in percentage: %) as a function of training (day 1 to 3) for each target consonant (male speaker).

Male	b	d	f	k	m	n	t	z	AVE
Day1	10.83	8.33	7.50	3.33	3.33	5.00	80.83	60.00	22.40
Day2	9.17	14.17	16.67	4.17	5.00	11.67	69.17	75.00	25.63
Day3	4.17	16.67	13.33	3.33	8.33	4.17	66.67	74.17	23.85

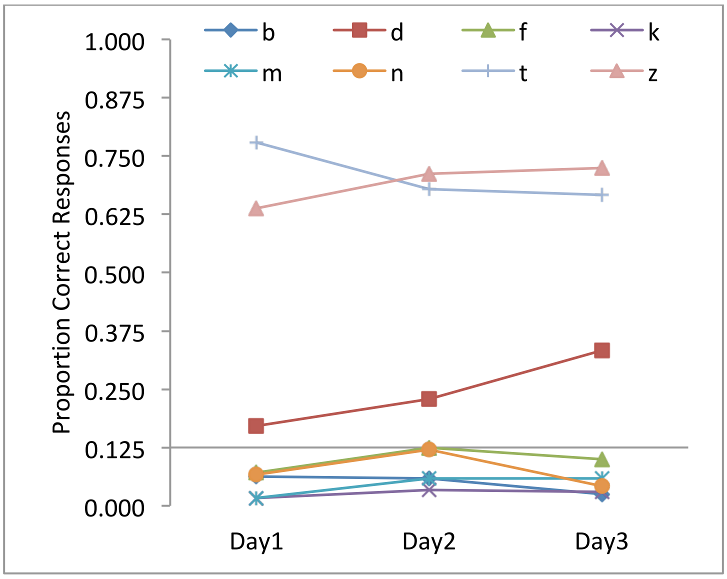


Fig. 4. Proportion of correct responses as a function of training (day 1 to 3), plotted separately for each target consonant (averaged across speakers). X-axis crosses at chance performance level (1/8 = 0.125).

According to the proportion of correct responses from listener's day to day training for each target consonant (/b/, /d/, /f/, /k/, /m/, /n/, /t/, /z/) in Fig. 4, listeners showed immediate improvement for most of the consonants from day one to day two. The consonants /d/ did in fact show very nice perceptual learning across three days, and this was exactly the consonant that started out slightly above chance. However, /t/ and /z/ started out nearby ceiling performance (/t/: 77.92%; /z/: 63.75%) and stayed there (/t/: 66.67%; /z/: 72.50%), all others start out at (or actually below) chance (12.5%), and never improve. Details about the participants' performance for each target could be seen from the following table 5.

Table 5: Proportion of correct responses (in percentage: %) as a function of training (day 1 to 3) for each target consonant (averaged across speakers).

	b	d	f	k	m	n	t	z
Day1	6.25	17.08	7.08	1.67	1.67	6.67	77.92	63.75
Day2	5.83	22.91	12.50	3.33	5.83	12.087	67.92	71.25
Day3	2.50	33.33	10	2.92	5.83	4.17	66.67	72.50

1.5 Short summary for the pilot study one

- **SNR level:** The initial correct responses for most of the eight consonants were below or around guess chance level, but they did not show any improvement across the training session. Perhaps the VCV stimuli in noise intelligibilities in this pilot study are too low to let listeners to obtain any improvement. Therefore, it is suggested that the identification of others consonants will be increased if the signal to noise ratio less difficult.

- **Target vowel:** Although the vowel /i/, which was used in this pilot study, was the best carrier vowel than the other vowels (*/a/*, */e/*, */i/*, */o/*, */u/*). Gutnick (1982) showed that the vowel /a/ is much easier to be detected in noisy background than the vowel /i/, and it may give a better clue to the timing of the following consonant. Probably, it is better to do another pilot study and use the vowel /a/ as the target vowel to explore for better perceptual learning performance.

- **Target consonants:** The consonants /t/ and /z/ were easier to reach their asymptotic identification performance, so the learning effect for these two consonants was harder to explore than the other consonants. Maybe it is better to replace these two consonants with another two harder ones. In literature, Phatak and Allen (2007) indicated that the identification of consonants within speech noise are separated in to three sets: high score consonants: /t/, /s/, /z/, /ʃ/, /ʒ/; intermediate ones: /n/, /p/, /g/, /k/, /d/; low score ones: /f/, /θ/, /v/, /ð/, /b/, /m/). In order to explore more about the auditory perceptual learning effects of vowel consonant vowel in speech noise, we should pay more attention about the identification intelligibility of consonant stimuli with background noise. When choosing the right target consonants for this auditory training study, the basic requirement probably is the initial identification of target consonants were supposed to be as low as its guess level. Based on this rule, maybe it is better to delete the high-scored consonants /t/ and /z/ and add another two intermediate ones, such as /g/, /p/.

In sum, the following questions and actions needed to consider after before we carried on the next pilot study:

1. Whether to combine the male and female voice together to continue the test or just choose one of them?

    Answer: Yes. Mixed the male and female voice together.

2. Whether to try to do another pilot study with SNR-24dB?

Answer: No. For the next study, the SNR should be changed to be easier than -24dB, it is suggested to be -18dB.

3. Regarding the consonants, whether to change the easier ones to be some lower score consonants?

Answer: Yes. Change consonats /t/and/z/ to be /g/and /p/

2. Pilot study two

2.1 Parameters, tests interface and results calculation

Following the VCV pilot study 1, two native speakers were recruited to participate for the VCV pilot study two. Several changes were made for the parameters of VCV stimuli from the pilot study one. Expect the changes about consonants, background noise type, SNR level and vowel used in the VCV pilot study two was different from VCV pilot study one. The test interface, test flowchart, and results calculation ways were same as the method used in the VCV pilot study one. Details of the changes were listed in the following part.

1. **Vowel**: /a/
2. **SNR**: -18dB
3. **Consonants**: /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/ (Compared with last pilot study, I removed /t/and/z/, because they are easily to obtain 100% correction. They were changed to /g/ and/p/.)

4. **Noise**: Babble noise. (As babble noise is much more real life like, it sounds like many people talking around)

2.2. Results

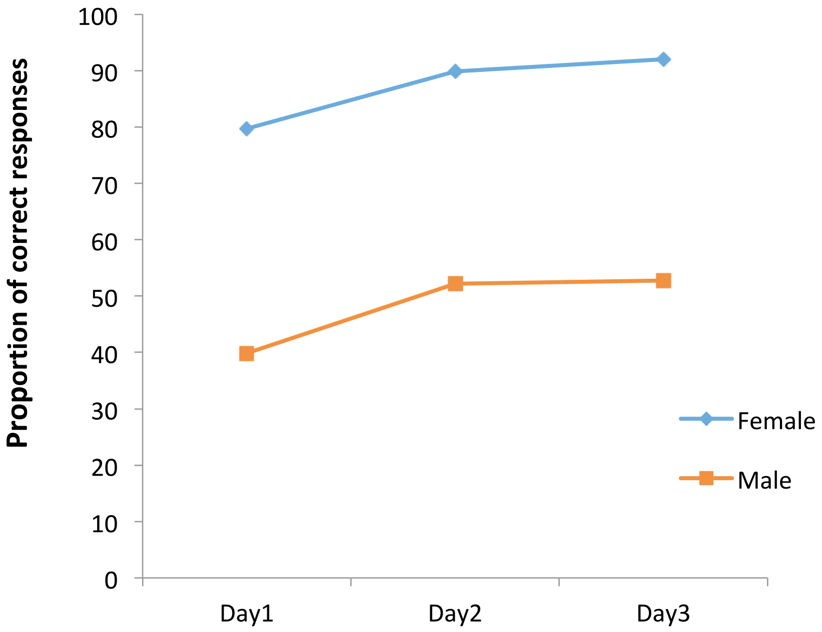


Fig. 5 Proportion of correct responses as a function of training (day 1 to 3), averaged across stimulus types and plotted separately for stimuli produced by a female (blue line) and a male (red line) speaker.

As shown in Fig. 5, based from the overall listeners' performance of the pilot study two, it seemed that the female talker leaded to better performance than the male talker one. Listeners' average score (across three days, three participants and eight consonants) for male talker were 48.28%, and for female talker was 87.19%. It showed that listeners accuracy of correct responses from the female talker is almost 38.91% higher that it from the male speaker. Regarding the overall performance from day1 to day 3, listeners' auditory perceptual performance improved for both the male and female voice (Male voice improvement: 12.97%; Female voice improvement: 12.34%), the improvement for the male talker was slightly more than the improvement for the female talker one. Details about the listeners' day one to day three performance across each consonants could be seen from table 6 (female speaker) and table 7 (male speaker). According to the following table 6 and 7, listeners could achieve 100% correction for their day one performance for consonants /d/ and/f/ with female talker speakers. It might prevent the learning effect for female speaker with consonants /d/ and /f/ and leaded to lower improvement for female talker's performance. This kind of ceiling effects should be avoided for perceptual learning design.

Table 6: Proportion of correct responses (in percentage: %) as a function of training (day 1 to 3) for each target consonant (female speaker).

Female	b	d	f	g	k	m	n	p	AVE
Day1	85.00	100.00	100.00	88.75	98.75	16.25	77.50	71.25	79.69
Day2	83.75	100.00	100.00	93.75	100.00	53.75	95.00	92.50	89.84
Day3	85.00	100.00	100.00	97.50	100.00	56.25	100.00	97.50	92.03

Table 7: Proportion of correct responses (in percentage: %) as a function of training (day 1 to 3) for each target consonant (male speaker).

Male	b	d	f	g	k	m	n	p	AVE
Day1	20.00	62.50	52.50	86.25	68.75	2.50	6.25	20.00	39.84
Day2	31.25	67.50	62.50	96.25	77.50	21.25	30.00	31.25	52.19
Day3	32.50	57.50	52.50	93.75	65.00	36.25	46.25	38.75	52.81

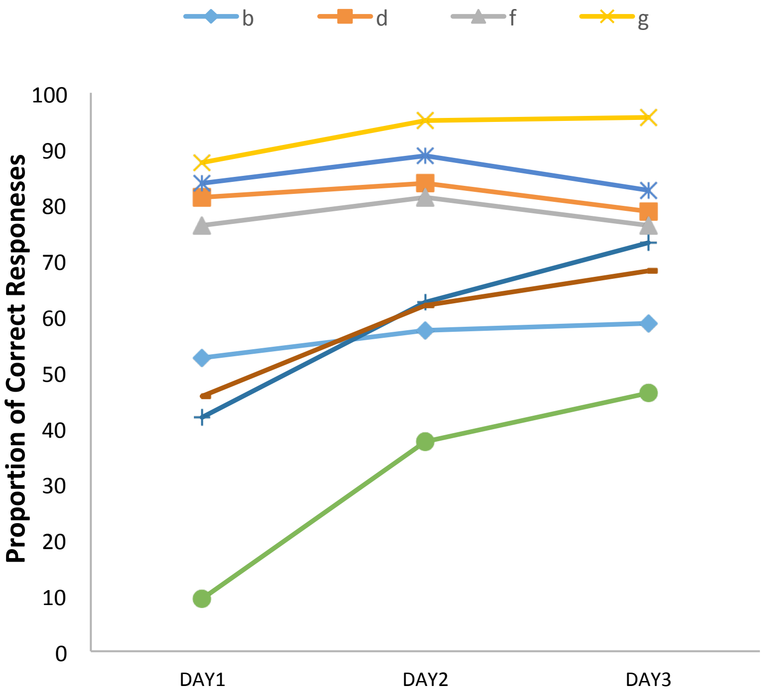


Fig. 6 Proportion of correct responses as a function of training (day 1 to 3), plotted separately for each target consonant (averaged across speakers).

According to the proportion of correct responses from listener's day to day training for each target consonant (/b/, /d/, /f/, /k/, /m/, /n/, /g/, /p/) in Fig. 6, most of the consonants showed immediate improvement from day one to day two. Regarding the performance from day one to day two for these eight consonants, the start point of the consonant /m/ was the lowest one (9.38%), while the initial performance for consonant /g/ was the highest one (87.5%). However, the consonants /m/ showed the largest improvement from day one to day two (almost 28.12%), the improvement for the consonant /g/ was only 7.5%. It suggested that the higher the start point of the accuracy of VCV stimuli was, the less of the improvement range was. Probably the consonant /g/ started out nearby its asymptotic performance. In this pilot study, except the consonant /m/

started below the guess level (12.5%), all others start out at (or actually) above 40% chance. Therefore, the VCV stimuli identification intelligibility for this test was probably too high to get the expected purpose.

Table 8: Proportion of correct responses (in percentage: %) as a function of training (day 1 to 3) for each target consonant (averaged across speakers).

	b	d	f	g	k	m	n	p
Day1	52.50	81.25	76.25	87.50	83.75	9.38	41.88	45.63
Day2	57.50	83.75	81.25	95.00	88.75	37.50	62.50	61.88
Day3	58.75	78.75	76.25	95.63	82.50	46.25	73.13	68.13

2.3 Short summary for pilot study two

According to listeners' performance results in this VCV pilot study two, even the average proportion (across consonants) correct responses of female talker was higher than the male voice one, the improvement were almost similar for these two speakers (Male speaker: 12.97%; Female speaker: 12.34%). Both of the two participants obtained 100% responses correction for the consonants /d/ and /f/ after they finished the day one female voice VCV in babble noise test. It showed that it was easier for listeners to achieve their ceiling performance with consonants /d/ and /f/, the VCV stimuli identification rate was probably too high to do a proper auditory perceptual training study. There were two reasons to lead to listeners ceiling performance, one is the target consonants ware too easy to be detected; the other one is the signal to noise ratio used in this pilot study is too high to make listeners to obtain their ceiling performance. As the consonants were all lower or intermediate score for speech in noise test, the signal to noise ratio was suggested to be reduced for the main VCV learning experiment.

Regarding the learning improvement, although the consonants /m/ belonged to be one of the low-scored speech in noise consonants, the consonant/m/obtained the largest improvement after 3 days' VCV fixed babble noise training. It indicated that the harder tasks might lead to better perceptual learning effect.

3. Outcomes and further plan

Although the participants felt the male voice one is easier than the female voice to detect, the proportion correct response of female voice is higher than the male's. Because the consonants /t/and/z/ are easily to obtain 100% correction in pilot study one, they were changed to the other two consonants /g/ and/p/. Regarding the background noise, babble noise is more real life like than speech shaped noise. It is suggested to keep using babble background noise in the main test for VCV perceptual learning study. Overall, learning effect obtained after three days fixed babble noise/speech shaped noise training from both SNR-18/SNR-24dB. For further study, it is better to choose the SNR=-24dB babble noise, training with (including eight consonants /b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/) and using both the male and female speakers.

Appendix 6 VCV study data

Table 1 The proportions of correct responses, incorrect responses, and don't know responses for each condition at each time point for VCV SNR -24 dB.

SNR-24		Pre	Day1	Day2	Day3	Post
Fixed (n=10)	Correct	40.31	44.70	46.34	50.38	51.25
	Don't know	21.72	5.72	4.31	1.86	9.14
	Incorrect	37.97	49.58	49.35	47.76	39.61
Random(n=10)	Correct	31.80	43.69	45.92	49.63	50.63
	Don't know	32.27	14.86	12.88	9.34	7.66
	Incorrect	35.93	41.45	41.20	41.03	41.71

Table 2 The proportions of correct responses, incorrect responses, and don't know responses for each condition at each time point for VCV SNR -30 dB.

SNR-30		Pre	Day1	Day2	Day3	Post
Fixed (n=10)	Correct	10.47	14.52	15.20	17.34	18.44
	Don't know	50.94	21.94	10.42	7.06	28.05
	Incorrect	38.59	63.55	74.38	75.59	53.52
Random (n=10)	Correct	11.88	19.00	24.11	25.64	30.23
Random (n= 9)	Correct	11.46	19.65	24.45	26.29	31.08
	Don't know	57.12	33.52	20.38	16.61	10.42
	Incorrect	31.42	46.83	55.17	57.1	58.5
Control (n=10)	Correct	11.80				16.88
	Don't know	29.85				21.95
	Incorrect	58.35				61.17

Table 3 The proportions of correct responses, incorrect responses, and don't know responses for each condition at each time block for short session VCV study with SNR -30 dB.

SNR-30		Block1	Block2	Block3	Block4	Block5
Fixed (n=12)	Correct	8.98	13.02	18.10	20.96	23.44
	Don't know	41.28	16.02	12.76	13.67	12.11
	Incorrect	49.74	70.96	69.14	65.36	64.45
Random (n=13)	Correct	9.62	14.42	16.71	20.43	24.88
	Don't know	57.57	50.24	43.39	42.07	36.66
	Incorrect	32.81	35.34	39.90	37.50	38.46