Transcribing nonsense words: The effect of numbers of repetitions and voices

Rachael-Anne Knight
Department of Language and Communication Science
City University London
Northampton Square
London
EC1V 0HB

Email: knight@city.ac.uk
Tel: 020 7040 8081
Fax: 020 7040 8577

Abstract

Transcription skills are crucially important to all phoneticians, and particularly for speech and language therapists who may use transcriptions to make decisions about diagnosis and intervention. Whilst interest in factors affecting transcription accuracy is increasing, there are still a number of issues that are yet to be investigated. The present paper considers how the number and type of voices, and the number of repetitions affects the transcription of nonsense words. Thirty two students in their second year of study for a BSc in Speech and Language Therapy were participants in an experiment. They heard two nonsense words presented ten times in either one or two voices. Results show that the number and gender of voices did not affect accuracy, but that accuracy increased between six and ten repetitions. Implications for teaching and learning, clinical practice, and further research are discussed.

Introduction

Phonetic transcription is, of course, the method by which sounds of speech can be recorded using symbols from the International Phonetics Alphabet (IPA) and its extensions (ExtIPA). As such, transcription is a crucial tool to all who use it, as it fixes the ephemeral speech signal in printed form and therefore allows a transient signal to be returned to at a later time.

It is important for clinicians and researchers that phonetic transcriptions are accurate and reliable. For clinicians, transcriptions will inform decisions about diagnosis and intervention. Howard and Heselwood (2002:373) comment on the 'status of the transcription as a necessary first step in assessment' and stress 'the importance of a high degree of accuracy and reliability'. For researchers, there is a similar importance to be placed on accuracy, as different theories and hypotheses might be evaluated against transcriptions. Because transcriptions are the basis for important, and potentially far-reaching, decisions in clinical practice and research, it is important for us to understand which factors might affect transcription accuracy.

Another important reason for needing to know how accuracy can be affected is for the purpose of enhanced teaching practices. Trainee transcribers must develop transcription skills to a level that is advanced enough to be useful in their future profession. However, in order for lecturers to teach students these skills, we need to know what factors might affect transcription accuracy. Furthermore, increased knowledge about the factors affecting accuracy might help lecturers allay some of the fears about what is perceived to be a difficult subject (Noble, submitted), develop strategies to make the process easier and more successful, and ensure that our assessment practices are fair, and comparable across institutions.

In order to fully appreciate the factors that might affect transcription accuracy we can consider the stages in the transcription process. Firstly the signal must be heard (and seen) and attended to. We can assume, suing recent models of working memory, that a representation of the signal is then stored in the phonological loop (Baddeley 1986), which is considered to be a subcomponent of working memory. Baddely, Gathercole and Papagno (1998: 158) describe the phonological loop as 'specialised for the retention of verbal information over short periods of time; it comprises both a phonological store … and a rehearsal process which serves to maintain decaying representations'. Whilst the item to be transcribed is stored in the phonological loop the transcriber must divide the signal into segments and analyse each segment by comparing the signal to representations stored in long term memory (although the form which these representations take is still at issue). Information about each segment (such as voice, place manner labels) will be accessed, as must the appropriate symbols which must be written down.

Differences in accuracy between transcribers can, therefore, be due to differences occurring at any stage of the transcription process. An obvious example is that a person who cannot hear the signal, (due to a physical barrier or hearing impairment) will not be able to move on to the later stages of the process unless other methods are found to bypass this initial stage. Storage and processing in working memory could also affect success as a more limited phonological loop capacity might mean that longer or more complex items cannot be remembered in detail (Maguire and Knight, in preparation). Segment identification might be affected by a variety of factors. For example, the degree to which symbols and sounds have been practiced will affect how they are stored in long term memory. In addition, abilities not directly related to transcription may also affect identification. Mackenzie Beck (2003: 2833) administered two types of aptitude tests to beginning speech and language therapy students and found that the musical aptitude test was 'at least as effective as the phonetics test in predicting later phonetic ability'. Finally the transcriber's ability to write symbols clearly will also affect the perceived accuracy of a transcription when it is examined by someone else.

Transcription accuracy may also be affected by factors concerning the item to be transcribed, and different types of material are likely to be transcribed more accurately than others. Clinical phrases, where a pseudoclinical version of a known English phrase is produced for transcription, may be difficult to transcribe as the phonological representation in the lexicon interferes with the analysis of the incoming signal. Oller and Eilers (1975) indicate that knowing the identity of an item to be transcribed can improve transcriptions, but can also lead to inclusion of features which are not in fact present but might be expected to occur in the item. The duration of the item to be transcribed may also be a factor affecting accuracy. In transcribing a passage phonemically, for example, the amount of material to be attended to will certainly exceed the capacity of the phonological loop and therefore more repetitions will be required for an accurate transcription.

The nature of the presentation is also likely to affect transcription accuracy. For example, an item presented with visual as well as auditory information will likely be easier to transcribe than one presented only in the auditory modality. Adding visual elements to the signal increases the signal-to-noise ratio significantly, at least in a noisy environment, (e.g. Sumby and Pollack, 1954, Middleweerd and Plomp, 1987) and gives some information about place of articulation and lip rounding. For this reason teachers usually encourage students to look at the face of the person giving a dictation, and to look at the face of clients in clinic or participants in research (although we do not currently have any data on the degree of improvement that might be expected when visual information is available in a transcription exercise).

The current paper deals with two additional aspects of presentation and their putative effects on the accuracy of transcription of nonsense words. The sections below summarise what is known, and what we might hypothesise, about the effect of the number of repetitions of a nonsense word, and the number and type of voices in which it is presented, on transcription accuracy.

Possible effects of repetitions

There is, to the author's knowledge, no empirical test of how the number of presentations of an item affects transcription accuracy. The ideal scenario in real-life settings is for as much information as possible to be gathered after one repetition of an item. Audio recordings are not always possible in clinic, and for both research and clinical purposes successive repetitions by an individual will not always be identical, especially when the individual is very young or has disordered speech (e.g. Holm, Crosbie and Dodd, 2007). Despite the ideal situation of transcribing accurately after one repetition, it is nevertheless the case that most educational institutions present far more than one repetition when they are training and examining students.

Although there is no previous research on how the number of repetitions affects transcriber accuracy, Amarosa, von Benda, Wagner, and Keck (1985) investigated how intertranscriber <u>reliability</u> was affected by single and multiple repetitions. Two phoneticians made live transcriptions of words spoken once by two children. At a later time the phoneticians retranscribed the session using recordings to which they could listen as many times as they felt necessary. On an item-by-item analysis, agreement between the transcribers rose by 12 and 16 percentage points for child 1 and 2 respectively (1985:283). Although accuracy was not addressed in the study, a qualitative analysis of the transcriptions made from multiple repetitions showed that they were less similar to the target than the live transcriptions. Amarosa et al. (1985: 285) suggest that the transcribers normalised less in this condition and argue that, if we are to produce transcriptions which are useful in establishing a child's phonological system, we must use groups of transcribers and allow them to listen to high quality audio tapes as often as is needed.

In a similar experiment Munson and Brinkman (2004) compared inter- and intra-rater reliability for judgements about the speech of children with phonological disorders. Listeners were asked either to transcribe a word or make a simple binary choice about whether it had been produced correctly or incorrectly, and items were heard either once or seven times. Whilst there was no difference in either inter- or intra-rater reliability for phonetic transcription, multiple presentations induced a small increase in inter-rater reliability for accuracy judgements.

So, there is some evidence that multiple repetitions may increase reliability and there is an intuitive appeal to the idea that more listenings will lead to greater accuracy in transcription. However, as Munson and Brinkman (2004: 342) point out there is a great deal of evidence in the literature that multiple repetitions of items can in fact lead to auditory illusions. In particular, the much replicated verbal transformation effect (first reported by Warren and Gregory, 1958) indicates that lexical tokens appear to change with multiple hearings. For example, Shoaf and Pitt (2002: 795) describe how listeners presented with multiple repetitions of the word 'truce' might report hearing transformations such as 'truth' as well as phonologically dissimilar words such as 'Esther', and even nonwords.

The verbal transformation effect also appears to be present to some degree when nonwords (both phonotactically legal and illegal) are presented. Warren (1961) indicates that a nonword gave rise to more transformations than four real words in all the age groups he tested. Likewise Natsoulous (1965) found both more transformations, and a shorter time for the first transformation to occur, in nonwords than real words.

From a teaching and learning perspective Wells (personal communication) suggests that transcriptions of nonsense words tend not to improve after about six repetitions, and may even get worse as the transcriber changes previously correct answers. This observation would seem to be consistent with the verbal transformation effect, although it has not been examined experimentally. From the literature then it appears that more repetitions may improve accuracy in the same way as they improve reliability, or that they may decrease accuracy in a similar way to the verbal transformation effect.

The experiment below examines the accuracy of student transcriptions after six and ten repetitions of a nonsense word.

Possible effect of number of voices

The number of voices in which an item for transcription is presented may also affect accuracy. Research suggests that normal speech processing is affected by whether words are spoken in one or multiple voices. Mullennix, Pisoni and Martin (1989), for example showed that spoken word recognition is impaired by hearing multiple voices. They found that participants were less accurate at recognising words, and were slower at a naming task, when they listened to lists of words when the speaker changed from trial to trial than when the speaker stayed the same.

Goldinger (1998) found similar results, in that words are recognised more quickly, and responded to more quickly in a lexical decision task if they have previously been presented in the same voice. In addition, participants in shadowing tasks are more likely to imitate a voice if they have heard it consistently than if different voices have been presented. Thus, hearing multiple voices seems to impair processing and hearing the same voice appears to improve processing.

There are two possible ways in which same voice effects might arise in normal speech processing. Goldinger, Pisoni and Logan (1991: 153) indicate that speech from multiple talkers is harder to process because listeners must compensate for different voices, which in turn diverts resources from the task of word recognition. In particular, Martin, Mullenix, Pisoni and Summers (1989) indicated that variability may affect encoding and rehearsal processes linked to working memory. In essence, they suggest that hearing different voices adds variability to the information which must be held in the phonological loop. Variability means that fewer words can be held in the loop, or that they are represented less accurately. If it is the case that different voices do indeed disrupt word recognition by adding variability to the phonological loop, then it is possible that multiple voices will a disrupt transcription of nonsense words as this process also relies on encoding and rehearsing items using the phonological loop.

Much of the later literature in the area (e.g. Goldinger 1998) suggests that same speaker effects are due to the long term storage and representation of words, rather than to their initial processing. Episodic theories of speech perception suggest that the mental lexicon is composed of multiple examples of each word. These examples, or exemplars, contain surface detail including information about voices, rather than abstracting away from this detail. When a person engages in word recognition the incoming signal is compared to these multiple examples. Words will be recognised more quickly the more similar they are to stored traces. As words heard again in the same voice will be very similar to stored traces they will be recognised more quickly than words repeated in a different voice (Jesse, McQueen and Page, 2007).

Whether or not the transcription of nonsense words is affected by multiple voices will depend on how information about non-native phonetic segments is stored in the brain. Nonsense words will not have been heard before and will not have a representation in the mental lexicon. However, individual segments must be represented in some way for recognition to occur and, like real words, are perhaps also represented by multiple exemplars. Recent findings in normal speech perception indicate that units smaller than the word are also subject to the effects of multiple talkers. Jesse McQueen and Page (2007) presented items for auditory lexical decision in the same or different voices. The items were either 1) repetitions of entire old words, 2) new words spoken in the same voice and containing the same phonemes as old words, or 3) new words containing previously unheard phonemes. Whilst repetitions of entire words were responded to most quickly, there was also some effect of hearing the same phonemes presented in a different order, indicating that there is an advantage for items repeated in the same voice even at a sublexical level.

Furthermore, Smith (2007: 1920) has shown that perceptual learning about voices affects items not only smaller than the word but also less abstract than phonemes. Hard-to-segment sequences that are phonemically identical but vary according to the placement of word boundaries (e.g. ‘pat sawed’ and ‘pat’s awed’) were played in noise before and after training with a voice. Subjects wrote down what they heard and improvement between the two testing phases was examined. Subjects who heard the same voice in training as in testing were shown to improve more than subjects who heard a different voice in the training phase. This was the case even though different tokens were used in the training and test phases. Smith's findings suggest, therefore, that the same-voice effect is not limited to words or even to phonemes.

Smith's and Jesse et al's work suggests that units smaller than the word are subject to same voice effects, and that these units are rather concrete. If sounds learnt in phonetic training are stored in a similar fashion to real words, then the number of voices in which a word is presented may have an effect on transcription accuracy. Sounds presented in a familiar voice will presumably activate very similar exemplars of that sound and thus be recognised quickly and accurately. Sounds presented in an unfamiliar voice will have to be matched to the closest exemplars available which will give some scope for error. Sounds may therefore be better identified if they are spoken in a similar voice to one for which exemplars are stored.

Therefore, it seems from the episodic speech perception literature that that the voices in which an item for transcription is presented may affect transcription accuracy. If multiple voices disrupt encoding and rehearsal processes in short term memory then a decrease in transcription accuracy may be found when items are presented in more than one voice. If effects arise from the storage of multiple exemplars of the sounds present in nonsense words, accuracy will depend on the degree to which the voices are similar to those voices for which a transcriber has stored exemplars.

From a teaching and learning perspective the number and type of voices in which an item is presented is also important. Students are often worried about how they will manage if they are examined by someone other than their lecturer, and teachers must be concerned about how best to encourage the development of robust mental representations for non-native speech sounds. The current work begins to investigate these issues by looking at if and how the number and type of voices in which a nonsense word is produced has an effect on transcription accuracy.

Overview

In summary then, the current paper aims to establish some initial conclusions about the effect of numbers of repetitions and the number and type of voices on transcription accuracy for nonsense words. These factors are assessed empirically by presenting students with nonsense words in one or two unfamiliar voices, one male and one female, and considering their accuracy after six and ten repetitions.

It is hypothesised that an increased number of repetitions may improve accuracy in the same way as it improves reliability, or that it may decrease accuracy due to verbal transformation effects. It is hypothesised that, if IPA sounds are represented in the same way as sounds in a subject's native language, then subjects hearing more repetitions in a female voice will do better overall because they have been trained solely by a female phonetician. Likewise, it is hypothesised that hearing repetitions in different voices might decrease accuracy because the transcriber has to handle different voices in working memory and compare to different tokens in episodic memory.

Method

The assessment of transcription accuracy lends itself rather well to an experimental approach, because, at least to some extent there are right and wrong answers and items can be varied in terms of content and presentation. In what follows students were asked to transcribe nonsense words presented ten times in either one or two unfamiliar voices, and their responses after six and ten repetitions were compared.

Participants

Participants were 32 second year undergraduates undertaking a course in Speech and Language Therapy. Students were all female, broadly reflecting the gender balance of the department and the profession, and were of mixed ability. All students had studied basic phonetics and phonology in their first year, and in their second year continue to work on ear training and clinical phonetics. Students were divided quasi-randomly into two equal groups of 16, depending on which side of the room they were seated in class when the experiment took place.

Stimuli

Two phoneticians (one male and one female) unknown to the students each recorded ten repetitions of two nonsense words. The phoneticians are of roughly the same age, and were trained at the same institution. The nonsense words were six segments in length and the sounds were selected from a set that the students were preparing for an upcoming class test. In this way the experiment also acted as revision for the test so that students felt motivated and appreciated that the exercise was useful for them as well as for the purposes of research. The set of sounds to be practiced for the test consisted of 30 consonants from the IPA including sounds produced on non-pulmonic airstreams, all the primary cardinal vowels, and secondary cardinals 1,2, 3 and 8. The nonsense words presented in the experiment were:

a. [ŋǁ ɔ dn u ʁ ɑ] and

b. [œ lˠ u ʌ a ts’].

Presentation

Nonsense words were presented over loudspeakers to each group separately. Group 1 heard word a first followed by word b, whereas group 2 heard word b followed by word a. Each group heard their first word presented in two voices (6 repetitions in one voice followed by 4 repetitions in another) and their second word presented in one voice. So, group 1 heard six repetitions of word b by the male, followed by four repetitions by the female. They then heard ten repetitions of word a by the female. Group 2 heard six repetitions of word a by the female followed by four repetitions from the male. They then heard 10 repetitions of word b by the male. There were 2.5 seconds between repetitions and each repetition was preceded by a tone to alert the students that the next repetition was about to start. Students wrote their transcriptions in blue pen for the first 6 repetitions and in black pen for the final 4. There was about one minute after the first six repetitions for the students to complete their answer so far and to change pens. There was also about a minute after the final repetition for students to finalise their answer.

Analysis

A transcribed segment was deemed to be correct if the correct symbol occurred in the correct position within the nonsense word. Marking was not determined by how close to the target the answer was, because of the restricted nature of the symbols to be chosen from (although this seems like a useful step to take in further experiments where the material is less restricted). No marks were given for incorrectly written symbols if they could be confused with another IPA symbol, or for metathesis errors (of which there were very few). For each student a mark out of six was given for each word for the response entered after 6 repetitions, and another for the response given after 10 repetitions.

Results

To investigate the effect of <u>number of repetitions</u> two paired samples t-tests were conducted, one for *word a* and one for *word b*. For *word a* students got an average of 2.1 segments correct after six repetitions compared to 2.8 segments after ten repetitions ($t(31)=-4.3$, $p<0.01$). For *word b* students got an average of 2.5 segments correct after six repetitions and 3.3 correct after ten repetitions ($t(31)=-5.01$, $p<0.01$). Most students made some improvement between six and ten repetitions, whilst a few did not make any improvement. Only one student performed more poorly after ten repetitions than after six, and this was only for one of the two words presented.

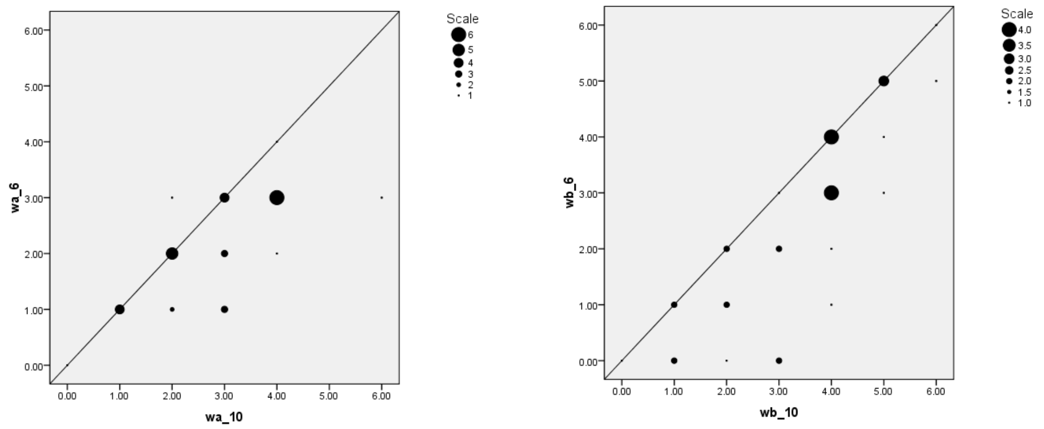

Figure 1 Correctly identified segments after six and ten repetitions for word a (left) and b (right)

In order to assess the effect of the <u>type of voice</u> (i.e. the sex of the speaker) an independent samples t-test was conducted on the total number of segments correct (the best scores in word a plus those in word b), with an independent factor of group. Results show no significant effect of group ($t(30)=1.44$, $p> 0.05$), indicating that those students hearing the majority of repetitions in the female voice did not outperform those hearing a majority of repetitions in a male voice.

In order to assess the effect of the *number of voices* a repeated measures MANOVA was conducted on the improvement scores (the score after ten repetitions minus that after six repetitions) with a single within-subjects factor of ‘word’ (2) and a between subjects factor of ‘group’ (2). An effect of number of voices would be indicated by a significant word*group interaction, because the two groups heard each word with a different numbers of voices. Results show that there is no significant effect of Word ($F(1,30) = 0.73$, $p>0.05$), or Group ($F(1,30)=3.7$ $p> 0.05$) and no significant Word*Group interaction ($F(1,30)= 3.2$ $p>0.05$). This shows that students found the words to be of equal difficulty that the groups are equally matched in terms of ability, and that the number of voices in which the word was presented did not affect transcription accuracy.

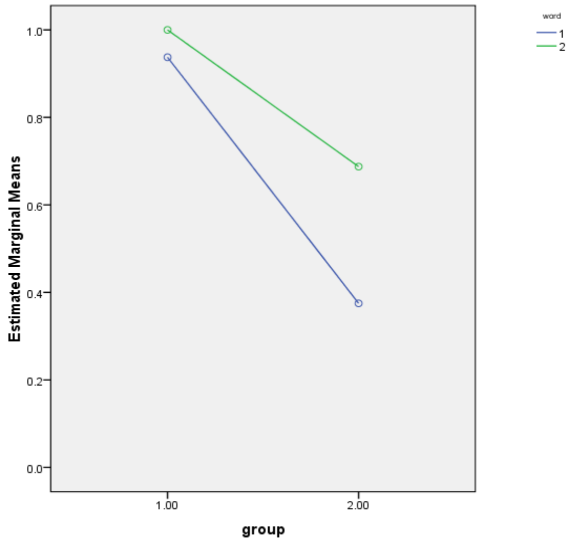

Figure 2 Improvement scores for word a (green) and b (blue) shown by group

Discussion

Repetitions

Results show that there is an effect of the number of repetitions on transcription accuracy, regardless of whether those repetitions occurred in the same or a different voice, or if the voice was similar to the one used for training. The majority of students improve in their accuracy for both words when they hear four extra repetitions.

There appears to be no verbal transformation effects like those found for words and nonwords in the literature cited above (e.g. Warren, 1961), as the vast majority of students' transcriptions improve rather that worsen after hearing an increased number of repetitions. There are a number of possible reasons why a verbal transformation effect has not been found. Firstly, transcribers hear multiple tokens, whereas the experiments that find verbal transformation effects use multiple repetitions of the same token. This account would fit with those explanations for the Verbal Transformation Effect (e.g. Snyder, Calef, Choban, & Geller, 1993), which propose that transformations occur due to habituation and satiation in the auditory system. Transformations are thought to be lessened if variation is present. In addition there are many more presentations in VTE experiments than when items are presented for transcription, so there may simply not be time for VTEs to arise.

Finally, the discrepancy may rest in the way nonsense words and IPA sounds are represented in the brain. The work cited above (particularly the work of Warren and colleagues) indicates that both real and nonwords are subject to VTEs. For real words, VTEs can be explained by habituation or satiation of lexical forms, and for nonwords the effect can be explained by the same processes operating on smaller units such as syllables or phonemes. Nonsense words, such as those presented here, do not have any lexical representations of their own as they have never been heard before. In addition many of the segments do not form a part of any words in the native language of the transcribers, so they also will not be subject to habituation spreading from other lexical items. It is certainly possible that a combination of these factors explains the lack of a VTE, and further experiments are necessary to clarify the situation, some of which are detailed below.

So, additional repetitions, at least up to 10, appear to improve transcription accuracy. There are two ways in which additional repetitions might benefit the transcription process. It may be that the nonsense words presented exceeded the capacity of the phonological loop so that students were not able to hold a detailed representation of the entire nonsense word in working memory after any one repetition. In this case, students might store different parts of the nonsense word after each repetition. More repetitions would therefore allow them to store, analyse and transcribe more parts of the nonsense word over the course of the exercise. The capacity is variously conceived in terms of items (around 7 ±2 items (Miller (1956)) or duration (1.5-2 seconds (Baddeley, Thomson and Buchanan (1975)). As the nonsense words were each of 6 segements and 1.2 (word a) and 1.4 (word b) seconds in duration, it is certainly possible that they exceed the capacity of the phonological loop.

There is some support for this suggestion by looking at the parts of the nonsense word that are attempted at different times. Segments at the beginning and the end of the word appear to be completed first (during the first six repetitions) than segments towards the middle of the nonsense word. This finding is reminiscent of serial position effects (e.g. Ebbinghaus 1885) where early and late items are remembered best when participants repeat lists of digits or words. The primacy effect is explained by initial items being rehearsed so often that they are transferred to long term memory, whilst recency is explained by final items still being in the phonological loop when the list has finished and has to be repeated. If this were the case it would suggest that a single nonsense word is treated more like a list of words or digits, which would seem intuitively to make sense, as each segment must be identified separately. This observation needs further testing by examining the accuracy and completeness of transcription after each repetition.

An alternatively explanation for why increased repetitions improve accuracy is that extra repetitions are useful because the different tokens contain variation. The method of presentation used here has ecological validity, as ear training classes (at least in the UK) tend to consist of the examiner repeating the same item multiple times rather than a recording where the same token is replayed. It is possible that variation occurring in the different tokens is helpful in the identification of segments because it refreshes the information stored in the phonological loop with a slightly different version after each repetition. So, for example, from one repetition to the next there may be slight changes in, say, voice onset time, which mean that the transcriber is presented with subtly different cues to a sound's identity after each repetition. If we assume that sounds learnt in class are stored as multiple exemplars, as is thought to be the case for real words and sublexical units, then variation makes it more likely that a token will match an exemplar in long term memory. However, if variation is the key to why increased repetitions improve accuracy we might expect transcription to improve more when those extra repetitions are spoken in a different voice, as this adds even more variability than repetitions in the same voice. As can be seen from the results, and is discussed below, this is not the case.

The current data do not allow us to tell whether repetitions improve accuracy due to the limited capacity of the phonological loop, due to variation supporting retrieval from long term memory, or to some combination of the two. Two further sets of experiments are planned which would allow us to further investigate these possibilities. Firstly, a comparison of repeated items and repeated tokens is necessary. Obviously a single repeated token would contain no variation, and therefore fewer chances for a similar exemplar of a segment to be found in long term memory. If accuracy is found to improve after more repetitions even for a repeated token, this would suggest that extra repetitions are not beneficial due to variability in the signal. Secondly, experiments using longer and more complex nonsense words are planned in order to further stretch the capacity of the phonological loop. If longer nonsense words (those that further exceed the duration of the loop) are found to benefit from additional repetitions even more than shorter words, this would suggest that extra repetitions allow different sections of the item to be attended to.

Voices

The vast majority of students were able to transcribe some parts of the nonsense words. The mean number of correct segments was just over 50% for the group, ranging from 0 to 12 segments correct out of a possible 12 (6 in each word). As both voices were unfamiliar to the students it is clear that some degree of generalisation is possible, and that this generalisation is not dependant on the type of voice heard (there was no difference between accuracy of transcription for male and female voices). However, performance in the experiment was poorer than in the class test that students took a week later. In the class test students got an average of 80% of segments correct when the same marking scheme was applied. This difference may be due to a week of extra revision, to the addition of visual cues in the class test, or to the voice in the class test being familiar.

However, from the other results presented above it seems that the discrepancy between the experiment and the class test is unlikely to be due to voice familiarity. If familiarity is an issue then we would suspect that voices similar to voices in which training has been conducted should be transcribed more accurately. As the entirety of the students training has been with one female lecturer it can be hypothesised that they would do better when transcribing a female voice. This was not the case in the experiment as group 1, who heard the majority of repetitions in the female voice, did not transcribe more accurately than group 2, who heard a male voice more often.

There is also no effect of the number of voices heard on transcription accuracy. It was hypothesised that hearing more voices could hinder transcription accuracy because transcribers have to allocate more working memory resources to dealing with variability, and make more searchers for similar tokens in long term memory. In the event it seems that there was very little difference between transcriptions produced when nonsense words were heard in one or two voices.

The reason that there is no effect of number of voices is somewhat unclear. It may be that the task of nonsense word transcription is not affected my multiple voices because it does not require the processing of indexical information. Mullenix and Pisoni (1990) suggested that indexical information is processed regardless of whether it is required for the task. They asked subjects to classify words by phoneme (/b/ or /p/) or by speaker sex, and found that the phoneme classification test was still affected by the number of voices heard. However, there is still an element of meaning involved in this task as /p/ and /b/ are phonemes in the participants' native language. For nonsense words, and the segments they contain, there is no equivalent linguistic function. Goldinger (1996: 1180) indicates that 'if voice details are tangential to the linguistic function of speech, they may be minimized in episodic traces'.

It seems likely, therefore, that the non-significant effect of number of voices is due to the nature of the task. Nonsense words are made up of sounds, some of which never play a meaningful role in the transcribers' native language. Due to the non linguistic nature of many of the IPA sounds to which students are exposed, there is no need to attend to indexical information in the encoding and rehearsal stages of working memory, or in the episodic traces of sounds in long term memory.

Conclusion and implications

From the work presented here, it can be seen that there is an effect of the number of repetitions of nonsense words on transcription accuracy but not of the number or type of voices. The facilitative effect of the number of repetitions may be due to variability found across tokens or to the limited capacity of the phonological loop. The lack of an effect of voices seems to be because transcription of nonsense words is very unlike normal speech processing in that the sounds do not have a linguistic meaning and therefore indexical information is not utilised in encoding, rehearsal, or stored representations. Further experiments are needed to investigate the role of repetitions in nonsense words of different lengths, and the difference between repeated items and tokens.

This work has some implications for clinical practice. We already know that multiple listenings can improve the reliability of transcriptions (e.g. Munson and Brinkman, 2004). If we assume that increased listenings to single token also improves accuracy then recordings should be taken in clinic and listened to multiple times at a later point.

There are also a number of implications for good practice in teaching and learning phonetic transcription. Firstly, as it seems that the number of repetitions can make a difference to transcription accuracy it is important that institutions move towards a standardised format for formative examinations so that no group of students is disadvantaged. The finding that the number and type of voices does not affect accuracy suggests that students can generalise across voices and, therefore, that we can continue to have students taught phonetics by only one member of staff. However, more work is necessary to establish the nature of mental representations for segments which are not part of a person's native languages and are learnt solely in ear training classes. As has been found in studies of second language learning (e.g. Logan, Lively and Pisoni, 1990), it is possible that variability in speakers heard during training will improve the robustness of mental representations for newly learned speech sounds.

Acknowledgements

This work was supported by a teaching and learning fellowship from City University London. Thanks to the participants at the British Association of Clinical Linguistics workshop in Reading, where parts of this work were presented in December 2008, and to Jo Verhoeven for useful discussions. Thanks to Mark J. Jones and Rachel Smith for recording the nonsense words, and to the students who took part in the experiment.

References

Amorosa, H., von Benda, U., Wagner, E. & Keck, A. (1985). Transcribing phonetic detail in thespeech of unintelligible children: A comparison of procedures. *International Journal of Language & Communication Disorders*, 20:3, 2, 81 — 287.

Baddeley, A.D. (1986). Working memory. Oxford: Clarendon Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 1, 158-173.

Baddeley, A., Thompson, N., & Buchanan, M. (1975). Word Length and the Structure of Memory. *Journal of Verbal Learning and Verbal Behaviour*, 1 , 575-589.

Ebbinghaus, H. (1885) Memory: A Contribution to Experimental Psychology. <http://psychclassics.yorku.ca/Ebbinghaus/> [accessed 16.03.09]

Goldinger, S (1988). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-279.

Goldinger, S. (1996) Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 5, 1166-1183.

Goldinger, S., Pisoni, D., & Logan, J. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 152–162.

Holm, A., Crosbie, S. & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language and Communication Disorders*, 42(4) 467 – 486.

Howard, S. & Heselwood, B. (2002). Learning and teaching phonetic transcription for clinical purposes. *Clinical linguistics & phonetics*, 16 (5), 371- 401.

Jesse, A., McQueen, J., & Page, M. (2007). The locus of talker-specific effects in spoken-word recognition. *Proceeding of the 16th Congress of the International Phonetics Association*, Saarbrucken, 1921-1924.

Logan, J., Lively, S. & Pisoni, D. (1991). Training Japanese listeners to identify Ennglish /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 2, 874-886.

Mackenzie Beck, J. (2003). Is it possible to predict students' ability to develop skills in practical phonetics? In *Proceedings of the 15th International Congress of Phonetic Sciences,* Barcelona, 2833 – 2836.

Maguire, E. & Knight, R-A. (in preparation). The effect of short term memory on the phonetic transcription ability of students.

Martin, C., Mullenix, J., Pisoni, D. & Summers, W. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 17, 152-162.

Middelweerd, M. & Plomp, R. (1987) The effect of speechreading on the speech-reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, 82, 6, 2145-2146.

Miller G. (1956) The Magical Number Seven, Plus or Minus Two. *The Psychological Review*, 63, 2, 81-97.

Mullennix, J., & Pisoni, D. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47, 379-390.

Mullenix, J, Pisoni, D., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.

Munson, B. & Brinkman, K. (2004). The influence of multiple presentations on judgments of children's phonetic accuracy. *American Journal of Speech-Language Pathology*, 13, 341–354.

Natsoulas, T. (1965). A study of the verbal-transformation effect. *American Journal of Psychology*, 78, 257-263.

Noble, P. (submitted). Can Phonetic Ability be Predicted? submitted to *International Journal of Language and Communication Disorders*.

Oller, D. & Eilers, R. (1975). Phonetic expectation and transcription validity. *Phonetica*, *31*, 288-304

Shoaf, L. & Pitt, M. (2002). Does node stability underlie the verbal transformation effect? A test of node structure theory. *Perception & Psychophysics*, 64 (5), 795-803.

Smith, R. (2007). The effect of talker familiarity on word segmentation in noise. *Proceedings of the 16th Congress of the International Phonetics Association*, Saarbrucken, 1917 – 1920.

Sumby, W. & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America*, 26, 2, 212-215.

Snyder, K. A., Calef, R. S., Choban, M. C. & Geller, E. S. (1993). Effects of word repetition and presentation rate on the frequency of verbal transformations: Support for habituation. *Bulletin of the Psychonomic Society*, 31(2), 91–93.

Warren, R. (1961). Illusory changes of distinct speech upon repetition—The verbal transformation effect. *British Journal of Psychology*, 52, 249-258.

Warren, R. & Gregory, R. (l958). An auditory analogue of the visual reversible figure. American Journal of Psychology, 71, 6l2-613.