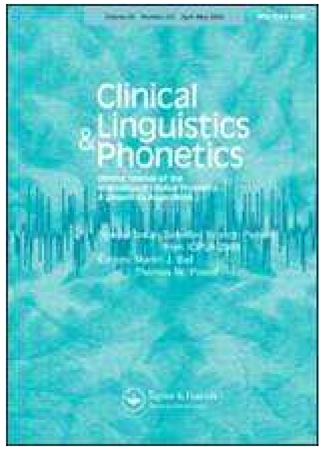

Transcribing nonsense words: The effect of numbers of voices and repetitions.

Journal:	*Clinical Linguistics & Phonetics*
Manuscript ID:	TCLP-2009-0087.R1
Manuscript Type:	Original paper
Keywords:	phonetics < speech, phonetic transcription < speech production measurement < speech, speech production measurement < speech, speech perception < hearing

Introduction

Phonetic transcription is, of course, the method by which sounds of speech can be recorded using symbols from the International Phonetic 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. For clinicians, transcription is often 'a necessary first step in assessment' (Howard and Heselwood 2002, 373), and therefore it is important for us to understand which factors might affect transcription accuracy. Knowledge of these factors may also help tutors and students in the process of learning and teaching, and ensure that assessment practices are fair and comparable across institutions.

The notion of transcription 'accuracy' can be problematic. As Heselwood and Howard (2008, 388) point out, giving the examples of [jʷ], [ɥ] and [wʲ], different transcriptions may in fact be functionally equivalent, and must be 'made, and interpreted, within a framework of phonetic theory'. That being said, there are certain conditions under which we can be confident that a single transcription is more accurate than alternatives. If a transcription is provided to a trained phonetician, who then produces the associated sequence after several practice attempts, we can probably be surer of the ideal transcription for that production than for an instance of spontaneous speech produced by a person with a speech disorder, for example. Whilst the notion of 'functional equivalence' will be relevant for both examples, it will be easier to decide if a given transcription is accurate or not in the former. Indeed, the former type of presentation is often relied upon when teaching and assessing transcription ability during phonetics training. In what follows, the notion of accuracy is taken to mean the degree to which a transcription is similar to the target in ideal circumstances, and when we can be sure of that target.

In order to fully appreciate the factors that might affect transcription accuracy (bearing in mind the caveats above) we can consider the various stages in the transcription process. Firstly, the signal must be heard (and seen) and attended to. We can assume, in line with 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 (for example voice, place and manner labels) will be accessed, as will the appropriate symbols, which must then be written down. Differences in accuracy between transcribers can, therefore, be due to differences occurring at any stage of this process.

Transcription accuracy may also be affected by a number of factors external to the transcription process. For example, knowledge of phonetic theory will likely determine what can be identified and transcribed. In addition, factors concerning the item to be transcribed will also affect accuracy, such as the duration of the item (which will affect how much can be stored in the phonological loop), and whether or not the transcriber knows the identity of the item (Oller and Eilers (1975)). The nature of presentation is also likely to affect 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, as adding visual elements increases the signal-to-noise ratio significantly, at least in a noisy environment, (e.g. Sumby and Pollack, 1954, Middleweerd and Plomp, 1987). The current paper deals with two additional aspects of presentation and their putative effects on the accuracy of transcription of nonsense words: the number of voices, and the number of repetitions.

Number of voices

The number of voices in which an item for transcription is presented may affect accuracy. Research suggests that normal speech processing is affected by whether words are spoken in one or multiple voices. Results have shown that there is an advantage in word recognition and naming (Mullennix, Pisoni and Martin (1989)), and for lexical decision (Goldinger 1998) when words are spoken in a voice that has been heard before.

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.

More recently, episodic theories of speech perception (e.g. Goldinger 1998) suggest that same-voice effects are due to the long-term storage and representation of words. These theories consider the mental lexicon to be composed of multiple exemplars of each word, and propose that these exemplars contain surface detail, including information about voices. During word recognition, the incoming signal is compared to these multiple exemplars, and words will be recognised more quickly the more similar they are to the exemplars. As words heard in a familiar voice will be very similar to stored exemplars, they will be recognised more quickly than the same words spoken in a new 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 (at least when the first repetition is heard). 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 work by Jesse McQueen and Page (2007), on normal speech perception, indicates that units smaller than words are also subject to the effects of multiple talkers. 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. The fact that small, concrete units are subject to same-voice effects, suggests that sounds learnt in phonetic training might also be subject to these effects.

If sounds learnt in phonetic training *are* stored in a similar fashion to words and segments learnt as part of language, then the number of voices in which a word is presented may have an effect on transcription accuracy. Words presented in more than one voice may be transcribed less accurately than those presented in a single voice as transcribers cope with matching a variable signal to multiple stored representations. It is hypothesised therefore, that if such sounds are represented in the same way as sounds in a subject's native language(s), then hearing repetitions in different voices will decrease accuracy in transcription because the transcriber has to handle different voices in working memory and compare them to multiple tokens in episodic memory.

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. Nevertheless, most educational institutions present far more than one repetition when they are training and examining students, and for this reason we need to know how repetitions affect accuracy.

Most of the literature does not address transcription accuracy, preferring to concentrate on issues of reliability, presumably because of the difficulties of deciding what constitutes an accurate transcription. Amarosa, von Benda, Wagner, and Keck's (1985) qualitative analysis of transcription errors indicates that transcriptions made from multiple repetitions were less similar to a know real-word target than items made after one repetition, and suggest that transcribers should be allowed to listen to items for transcription as often as they find necessary. However, Shriberg, Kwiatkowski and Hoffman (1984) suggest that, when transcribers disagree, items should not be listened to more that three times, because repeated listenings can lead to auditory illusions. Furthermore, Munson and Brinkman (2004: 342) suggest that repeated listening can lead to the verbal transformation effects (VTEs) first reported by Warren and Gregory (1958). For example, multiple repetitions of the word 'truce' might result in listeners hearing transformations such as 'truth' as well as phonologically dissimilar words such as 'Esther', and even nonwords (Shoaf and Pitt (2002: 795)). VTEs are also found when nonwords (both phonotactically legal and illegal) are presented. In fact nonwords may give rise to more transformations (Warren, 1961), and the first transformation may occur more quickly, than is the case for real words (Natsoulous, 1965). These findings seem to suggest that transcription accuracy may decrease with multiple hearings of an item.

From a teaching and learning perspective J.C. Wells (personal communication) also suggests that transcriptions of nonsense words tend not to improve after about six repetitions, and accuracy may even decrease with extra repetitions 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. The experiment below examines the accuracy of student transcriptions after six and ten repetitions of a nonsense word, with the hypothesis that increased repetitions will lead to decreased accuracy.

Method

The assessment of transcription accuracy lends itself to an experimental approach, because 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. They had passed first year courses in phonetics and phonology but were still several months ahead of their final transcription assessments. Students were all female, broadly reflecting the gender balance of the programme and the profession, were of mixed ability by their own report (see Knight, 2009 for a survey with the same students), and their ages ranged from 19-42. Students were divided quasi-randomly into two equal groups of 16, depending on which side of the room they were seated on the day of the experiment.

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 made up of sounds from a set that had been introduced to students in preparation 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 cardinal vowels 1,2, 3 and 8. The nonsense words the speakers were asked to produce were:

a. [ŋǁ ɔ dn u ˈʁ ɑ] and

b. [œ lˠ u ˈʎ a ts’].

Sounds such as [dn] and [ŋǁ] had been introduced to the students as single segments, hence each word can be considered as six segments in length. The nonsense words produced by the phoneticians were then transcribed by two other trained phoneticians to ensure that the words had been accurately rendered, and to see if they were equivalent in the two voices. The expert transcribers disagreed on three of the segments. In word A one transcriber identified the first vowel as cardinal 6 and one as cardinal 7, and for the final vowel, one transcriber identified it as cardinal 4 and one as cardinal 5. For word B, one transcriber identified the palatal approximant as a lateral while the other did not. In each case the transcribers thought that the words were spoken in the same way by both speakers, but the transcribers disagreed about the identity of the segments. This disagreement was taken into account in scoring the students’ transcriptions, as described below. The expert transcribers were also asked to compare the voices of the speakers. They both indicated that the male voice was lower in pitch and less breathy, but that the durations of each segment, and the prosody of the nonsense words were equivalent in the two voices.

Presentation

Nonsense words were presented over loudspeakers to each group of students separately. Group 1 heard word B first followed by word A, whereas group 2 heard word A followed by word B. Each group heard their first word presented in two voices (six repetitions in one voice followed by four 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 ten 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. This inter-stimulus interval was chosen by reference to the normal procedure in class, and had been piloted with students not taking part in the current experiment.

Students were told that each nonsense word would contain six segments, as would be the case in their upcoming class test, and completed their transcriptions under exam conditions. They wrote in blue pen for the first six repetitions and in black pen for the final four, so that transcriptions at different stages could be identified. 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 answers.

Analysis

Student transcriptions were scored using a version of a feature matrix (e.g. Cucchiarini (1996)). Each segment was assigned a maximum score based on its features; essentially one mark was assigned for each feature in a sound's label. To begin with vowels were scored out of three, reflecting dimensions of backness, openness, and roundness. Consonants were also scored out of at least three, reflecting voice, place and manner labels, and features such as non-pulmonic airstream mechanisms, and secondary articulations were also assigned a mark. This method of scoring meant that discrepancies between the expert transcribers could be taken into account. The vowels for which there had been disagreement were each scored out of two, disregarding the feature that was not agreed upon. The approximant in word B was scored out of three, removing the lateral feature which the expert transcribers had not agreed upon. In this way students were not penalized for choosing either of the alternatives for disagreed upon sounds. For each student, a 'percentage correct score' was calculated for each word after six and ten repetitions.

Results

A repeated measures MANOVA was conducted on the percentage correct scores, with within-subjects factors of 'word' (2) and 'repetitions' (2), and a between-subjects factor of 'group' (2). Results show that there was no significant effect of Word ($F(1,30) = 0.28$, $p>0.05$), or Group ($F(1,30)=0.01$ $p> 0.05$), indicating that the words were of equal difficulty and that the groups were equally matched in terms of transcription accuracy. There was however a significant effect of repetitions ($F(1,30)= 45.2$, $p<0.01$), as shown in Figures 1 and 2. Students scored an average of 52% correct after six repetitions and 68% after ten repetitions. There was also a significant interaction between repetitions and group (F(1,30) = 6.16, p<0.05). Compared to group 2, group 1 performed less well after six repetitions, but better after ten repetitions for each word. An effect of the number of voices heard would be indicated by a significant word*group interaction, because the two groups heard each word with a different numbers of voices. However no such interaction was found (F(1,30)= 0.58 p>0.05). No other interactions were significant.

Figures 1 and 2 about here

Results for individual segments

The significant effect of number of repetitions was further investigated. In order to see if individual segments are equally affected by extra repetitions, paired sample t-tests between six and ten repetitions were conducted for each segment. The mean scores after six and ten repetitions, and the results of the t-tests, are shown in Tables 1 and 2.

Tables 1 and 2 about here

Improvements were found for all segments when more repetitions were heard, and significant improvements were found for eight.

Errors

A full error analysis was not undertaken because the students did not have a full choice of all IPA sounds. Their responses were constrained by the set of sounds they knew might appear, and their transcriptions reflected this. However, a brief description of errors is given for the three sounds that were the most poorly transcribed after ten repetitions, as this may indicate future avenues for research. For [dn], the most common errors were failing to write anything down, and transcribing the sound as a voiceless alveolar nasal, which was one of the sounds present in the set that might appear. For [œ], the most common errors were [ɛ] and [ɜ]. For [u] in word B, the most common errors were [i], and to leave the symbol out entirely.

Discussion

The aim of the reported experiment was to assess the effect on transcription accuracy of the number of voices and repetitions in which a nonsense word is presented.

Number of Voices

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

The reason that there was no effect of the number of voices is somewhat unclear, however. It might be that two voices were not enough to trigger same-voice effects. We are often not told how many voices are used in episodic-memory-type experiments, although Goldinger, Kleider and Shelley (1999), at least, found effects with just two voices for real word material. Alternatively, the lack of an effect may be because voice details are not processed in this task. The most extreme version of this explanation would be that IPA sounds are not represented as multiple exemplars and therefore voice details are not stored in memory, or processed in nonsense word transcription. A less extreme suggestion would be that IPA sounds are stored as multiple representations, but only details such as a sound's phonetic environment are present and voice details are ignored. Yet another alternative is that voice details of IPA sounds *are* stored and processed, but were just unhelpful in this task as there were no exact voice matches to the stored exemplars (because both voices were unfamiliar). An exemplar store of IPA sounds would be very different from the native lexicon, containing only relatively few voices and tokens of each sound. Therefore, unless, there is an exact match to be found, the number of voices heard may make very little difference to transcription accuracy.

It is not possible to choose between these different alternatives from the current results and clearly more work is needed to establish the nature of mental representations for IPA sounds. An experimental approach to this issue would need to vary factors such as the number of voices presented, and the familiarity of those voices. This is particularly important for teaching and learning, and for clinical practice, where it will be important to understand the relationship between transcribing known and new voices.

Number of Repetitions

Results show that students produce more accurate transcriptions after ten repetitions than after six. This was contrary to the hypothesis that extra repetitions would lead to decreased accuracy. The hypothesis was based on the findings of studies showing Verbal Transformation Effects (VTEs), whereby multiple hearings cause changes in perception away from the target. No verbal transformation effects were found in the current results, as the vast majority of students' transcriptions improve rather that worsen after hearing an increased number of repetitions, and there is at least a trend for the accuracy on each segment to improve. The lack of VTEs may be because students heard multiple tokens, whereas experiments demonstrating VTEs usually present multiple repetitions of the same token. In addition VTE experiments present many more repetitions of items than were used here (although exactly how many is sometimes unclear), so there may simply not have been time for VTEs to arise. It will be useful in future to present many repetitions of a single token to see if VTEs ever arise in nonsense word transcription.

The effect of extra repetitions was, in fact, to improve accuracy, as students produced transcriptions that were closer to the target item. Improvements were found for all segments, and significant improvements were found for eight. There does not seem to be an obvious factor linking the four segments for which no significant effects were found, either in terms of phonetic identity, position in the word, familiarity, or accuracy (some segments for which results were non-significant, like [lˠ] were accurately identified after both six and ten repetitions, while others, like [u] in word

B, were not accurately identified at any point). Testing with a wider variety of sounds, in different positions, and in nonsense words of different lengths, would help to reveal why the transcription of some sounds does not improve significantly.

So, why do extra repetitions improve accuracy? One explanation might be that our previous hypothesis misinterpreted the role of variability in the signal, and that variability in different tokens actually has a facilitative effect. Whilst we hypothesized that variation between voices might hinder transcription, it could be possible that variation helps transcribers identify segments. As the repetitions were made up of different tokens, they would have contained subtle variations, even when in the same voice. This variation might have been helpful because it refreshed the information stored in the phonological loop with a slightly different version after each repetition, meaning that the students were presented with subtly different cues to a sound's identity. However, if variation is the key to why increased repetitions improve accuracy, we might expect transcriptions to improve more when extra repetitions were spoken in a different voice, as this would add even more variability than repetitions in the same voice. As can be seen from the results above this was not the case.

An alternative explanation is that extra repetitions may have been helpful because the nonsense words exceeded the capacity of the phonological loop. The loop is thought to have a capacity of around 7 (±2) items (Miller (1956)), or 1.5-2 seconds (Baddeley, Thomson and Buchanan (1975)). As the nonsense words were each of six segments, and 1.2 (word A) and 1.4 (word B) seconds in duration, they may have stretched the capacity of the phonological loop. If loop capacity were exceeded, transcribers would be able to focus only on *sections* of the nonsense word during each repetition, and therefore more repetitions would allow them to focus on more sections.

Some support for this suggestion comes from identifying when different parts of the nonsense words were attempted. Segments at the beginning and the end of the words appear to have been completed before (during the first six repetitions) segments towards the middle. 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 is also happening when people transcribe it would suggest that a single nonsense word is treated more like a list than a single word, which has intuitive appeal, because each segment must be identified separately. This observation may not be accurate for all transcribers, who we know often develop individual strategies (e.g. Fokes and Bond, 1995), but is worthy of further testing by examining the accuracy and completeness of transcriptions after each repetition.

The current data suggest a number of further experiments in order for us to identify definitively the role of variation and memory limitations in transcription accuracy. If memory limitations are important then we may expect transcribers with better working memory capacities to be better transcribers. Preliminary data (Maguire and Knight, in preparation) seems to suggest that this is the case. In addition, presenting longer and more complex nonsense words would further stretch the capacity of the phonological loop. If longer nonsense words are found to benefit from additional repetitions even more than shorter words, this would suggest that extra repetitions do indeed allow different sections of the item to be stored in the loop and attended to (see e.g. Ashby, Maidment and Abberton, (1996) for a similar teaching and learning approach). A further comparison of repeated items and repeated tokens is necessary to see if Verbal Transformation Effects arise. However, if accuracy improves after more repetitions even for a repeated token, this would confirm that extra repetitions are not beneficial due to variability in the signal.

Conclusion and implications

In conclusion, there is an effect of the number of repetitions of nonsense words on transcription accuracy, but no effect of the number of voices in which they are presented. The lack of an effect of voices seems to suggest that voices details were not processed, but further experiments are necessary to decide if IPA sounds are not stored as multiple exemplars, or if effects did not arise because both voices were unfamiliar. The facilitative effect of extra repetitions is likely due to the limited capacity of the phonological loop, but further experiments are needed to investigate the role of repetitions in nonsense words of different lengths, and the difference between repeated items and tokens.

The results have some implications for good practice in teaching and learning phonetic transcription. As the number of repetitions can make a difference to transcription accuracy it is important that institutions move towards a standardised format for examinations, so that no group of students is disadvantaged. The finding that the number of voices in which a word is presented does not affect accuracy suggests that voice details may not be processed in transcription, and perhaps, therefore, that it is appropriate for students to be taught phonetics by a single member of staff, as is quite often the case. However, we must be cautious, because we are still unsure of the role of voice familiarity, and because studies of second language learning (e.g. Logan, Lively and Pisoni, 1990), have shown that variation in the speakers heard during training improves the robustness of mental representations for newly learned speech sounds. Therefore, not only is more work necessary to establish the factors that affect transcription accuracy, but also 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.

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 and Susanna Martin for useful discussions and for transcribing the nonsense words. Thanks also to Mark J. Jones and Rachel Smith for recording the nonsense words, and to the students who took part in the experiment.

References

Ashby, M., Maidment, J. and Abberton, E. (1996). Analytic listening: a new approach to ear-training. *Speech, Hearing and Language* 9: 1-10.

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.

Cucciarini, C. (1996). Assessing transcription agreement: methodological aspects. *Clinical Linguistics & Phonetics*, 10, 2, 131-155.

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

Fokes,J. & Bond, Z. (1005) Protocol analysis and the process of transcription, *Proc. of 13th International Congress of Phonetic Sciences*, 528-531.

Goldinger, S (1998). 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.

Goldinger, S., Kleider, H., and Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory and Cognition, 27 (2)*, 328-338.

Heselwood, B. & Howard, S. (2008) Clinical phonetic transcription. In Ball, M.J., Perkins, M., Müller, N. & Howard, S. (eds.) *The handbook of clinical linguistics*. Oxford: Blackwell. pp.381-399.

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.

Knight, R-A. (2009) Feeling confident about transcription (?): A student survey concerning numbers of repetitions and new voices. *Proceedings of the Phonetics Teaching and Learning Conference*, University College London.

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

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

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.

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.

Shriberg L D; Kwiatkowski J; Hoffmann K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research*, 27, 3, 456-65.

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.

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

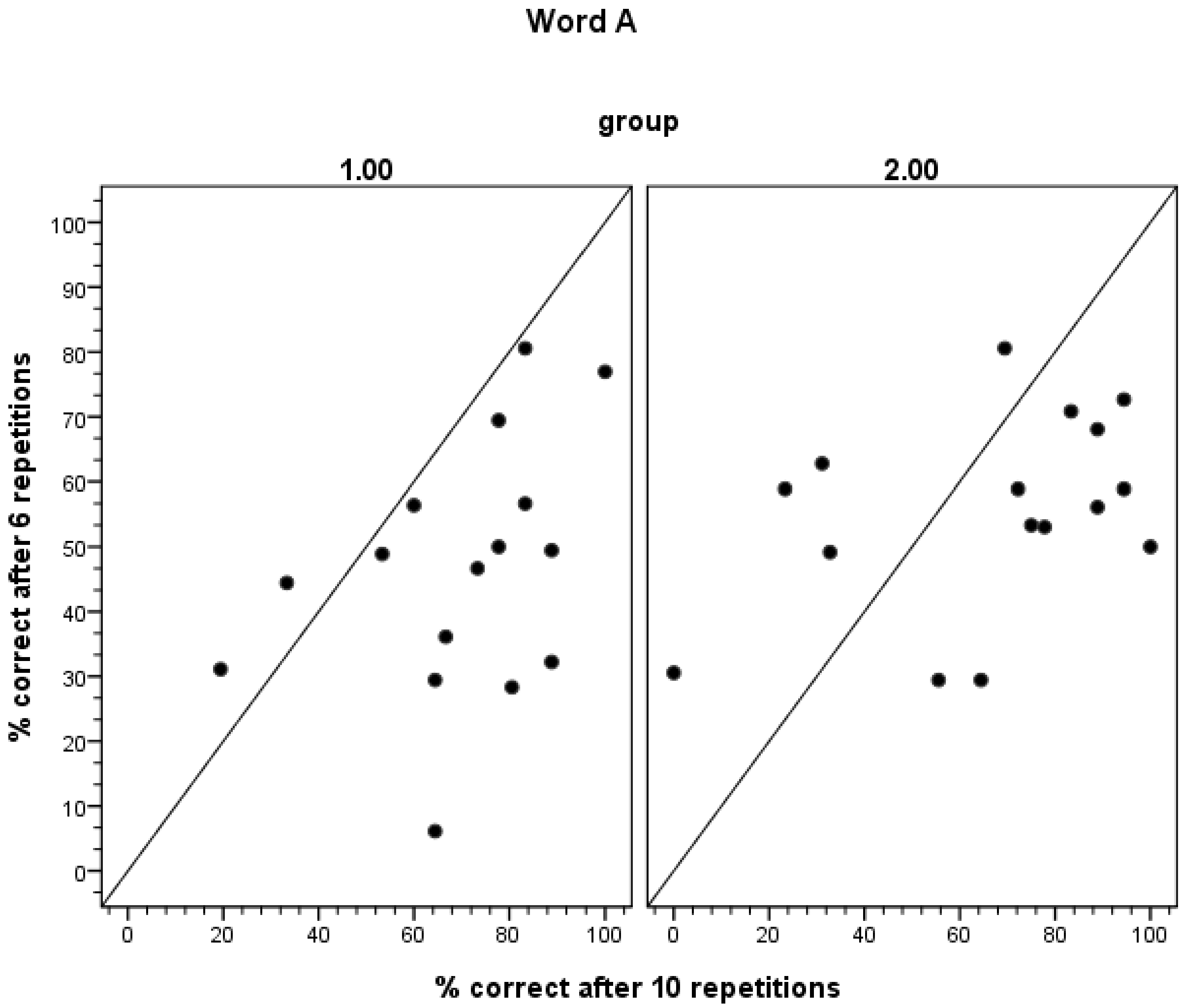

Figure 1 Scatter plot of each subject's score after 6 and 10 repetitions for word A.

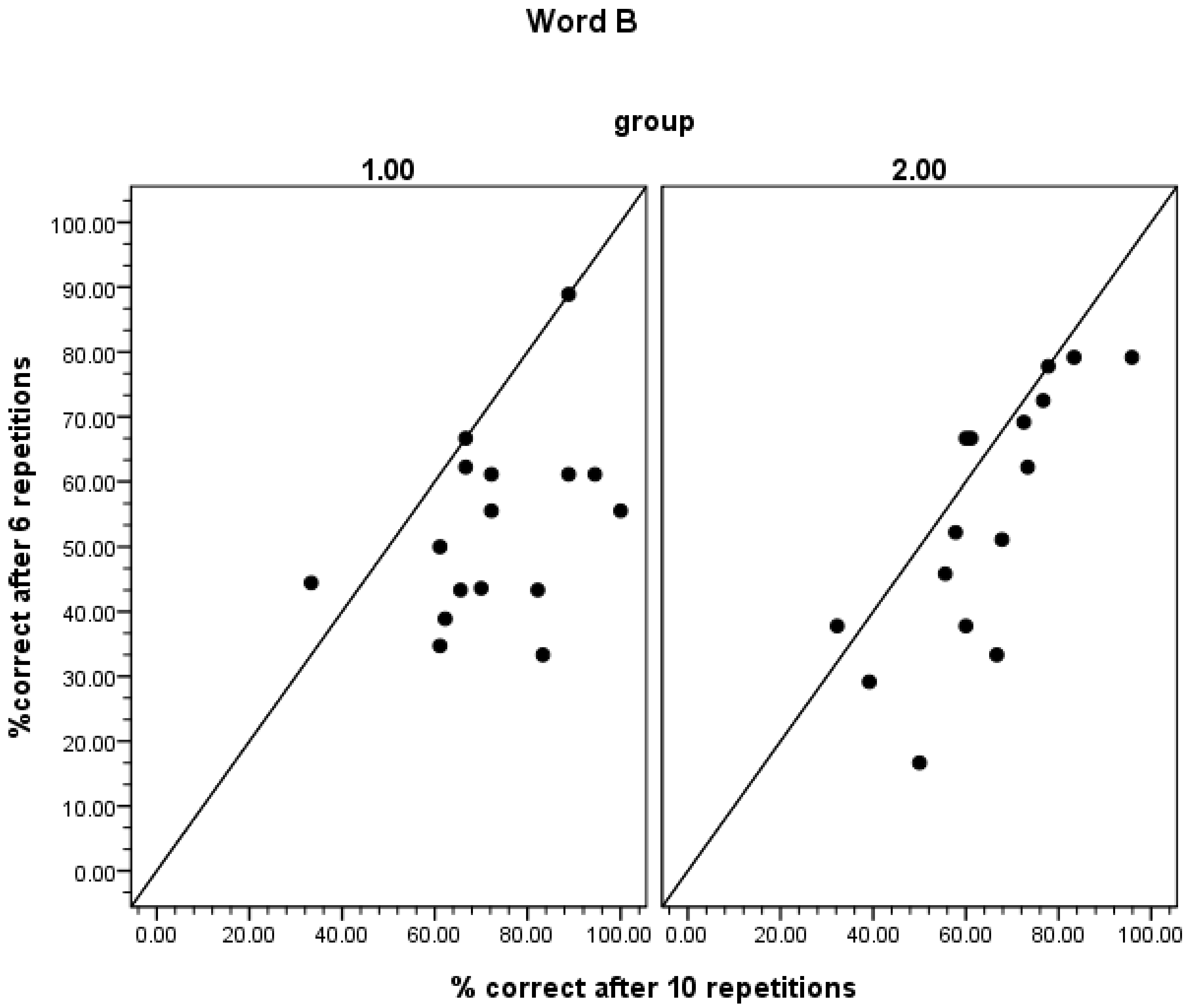

Figure 2 Scatter plot of each subject's score after 6 and 10 repetitions for word B.

Word A

Segment	ŋǁ	ɔ	dn	u	ʁ	ɑ
% correct after 6 repetitions	72	64	29	36	42	63
% correct after 10 repetitions	79	81	45	67	51	83
significance	<0.05	<0.01	<0.01	<0.01	>0.5	<0.01

Table 1. Overall percentage total correct, and t-test result, for each segment in word A.

Word B

Segment	œ	$l^{ɣ}$	u	ʎ	a	ts’
% correct after 6 repetitions	41	76	27	38	56	91
% correct after 10 repetitions	49	81	38	63	84	98
significance	>0.05	>0.05	>0.05	<0.01	<0.01	<0.05

Table 1. Overall percentage total correct, and t-test result, for each segment in word B.