**Effect of Variability in Talkers on Spoken Word Recognition and Recall**

Sumanth P.\* $, Pooja C.\*, & Rajasudhakar R.\*\*

**Abstract**

How do listeners integrate temporally distributed phonemic information into coherent representations of syllables and words when it is spoken by different speakers the same linguistic units? In day to day life, individuals must comprehend speech produced by many different talkers (Hager, & Amanda, 2013). Perceptual invariance of speech signal evolves an interesting topic to study among single and multiple talkers. Little, however, is known about the influence of talker variances in spoken word recall ability in Indian context. The present study aimed to evaluate the spoken word recognition and recall ability under single and multi-talker variations in adult neuro-typical individuals. 30 participants (15 males and 15 females) whose age was between 18 to 21 years were chosen for the study where each of them were presented with word list of single (List A) and multi talkers (List B) in Kannada. Results indicated that the participants were able to recall more number of words from List A compared to List B. Study re-affirms the serial order position effect where, participants were able to recall words better in initial and final position of a list than the middle ones. These results provide new insight into the perceptual constancy and its difficulties when spoken by multiple speakers.

**Keywords:***Spoken word recognition, recall ability, word perception, talker variability.*

**Introduction**

Speech perception is the process of imposing a meaningful perceptual experience on an otherwise meaningless speech input (Massaro, 2001). It is a complex process that involves converting highly encoded physical stimulus into some kind of abstract neural representation. A fundamental theoretical problem that remains unsolved despite decades of speech research is the lack of invariance between the speech signal and perceptual categories.When differences in talker and speaking rate are considered, the problem of lack of invariance becomes considerably worse, because these factors provide additional sources of many-to-many mappings. Across talkers, the same acoustic pattern can correspond to different perceptual categories, whereas different talkers’ productions of the same phoneme can differ acoustically (Peterson & Barney, 1952; Dorman, Studdert-Kennedy, & Raphael, 1977).

Mental representations of speech are much more detailed and more elaborate and they contain several sources of information about the talker's speech characteristics. Speech variability has an effect on both speech processing and representation (McLennan & Luce, 2005), maintaining detailed representations of talker-specific characteristics assists word identification for known talkers (Nygaard, Sommers&Pisoni, 1994). In contrast, both intra-talker and inter-talker variability can cause performance decrements. For example, in word identification and memory tasks, accuracy suffers when stimuli vary in speaking rate or style, or include multiple talkers or tokens (Mullennix, Pisoni& Martin, 1989; Nygaard, Sommers, &Pisoni, 1995; Uchanski&Braida, 1998; Bradlow, Nygaard&Pisoni, 1999; Sommers&Barcroft, 2006; Magnuson &Nusbaum, 2007).

Speech perception and spoken word recognition accuracy depend on a wide range of talker, listener and utterance-related characteristics, all of which can vary across communicative situations. Speech signals display a great deal of variability primarily because of factors that influence the production of spoken language. Among these are within and between talker variability, changes in speaking rate and dialect, differences in social contexts, syntactic, semantic and pragmatic effects, as well as a wide variety of effects due to the ambient environment such as background noise, reverberation and microphone characteristics (Klatt, 1986). In one experiment, Sommers, Nygaard&Pisoni(1992) found that words produced at different speaking rates (i.e., fast, medium and slow) were identified more poorly than the same words produced at only one speaking rate. Indeed, in current word recognition models, much emphasis is placed on factors such as word frequency and lexical structure and its effects on spoken word recognition performance using a variety of experimental paradigms (Landauer& Streeter, 1973; Eukel, 1980; Luce, 1986).

Creelman (1957) investigated the effects of talker variability on the recognition of spoken phonetically balanced (PB) words. Author played the monosyllabic words in noise to a group of five listeners which was recorded by 1, 2, 4, 8, or 16 talkers. The results showed that the words presented in the lists spoken by two or more talkers were identified less accurately than words presented in the list spoken by only a single talker and the differences in performance were relatively small: on the order of 7% - 10%.

Mullennix et al. (1989) investigated spoken word recognition in continuation of the Creelman (1957) study where, thirty-seven undergraduate students from introductory psychology courses at Indiana University volunteered as participants. Fifteen participants served as talkers to produce the stimulus materials and another 22 participants served as listeners in the perceptual experiment. All participants were native English speakers and reported no history of speech or hearing disorders at the time of testing. The stimulus consisted of 68 spoken words (CVC monosyllabic English words containing wide variety of consonants) obtained from each 15 different talkers of same language with signal-to-noise ratios (SNRs) +10, 0 and -10 dB as background noise variation and for other 22 participants were instructed to type the word corresponding to what they thought they had heard on each trail. A 2 minutes rest period occurred between each block and at different SNR in each particular block. Results showed that word recognition accuracy decreased across the SNR from +10 to -10 in both the talkers and better word recognition was reported for single talker than the multi talker’s conditions.

Variability in speaking rate has also been shown to affect speech perception and spoken word recognition (Miller &Liberman, 1979; Summerfield, 1981; Miller &Volaitis, 1989; Volaitis& Miller, 1992). For example, Sommers, Nygaard&Pisoni (1994) showed a decrease in word identification scores for mixed-speaking-rate lists relative to single-speaking-rate lists. Interestingly, however, no decrease in word identification scores was found for lists of words that had mixed overall amplitudes relative to lists of words presented at single overall amplitude. These results suggest that variability due to talker characteristic and speaking rate is both time and resource demanding. As the talker or speaking rate changes from trial to trial in these tasks, fewer processing resources are available for extracting the phonetic content of each word, resulting in higher error rates and longer response times rather than low variability contexts.

Recall can also be defined as a process where information can be recollected or retrieved from storage at will. Specific encoding operations performed on what is perceived determine what retrieval cues are effective in producing access to what is stored (Tulving& Thomson, 1973). Serial recall is the ability to recall items or events in the order in which they occurred. Primacy and recency effects are generally seen in serial order recall, where primacy effects refer to better recall of items earlier in the sequence, while recency effects refer to better recall of the last few items.

Nygaard, Sommers and Pisoni (1992) reported that subjects recall words from lists produced at a single speaking rate was better than the same words produced at several different speaking rates. Interestingly, the differences appeared in the primacy portion of the serial position curve suggesting greater difficulty in the transfer of items into long-term memory.

Goldinger, Pisoni and Logan (1991) investigated the nature of talker variability effects on serial recall of spoken word lists by varying the rate of presentation of the items in the list to be recalled. Authors hypothesized that rate of presentation would selectively affect the listener's ability to encode the distinctive talker information for multiple-talker lists. If given enough rehearsal time, listeners might be able to use the distinctive talker information as an additional retrieval cue, and thus the multiple-talker lists would be more accurately recalled than the single-talker lists. Indeed, Goldinger et al. (1991) reported that at fast presentation rates (one word every250 msec), words in the primacy portion of the single-talker lists were more accurately recalled than those from multiple-talker lists, whereas at slow presentation rates (one word every 4,000 msec), this difference in recall accuracy was reversed. These results showed that information about a talker is encoded into long-term memory and can be used as an effective retrieval cue under optimal conditions. Goldinger (1996) reported that detailed talker information appeared to be retained in memory and used in perceptual identification (implicit task) and recognition memory (explicit task) tasks.

Experiments on serial recall of lists of spoken words by Martin, Mullennix, Pison& Summers (1989) and Goldinger et al. (1991) demonstrated that specific details of a talker's voice are also encoded into long-term memory. Using a continuous recognition memory procedure, Palmeri et al. (1993) found that detailed episodic information about a talker's voice is also encoded in memory and is available for explicit judgments even when a great deal of competition from other voices is present in the test sequence.

**Need for the study**

Listeners integrate auditory cues in perception of speech when communicating in both normal and compromised listening environments. Three factors affect the success of interaction in communication situations: characteristics of the talker, characteristics of the listener, and characteristics of the speech signal itself. In day to day life, individuals must comprehend speech produced by many different talkers (Hager, & Amanda, 2013). Also to determine whether perceptual effects of stimulus variation are due to time consuming normalization process or due to the increased encoding time for perceptual detail, studies need to be done to investigate the effects of stimulus variability on memory for spoken word recognition. This demonstrates that talker variability is an important factor in speech perception that must be incorporated into current conception of spoken word recognition. Published empirical studies in Indian context/literature is meager in this line where most of the languages are syllable based and it is very much tedious for listeners to recognize and memorize each phonemes of the word in ongoing speech on a noisy background for better speech perception.

**Aim of the study**

The present study was plannedto determine the effect of talkers’ variability on spoken word recognition and recall in adult neuro-typical individuals whose native language is Kannada.

**Objectives**

1. To evaluate the spoken word recognition and recall abilities under single and multi-talker variations.
2. To compare the spoken word recognition and recall abilities between males and females.
3. To find out the serial position effect using single and multi talkers as one of the variable for spoken word recognition and recall.

**Method**

***Participants***

A total of 30 neuro-typical adult individuals were considered for the study and are classified into two groups: Group 1 (15 males) and Group 2 (15 females).All the participants were within the age range of 18-21 years with mean age of 19.4 (males) and 19.3 (females).All the participants were native speakers of Kannada language and who reported no history of speech, language, hearing, cognitive and communication disorders at the time of testing.

***Stimulus material***

The test material consisted of two sets of word lists: List A and List B, respectively in which both had 15 tri-syllabic meaningful Kannada words each (see Appendix A). Words in both the lists were semantically unrelated to each other and are of different categories of words. List A was composed of a single talker and List B on other hand; it was prepared by multi talkers’ utterances.

Words in the first list (List A) were uttered by a male of 21 years of age whose native language is Kannada. His speech clarity was assessed by 3 SLP’s where, they rated his speech as 100% intelligible. Whereas, words in the second list (List B) were uttered by 15 different individuals (both males and females) age ranged between 18-23 years of native Kannada speakers. Similarly, the multi talker’s speech was assessed by 3 SLP’s and out of fifteen words only one word which is 100% intelligible was considered from each different individual’s utterance for preparation of stimuli for List B.

All these recordings were carried out in a sound treated room using CSL software (CSL 4500 Model, Kay Pentax, New Jersy, USA) and with shure microphone, where the input signal is digitized online via a 16 bits analog to digital convertor at 44.1 KHz sampling rate and was saved in .wav format. Both the List A and B recorded stimuli was edited to maintain a 3 seconds inter-word interval duration and the whole stimulus was superimposed with speech babble as a background noise using Aux Viewer Version 1.41 (Win 64) software. Both the signal words and speech babble noise were presented at 0 dB SNR. The final version of the recorded stimuli for the study thus, prepared including List A& List B was made.

***Procedure***

Presentation of stimuli

The aim and objective of the study was explained and took a written consent from the participants. The participants were made to seat comfortably in noise free room. The stimulus was controlled online using laptop and presented through head phones (ZEB-100HM) binaurally at a comfortable loudness (60dB SPL). Each word was presented only once with no possibility of repetition and the order of list presentation (List A and List B) was counterbalanced across listeners and 3 – 5 minutes break was given between the two lists presentation. After the experiment, participants were given a questionnaire that contained five5 questions (4 close ended and 1 open ended) to answer about their feedback on study related participation.

***Instruction***

Participants were instructed to listen to the words presented from the each list completely and to write the words immediately by recalling it. Upon completion of the task, participants were asked to complete a post experiment questionnaire by writing close-ended answers to the questions, majorly to check the experience and perception capacity of spoken word recognition and recall of the present task (see Appendix B).

***Statistical analysis***

Number of words recollected was measured and for each recollected word, a score of one was assigned and words which are not related to lists were not scored. Also words which are closed to the target were scored as one. Statistical analysis was carried out using SPSS software Version 20, to find out the statistical difference between lists (List A versus List B) and gender (Males versus Females) in spoken word recognition using Mixed ANOVA.

**Results**

To test the normality, Shapiro-Wilk test was carried out where it showed that the data fulfilled the normality assumptions. Hence, Mixed ANOVA was carried out to see the significant difference within subject effect (Lists), between subject effect (Gender) and interaction effect.

The mean (M) number of words recollected by the two groups (Group 1 and Group 2) across the two tasks of single and multiple talkers are obtained with respective standard deviation (SD) values are shown in table1. Descriptive statistics revealed that the mean number of words that were recollected was relatively more by participants for List A compared to list B. The same was observed in both males and females. The SD values are higher in List B than List A indicating more variability in performance by participants for multi-talker list. Also, females (Group 2) recollected more number of words in List A& B compared to males (Group 1). Further, SD values are higher in females in List A& B indicating wider variability in performance.

*Table 1: Mean and standard deviation of words recalled across lists & gender*

|  |  |  |
| --- | --- | --- |
| **No. of words recalled** | **Single Talker** (List A) | **Multi Taker** (List B) |
| **Group 1**(Males) | **Group 2**(Females) | **Group 1**(Males) | **Group 2**(Females) |
| Mean | 7.46 | 8.06 | 6.53 | 7.2 |
| SD | 0.91 | 1.86 | 1.72 | 1.89 |



*Figure 1: Mean scores of word recalled between two groups on List A& List B.*

Within subject comparison, there was a significant difference between the single talker (List A) and multi talker list (List B), F (1, 28) = 4.837, p < 0.05. The same is depicted in figure 1. Also, results of Mixed ANOVA revealed that there is no significant interaction effect [F (1, 28) = 0.07, p > 0.05] found for lists (tasks) and groups (gender). Table 2 shows the results of Mixed ANOVA. Even, between subject comparison, there was no statistically significant difference between Group 1 and 2 [F (1, 28) = 2.043, p > 0.05], that is, no gender difference found.

*Table 2: Results of Mixed ANOVAfor different varibles*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **df** | **F** | **sig.** |
| Lists  | 1 | 4.83 | 0.03 |
| Gender | 1 | 2.04 | 0.16 |
| Lists \* Gender | 1 | 0.07 | 0.93 |
| Error | 28 |  |  |

In analyzing the serial order position effect on word recall ability, most of the participants have increased percentage of word recall at the initial (primacy effect) and the final position (recency effect) compared to the middle position of words in both the lists (List A & B).Figure 2 shows the serial order positioning effect for males (a) and females (b) between list A and list B. From the figure 2, it can be noted that the retain ability of words were superior in the initial and final position of list of words when it was read by multi-talker rather than a singer talker. Both the primacy and recency effects were clearly noticed in group 2 (females) than group 1 (males).



2 (a)



2 (b)

*Figure 2: (a) Percentage of correctly recalled words for both single and multi talker lists as a function of serial order position effect in males and (b) in females.*

*(WRST – Word Recall in Single Talker List; WRMT – Word Recall in Multi Talker List)*

The post experimental questionnaire responses or feedbackobtained from participants were tabulated & analyzed. The obtained data was analyzed into percentages and the results revealed the following points;

66.66% of participants reported that List A was easier compared to List B [figure 3(a)].83.33% of participants reported words read by male voice was easier to recollect than female voice [figure 3(b)]. 60% of participants have mentioned that they require more time to recall words in List B compared to List A [figure 3(c)].93.33% of participants have mentioned that they can recall the word in the initial position of the list easily compared to other positions [figure 3(d)] and 26.66% of participants reported that they would be able to recall approximately 8 words from talker’s list which matches with the present study findings [figure 3(e)].

3(a)

3(b)

3(c) 

3(d)

3(e)

*Figure 3: (a) Percentage of words recollected easily between lists; (b) Percentage of word recollected based on voice; (c) percentage of time required for recalling of words; (d) Percentage of recall of word with respect to position in the list; (e) Percentage of perceived correct number of words recalled.*

**Discussion**

Spoken word recognition is one of the major components in speech perception which would be affected by various factors like talker’s variability, environmental parameters and perceptual recalling ability of an individual. The results of the present study found that participants were able to recognize and recall the words better from the single talkers list compared to the multi talkers list. That is, participants recalled significantly more words from list A than List B, which indicated that participants might formed relatively an acoustic vowel space for a single talker that influence them to encode, process, store and retrieve words more easily than forming multiple acoustic vowel spaces for different speakers as in case of List B (Multi talkers). The present study findings supports the findings of Mullennix et al. (1989) who reported that better word recognition for single talker than multi talkers. This is also correlating with the self feedback of the participants that 66.66% of them can recognize the words easily from List A compared to List B and found that 60% of participants needs longer time to recalwords from List B (multi-talker list).

Females (Group 2) in the present study recollected relatively more words than male participants (Group 1), albeit, the difference is not statistically significant. Hence, the present study did not find any gender difference in spoken word recognition and recollect task. The present findings is contradicting with the fMRI study of [Schirmer](https://scholar.google.co.in/citations?user=59IHakYAAAAJ&hl=en&oi=sra), Zysset, [Kotz](https://scholar.google.co.in/citations?user=2wVCGFUAAAAJ&hl=en&oi=sra) and Cramon (2004) who reported that the left inferior frontal gyrus has been repeatedly implicated in semantic processing and has evidence that semantic processing in women is more susceptible to influences from emotional prosody than semantic processing in men. More number of participants warrented in each of the group in future for gender related differences in spoken word recognition investigation.

The serial order position effect in the present study revealed that most of the participants were able to recall the initial and final words compared to the middle words in both the lists.Further, Group 1 showed that the percentage of initial and final words recall ability of the multiple talkers list was higher compared to the middle words. Apparently, the distinctive information provided by each of the different voices associated with the words in the list allowed listeners to remember both the word and its temporal position in the list. Participants were encoding distinctive talker information along with the linguistic content of the spoken words into a memory representation that could be used for serial recall positioning. These findings are in consonance with the study findings of Goldinger et al. (1991), Palmeri et al. (1993) and Nygaard et al. (1994) who reported that talker-specific information is retained in long-term memory and can be used not only to aid recognition memory but also to facilitate the subsequent perceptual analysis of the phonetic content of a talker's novel utterance.These distinct patterns of results for each stimulus dimension investigated in this study suggests that each source of variability encountered in the speech signal may engage different adaptive mechanisms in the perceptual system and may affect different levels of processing.

**Conclusion**

The present study found no gender difference in word recall or spoken word recognition task as a factor of talker variability. Also, the present study found spoken word recognition was better in single talker than multi talkers. This study added the evidence on to the literature of speech perception in Indian (Kannada) language/context. The output of the results can be generalized to understand, how speech perception is difficult in typical in terms of talkers’ variability.

Further studies can be planned in individuals with Hearing Impairment (HI) whether they are able to recognize and process the spoken words of multi talkers easily or do they require tunings of their speech perception or improving their meta-cognitive skills for better spoken word recognition and recall ability that’s needs to be explored.

Appendix A

List of tri-syllabic words used in the study (in Kannada)



Appendix B

POST-EXPERIMENT QUESTIONNAIRE



*Figure 1: Mean scores of word recalled between two groups on List A& List B.*



*Figure 2: (a) Percentage of correctly recalled words for both single and multi talker lists as a function of serial position effect in males and (b) in females.*

*(WRST – Word Recall in Single Talker List; WRMT – Word Recall in Multi Talker List).*

2 (a)

2(b)

*Figure 3: (a) Percentage of words recollected between lists; (b) Percentage of words recollected based on voice; (c) percentage of time required for recalling of words; (d) Percentage of recall of words with respect to position in the list; (e) Percentage of perceived correct number of words recalled.*

3(a)

3(b)

3(c) 

3(d)

3(e)

*Table 1: Mean and standard deviation of words recalled across lists & gender*

|  |  |  |
| --- | --- | --- |
| **No. of words recalled** | **Single Talker** (List A) | **Multi Taker** (List B) |
| **Group 1**(Males) | **Group 2**(Females) | **Group 1**(Males) | **Group 2**(Females) |
| Mean | 7.46 | 8.06 | 6.53 | 7.2 |
| SD | 0.91 | 1.86 | 1.72 | 1.89 |

*Table 2: Results of Mixed ANOVA for different variables*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **df** | **F** | **Sig.** |
| Lists  | 1 | 4.83 | 0.03 |
| Gender | 1 | 2.04 | 0.16 |
| Lists \* Gender | 1 | 0.07 | 0.93 |
| Error | 28 |  |  |