Temporal Jitter Mimics the Effects of Aging on Word Identification and Word Recall in Noise

Sasha Brown & M . Kathy Pichora-Fuller
School of Audiology and Speech Sciences, University of British Columbia

Difficulty understanding speech spoken in noise is a particular problem for older adults, even when they have no clinically significant elevation in pure-tone audiometric thresholds (CHABA, 1988). Because spectral models of hearing loss are inadequate to account for their particular difficulties (for a review see Schneider & Pichora-Fuller, 2000), researchers have devoted increasing effort to the investigation of the nature of behavioural and physiological declines in auditory temporal processing with age and the effects this could have on speech perception.

Physiological studies have yielded converging evidence suggesting that there is age-related loss of neural temporal synchrony at various levels of the auditory system (for a review see Schneider, 1997). Such a loss of synchrony has been implicated in age-related changes on a number of perceptual measures relevant to the extraction of temporal fine structure speech cues. Monaurally, loss of synchrony could explain why age-related increases in frequency difference limens (DL) are greater for low frequencies than for high frequencies (Abel, Krever, & Alberti, 1990). Because frequency DL is thought to depend on phase-locking at low frequencies, a loss of synchrony would have more of an effect at these frequencies. Binaurally, age-related changes in masking-level differences have been observed for both non-speech and speech signals and have been attributed to an increase in temporal jitter or a loss of temporal synchrony (Pichora-Fuller & Schneider, 1992). Thus, loss of synchrony in aging auditory systems may have wide-reaching perceptual consequences, including disruption of the fine structure cues important for understanding language spoken in noise.

Furthermore, such an asynchrony may affect cognitive processing beyond the level of word recognition. If the fidelity of the auditory signal is compromised in transmission up the auditory pathway, more cognitive resources may be allocated to interpreting the signal. It has been found that older adults are better able to use contextual cues in order to compensate for declines in perceptual acuity (Pichora-Fuller, Schneider, & Daneman, 1995). Such an allocation of resources, however, leaves fewer cognitive resources for other processes necessary to the integration and retention of spoken language. Thus, a perceptual difficulty can potentially cascade into a reduction in working memory capacity, which can be manifest in problems recalling speech even if it is correctly perceived (Pichora-Fuller, et al., 1995).

The current studies attempt to simulate auditory aging by introducing an asynchrony to the speech signal and presenting it to young adults with normal hearing ability[1]. The results were then compared to those obtained in an earlier study (Pichora-Fuller et al., 1995) in which older adults with normal pure-tone thresholds listened to intact stimuli. The first experiment focused on word identification performance in four different S/Ns, while the second added a memory task in order to determine if the increased perceptual difficulties would impact working memory span.

EXPERIMENT 1.

Methods:

Participants. The participants were twelve young adult paid volunteers (mean age = 24.7 years, S.D. ± 3.2), with clinically normal hearing (pure-tone thresholds from 250 to 8,000 Hz ≤ 20 dB HL). All participants provided informed consent and their rights as participants were protected.

Stimuli and Apparatus. In this study the effects of temporally jittering the sentences of the Revised Speech Perception in Noise test (SPIN-R; Bilger, Neutzel, Rabinowitz, & Rzeczkowski, 1984) were investigated. The SPIN-R test consists of 8 lists of 50 sentences. The listener's task is to report the last word of the sentence immediately following its presentation. In half of the sentences in each list, the last word is predictable from the sentence context (e.g. *The wedding banquet was a feast.*) and in the other half it is not predictable (e.g. *We could consider the feast.*). Performance with low-context sentences is thus meant to assess ability when only auditory cues are present, while an improvement in performance with contextual cues would indicate the participant's ability to deploy cognitive resources to rescue a degraded signal.

The SPIN-R sentences were presented both intact and jittered. Visualizing the speech signal as a graph portraying changes in amplitude over time, jitter can be seen to cause slight alterations in the timing of each point. The program used allowed the experimenter to determine both the range of changes and the rate at which these changes occur. Within these parameters, the alterations occur based on the Gaussian distribution of band-limited white-noise. For each sample in the sound file (20 kHz sampling rate), a delay value is selected by referring to the distribution of noise, determining the amplitude of the noise at the corresponding point in time and the using this amplitude as a delay value. The delay value then determines the position of the time sample in the original speech waveform whose amplitude value is to be substituted in for the value at the time position in question. The chosen bandwitdth of the low-passed noise determines the rate at which the delay values will change and the standard deviation alters the amplitude of the noise signal thereby increasing the range of delay values that can be selected. To create the low-frequency jittered stimuli a Fast Fourier Transform was used to separate the speech signal into its component frequencies. The signal was then divided into two bands, one above and one below 1,200 Hz. Each band was converted back to the time domain using an Inverse Fast Fourier Transform. The lower band was jittered based on a LP noise with a 500-Hz bandwidth and .25 msec standard deviation; the upper band was not jittered. The two bands were then recombined.

The speech signal remained at 70dB HL, and the babble was adjusted to create the S/N conditions presented to each participant through TDH-39 10W earphones in a sound attenuating IAC booth. The experimenter controlled the SPIN-R form number, jitter condition, S/N and presentation level of the stimuli through an in-house computer program on an IBM-compatible personal computer. The digital signals were routed through the Tucker Davis Technologies DDI, FT5, PA4, SM3 and HB5 modules before reaching the listener's earphones.

Procedure. Participants were required to repeat the sentence-final word from each sentence. All participants completed the lists in a fixed order during two sessions each lasting about one hour. In session one, they heard intact sentences at +8 and then at +4 dB S/N, followed by jittered sentences at +8 and then at +4 dB S/N. In the second session, they heard intact sentences at 0 and –4 dB S/N, and then jittered sentences at 0 and –4 dB S/N. The order of SPIN-R lists differed for each subject; the first four SPIN-R lists were counter-balanced over the +8 and +4 dB S/N presentations, and the last four SPIN-R lists were counter-balanced over the 0 and –4 dB S/N presentations.

Results and Discussion:

Analyses of variance confirmed that, consistent with previous literature (e.g., Pichora-Fuller et al., 1995), the participants in the study were better able to correctly identify the sentence-final words when the S/N was higher ($F(3,33)=424.1$, $p<0.001$), and when the context of the sentences provided cues ($F(1,11)=416.8$, $p<0.001$). When jitter was introduced to the speech signal, performance was

Figure 1
Mean Percent-Correct Word-Identification Scores for Young Participants with Jittered Stimuli and Old Participants with Intact Stimuli with High- and Low-Context Sentences

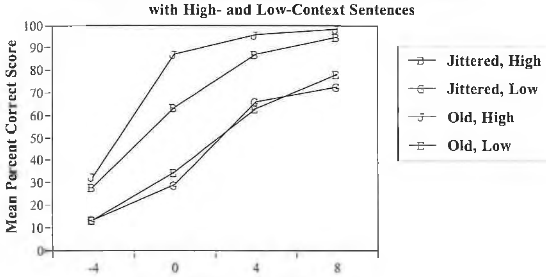
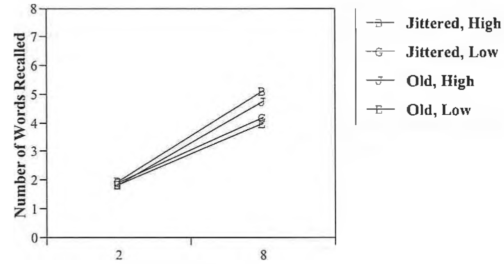


Figure 2
Number of Words Correctly Recalled for Each Recall Set Size: Young Participants with Jittered Stimuli at S/N=4 dB, and Old Participants with Intact Stimuli at S/N=5 dB with High- and Low-Context Sentences



significantly effected ($F(1,11)=262.5$, $p<0.001$). There were also interaction effects between jitter and S/N ($F(3,33)=6.7$, $p<0.001$) and S/N and context ($F(3,33)=20.6$, $p<0.001$). No other interactions were found to be significant. It is thus apparent that the introduction of a temporal asynchrony does have a deleterious effect on word identification. The results are also remarkably similar to those obtained from old listeners with normal pure-tone thresholds in the Pichora-Fuller et al. (1995) study. Figure 1 displays the mean percent-correct sentence-final word-identification scores for jittered sentences in this experiment as well as for the elderly listeners in the 1995 study so that the results can be easily compared. For low context sentences (empty symbols), the resemblance is immediately striking. The young subjects in the current study were able to identify close to the same number of sentence-final words when presented with jittered sentences as were the older adults with intact sentences. The jitter manipulation thus appears to mimic the effect of age in word identification, at least when contextual cues are minimal.

Recall that performance on the low-context SPIN-R sentences is believed to represent auditory skill, whereas increased performance on high-context sentences represents the listener's ability to make use of contextual cues. Pichora-Fuller et al. (1995) found that older adults were better able to make use of context than were younger adults, especially as the listening conditions became less favourable. In Figure 1, it can be seen that a performance difference occurs when the S/N is moderately difficult (+4 dB and 0 dB). These listening conditions likely mimic the more difficult ones encountered in every-day life, in which the older adults have more experience relying on contextual cues to rescue degraded signals.

EXPERIMENT 2.

Methods:

Participants. The participants in this experiment were sixteen young adult paid volunteers (mean age = 26.8 years, S.D. $\pm$ 1.78), whose hearing was clinically normal (pure-tone thresholds from 250 to 8,000 Hz $\leq$ 20 dB HL). All participants provided informed consent and their rights as participants were protected.

Stimuli and Apparatus. As with Experiment 1 a temporal asynchrony was applied to the eight forms of the SPIN-R test, and the participants listened to both intact and jittered sets of each. Again, the jitter was applied only to the frequency components below 1200 Hz, but in this experiment the S/N conditions were restricted to +8 and +4 dB. The signal was presented in the same manner outlined for Experiment 1.

Procedure. Again listeners were required to repeat the sentence-final word. In this experiment, however, participants were also instructed to judge the predictability of the sentence final word in order to ensure that they attempted to comprehend the entire sentence rather than attending to only the sound of the last word. Participants were also required to recall as many words as possible from the just completed set of either two or eight sentences. The set sizes remained constant for a given SPIN-R form and the listener was told ahead of time whether they would be required to remember words from a set of two or eight.

The first half of the experiment consisted of intact SPIN-R forms, at both S/Ns and both recall set sizes, progressing from most easy to most difficult condition. The second half consisted of the same order of conditions, but the sentences were jittered. The forms were balanced across conditions and participants.

Results and Discussion:

Performance on the word identification task was similar to that obtained in Experiment 1 for the same S/N conditions; when the sentence-final word-identification scores for Experiment 1 and Experiment 2 were compared, it was found that there was no significant difference between the participant groups ($F(1,26)=1.8$, $p>0.10$). Again analyses of variance confirmed that main effects were found for jitter ($F(1,15) = 86.4$, $p<0.001$), S/N ($F(1,15) = 70.2$, $p<0.001$), and context ($F(1,15) = 330.5$, $p<0.001$) in Experiment 2. Recall set-size had no main effect ($F(1,15)=0.16$, $p>0.1$), suggesting that an increased memory load does not effect perceptual processing. Significant interaction effects were found for jitter x context ($F(1,15)=7.1$, $p<0.05$), and S/N x context ($F(1,15)=24.3$, $p<0.001$). No other interactions were found to be significant.

Participants were also required to indicate whether they thought the sentence-final word was predictable from the context of the sentence. Mean scores on this task remained above 94%, suggesting that the listeners were comprehending the entire sentence rather than simply attending to the sentence-final word.

When the young listeners were required to recall the sentence final words, it was found that there were main effects for jitter ($F(1,15)=43.72$, $p<0.001$), recall set-size ($F(1,15)=286.7$, $p<0.001$) and context ($F(1,15)=53.6$, $p<0.001$). The young listeners in the current study performed very similarly to those in the study of Pichora-Fuller et al. (1995)[2], and when the sentences were jittered, the young subjects' performance was reminiscent to that of the elderly listeners when presented with intact stimuli. Figure 2 compares the number of words correctly recalled for the young people in the current study when presented with jittered stimuli in a S/N of 4 dB and those for elderly listeners with intact stimuli at a S/N of 5 dB.

It appears that the introduction of jitter decreased word-recall ability in young subjects in a manner similar to the decreased ability in older adults, thereby supporting the hypothesis that external jitter resembles the internal jitter inherent in the aged auditory system, and results in similar processing demands. Presumably for both older adults with intact stimuli and younger adults with jittered stimuli, fewer cognitive resources are left for remembering. The fact that a slight difference in recall performance is noted between elderly adults and young adults with jittered stimuli suggests that although the increased resources allocated to the perceptual channel may explain a large part of the word-recall deficits, something else further contributes to the noted recall difficulties. This is in keeping with a large body of research that proposes that the memory difficulties experienced by older adults have a cognitive component and that an age difference is noted more often as task complexity increases (e.g., Obler, Fein, Nicholas, & Albert, 1991). The performance difference between young adults listening to jittered stimuli and older adults listening to intact stimuli occurred when the memory load was high (recall set-size of 8), and when the listening conditions were fairly difficult (S/N of 4 to 5 dB). It is thus not surprising that a slight difference occurred and indicates that perceptual and cognitive factors cannot be considered in isolation, particularly in the aged population.

Although it appears that both perceptual and cognitive deficiencies contribute to the noted word-recall difficulties in older adults, the findings of the current study and those of Pichora-Fuller et al. (1995) suggest that perceptual processing receives priority in resource allocation. In both of these experiments, word identification performance was not affected by an increased memory load, even though increased perceptual stress reduced recall. From the perspective of an information-processing model, it makes sense that priority is given to perceptual processing; preserving resources for memory and other cognitive processes would be of little benefit if the cost were a loss or reduction in quality of incoming information

General Discussion:

This simulation of the neural jitter that is thought to disrupt the ability to phase-lock to lower frequencies in the aging auditory system was successful. The applied jitter was determined to affect young adults' performance such that it resembles that of older adults in two different tasks. Such a finding helps explain why there tends to be an age-related decline in the ability to perceive speech, particularly in the presence of background noise, when no hearing loss is evidenced by standard clinical pure-tone audiometry. The findings also help explain some aspects of working-memory problems experienced when information is presented auditorily under challenging listening conditions

Furthermore, this series of experiment provides evidence for the inter-relationship between perceptual (bottom-up), and cognitive (top-down) channels. It was demonstrated that degraded perceptual processing affects processes, such as memory, which are traditionally believed to be cognitive. Cognitive processes, such as the use of available semantic cues to help one predict content, were also shown to be employed to rescue signals degraded at the perceptual level. It is thus evident that as people age numerous factors interact and affect the way speech is heard, interpreted and remembered.

Abel, S.M., Krever, E.M., & Alberti, P.W. (1990), *Scand. Aud, 19*, 43-54;
CHABA (1988), *JAAA, 7,* 190-202;
Obler, I.K., Fein, D., Nicholas, M., & Albert, M.L (1991), *Applied Psycholinguistics, 12,* 433-452;
Pass, H. (1998) M.Sc. Thesis, UBC;
Pichora-Fuller, M.K., & Schneider, B.A. (1992), *JASA, 91*, 2129-2135;
Pichora-Fuller, M.K., Schneider, B.A., & Daneman, M. (1995), *JASA, 97,* 593-607;
Schneider, B.A., (1997), *JSLPA, 21*, 111-124;
Schneider, B.A., Pichora-Fuller, M.K., (2000), In F.I.M Craik & T.A. Salthouse (Eds) Handbook of Aging and Cognition, pp 155-219.

[1] A previous study (Pass, 1998) has already applied jitter to the SPIN-R sentences, but the means of application resulted in spectral splatter that masked the high frequencies of the speech signal, making it difficult to determine the relative contributions of the jitter and the masking. The present study refined this jitter by applying it only to the frequencies below 1200 Hz.

[2] Because different S/N conditions and recall set sizes were used in the 1995 experiment, the findings cannot be directly compared, although it is still possible to discuss the trends.