Available online at www.sciencedirect.com
ScienceDirect
Hearing Research 223 (2007) 114–121
Hearing Research
www.elsevier.com/locate/heares

Research paper

Temporal jitter disrupts speech intelligibility: A simulation of auditory aging

M. Kathleen Pichora-Fuller [a,*], Bruce A. Schneider [a], Ewen MacDonald [b], Hollis E. Pass [c], Sasha Brown [c]

[a] Department of Psychology, University of Toronto, 3359 Mississauga Rd., Mississauga, Ont., Canada L5L 1C6
[b] Institute for Biomaterials and Biomedical Engineering, University of Toronto, 164 College St., Toronto, Ont., Canada M5S 3G9
[c] School of Audiology and Speech Sciences, University of British Columbia, 5804 Fairview Ave., Vancouver, B.C., Canada V6T 1Z3

Received 7 November 2005; received in revised form 19 October 2006; accepted 30 October 2006
Available online 8 December 2006

Abstract

We disrupted periodicity cues by temporally jittering the speech signal to explore how such distortion might affect word identification. Jittering distorts the fine structure of the speech signal with negligible alteration of either its long-term spectral or amplitude envelope characteristics. In Experiment 1, word identification in noise was significantly reduced in young, normal-hearing adults when sentences were temporally jittered at frequencies below 1.2 kHz. The accuracy of the younger adults in identifying jittered speech in noise was similar to that found previously for older adults with good audiograms when they listened to intact speech in noise. In Experiment 2, to rule out the possibility that the reductions in word identification were due to spectral distortion, we also tested a simulation of cochlear hearing loss that produced spectral distortion equivalent to that produced by jittering, but this simulation had significantly less temporal distortion than was produced by jittering. There was no significant reduction in the accuracy of word identification when only the frequency region below 1.2 kHz was spectrally distorted. Hence, it is the temporal distortion rather than the spectral distortion of the low-frequency components that disrupts word identification.

Keywords: Aging; Temporal jitter; Neural synchrony; Speech intelligibility; Word identification

1. Introduction

Some older adults experience greater difficulties than their younger counterparts in understanding language spoken in noise, even when they have no clinically significant elevation in pure-tone audiometric thresholds in the speech range (e.g., Pichora-Fuller et al., 1995). Spectral models of hearing loss are inadequate to account for their particular difficulties (for reviews see Divenyi and Simon, 1999; Pichora-Fuller and Souza, 2003; Schneider and Pichora-Fuller, 2000). In search of other possible explanations, researchers have devoted increasing effort to the investigation of the nature of behavioural and physiological declines in auditory temporal processing with age (e.g., Fitzgibbons and Gordon-Salant, 1996; Frisina and Frisina, 1997; Frisina et al., 2001). One such decline, age-related loss of neural synchrony, is believed to be independent from other age-related declines which are associated with outer hair cell damage or changes in endocochlear potential (Boettcher et al., 1996; Gates et al., 2003; Mills et al., in press).

Abbreviations: ANOVA, analysis of variance; BW, bandwidth; dB HL, decibels hearing level; dB S:N, decibels signal-to-noise ratio; dB SPL, decibels sound pressure level; D-to-A, digital to analog; DC, direct current; DL, difference limen; FFT, fast Fourier transform; Hz, Hertz; IFFT, inverse fast Fourier transform; kHz, kilohertz; M, mean; ms, millisecond; RMS, root mean square; SD, standard deviation; S:N, signal-to-noise ratio; SNK, Student–Newman–Keuls test of multiple comparisons; SPIN-R, speech perception in noise test – revised; STEP, spectro-temporal excitation pattern

* Corresponding author. Tel.: +1 905 828 3865; fax: +1 905 569 4326.
E-mail address: kpfuller@utm.utoronto.ca (M.K. Pichora-Fuller).

0378-5955/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2006.10.009

There is converging evidence from physiological studies suggesting that an age-related loss of neural temporal synchrony can occur at various levels of the auditory system (for reviews see Frisina et al., 2001; Schneider, 1997). Age-related decreases in compound action potentials have been observed in quiet-reared gerbils (Hellstrom and Schmiedt, 1990). Age-related neuro-chemical changes and changes in the latency and definition of waves IV and V of the auditory brainstem response consistent with a decrease in neural synchrony have been observed in quiet raised rats (Backoff and Caspary, 1994; Raza et al., 1994).

Age-related loss of temporal synchrony could also explain why the performance of older adults is poorer than that of younger adults on a number of non-speech and speech tasks that are dependent on the extraction of temporal fine structure from the acoustic stimulus. Monaurally, loss of synchrony could explain why age-related increases in frequency difference limens (DL) are greater for low frequencies than for high frequencies (e.g., Abel et al., 1990; He et al., 1998; Moore and Peters, 1992). Because the frequency DL is thought to depend on phase-locking at low frequencies, a loss of synchrony would differentially affect DLs for low-frequency signals. Loss of synchrony might also contribute to age-related declines in the ability to detect a mistuned harmonic (Alain et al., 2001) or to segregate concurrent vowels (e.g., Snyder and Alain, 2005; Summers and Leek, 1998; Vongpaisal and Pichora-Fuller, in press). Binaurally, age-related changes in masking-level differences have also been observed for both non-speech and speech signals (Grose, 1996; Pichora-Fuller and Schneider, 1991, 1992, 1998), and these changes have been attributed to an age-related increase in temporal jitter or a loss of temporal synchrony (Pichora-Fuller and Schneider, 1992). Thus, there is psychoacoustic as well as physiological evidence pointing to a loss of synchrony in aging auditory systems.

Given the observations that older adults have difficulty understanding speech in noise and that a loss of synchrony may be a feature of auditory aging, it is tempting to consider whether or not loss of synchrony may contribute to age-related difficulties in understanding speech in noise. The aim of the present study is to investigate the hypothesis that disruptions in periodicity result in disruptions in speech intelligibility. Specifically, we simulate a loss of synchrony in younger adults by modifying the fine structure of the acoustic signal so that it becomes less periodic, and we investigate the consequences of this reduction in periodicity for identifying words spoken in noise. As far as we know, the effects of distorting the signal in this way have not been studied previously.

Loss of neural synchrony can be modeled as an increase in temporal jitter (e.g., Durlach, 1972; Pichora-Fuller and Schneider, 1992). Consider that without jittering the signal, each period of a 500-Hz tone would be exactly 2 ms, such that if the neural response were precisely phase-locked to this stimulus, spikes would always occur exactly at 2 ms intervals. However, in healthy younger listeners, phase-locking is not perfectly precise. Inter-spike intervals to a 500-Hz tone will have a mean value of 2 ms, but inter-spike intervals will be distributed around this mean. Therefore, we can model the temporal jitter that occurs in healthy younger auditory systems as a normal distribution of inter-spike intervals with a relatively small standard deviation, and simulate a loss of synchrony by increasing the standard deviation of this distribution. To simulate reduced phase-locking, we introduced delays in the signal where the delays vary randomly with time according to a normal distribution.

Specifically, if $x(t)$ is the intact waveform, the jittered waveform, $y(t)$, is a time-delayed version of $x(t)$ in which the time delay, δ, varies over time such that $y(t) = x[t - \delta(t)]$. There are two factors contributing to the jitter. The first is the distribution of delays that might occur over a long period of time. The second is the rate at which the delay changes with time. So that the delay, $\delta(t)$, would have a random normal distribution, we make $\delta(t)$ proportional to the amplitude of a low-pass noise (the amplitude of a noise is a normally distributed random variable), with the root mean square (RMS) amplitude of the noise being equal to the standard deviation of the distribution of delays. To vary the rate at which the delay changes with time, we vary the high-frequency cut-off of the noise, with a lower cut-off producing a slower rate of change. The bandwidth (BW) of the noise is thus determined by the high-frequency cut-off. In earlier pilot experiments we found that intelligibility is severely reduced by a small amount of jitter (RMS of 0.25 ms and BW of 500 Hz, Pichora-Fuller et al., 1999). Importantly, there is no noticeable change in the level nor the envelope of the speech signal. Note that this conceptualization focuses on loss of synchrony in terms of the loss of periodicity of firing by a fiber across time rather than as the loss of synchronization of firing across fibers at a given time.

1.1. Spectral distortion

If loss of synchrony is solely of neural origin, then events prior to spike generation would be unaffected. Specifically, loss of neural synchrony would have no effect on the pattern of activation along the basilar membrane. However, jittering the stimulus produces spectral splatter that will broaden the pattern of activation on the basilar membrane. Therefore, an unwanted effect of jittering the speech signal is to spectrally smear the energy so that spectral features such as formants may become obscured, especially at high frequencies in the most extreme jitter conditions. Thus, even though the temporal envelope of the waveform is largely preserved, the disturbance of the temporal fine structure has both temporal and spectral consequences, and the adverse effects of the jitter simulations on word identification that were observed in the pilot studies might have been due to spectral changes rather than, or in addition to, temporal disruptions of the fine structure.

To control for the amount of spectral splatter in the present study, we restricted the application of jitter to a low-frequency band and we also used a control condition in which there was a similar amount of spectral splatter but negligible temporal distortion. By altering only the components below 1.2 kHz, we focused the simulation on a frequency region where phase-locking and synchrony coding are known to occur and where we could most effectively minimize the potentially audible high-frequency spectral splatter. To achieve the control condition, we modified the smearing procedure developed by Moore and colleagues to simulate the loss of frequency selectivity associated with broadening of auditory filters in cochlear hearing loss (Baer and Moore, 1993, 1994; details are provided in Appendix A) so that the control condition had the same spectral distortion as the jitter condition, but negligible temporal distortion. The advantage of using the procedure of Baer and Moore (1993) to smear the spectrum is that it was specifically designed to produce spectral smear, but it preserves periodicity. To ensure comparability in spectral distortion, we restricted the smearing to the same frequency region for both the jitter and smear conditions. Thus, the two conditions differ markedly only in the degree of temporal or phase distortion that was produced.

Spectrograms provide a standard method of estimating the spectro-temporal energy distribution of a signal, but spectro-temporal excitation patterns (STEPs) provide an alternative method to estimate spectro-temporal energy distribution based on properties of the human auditory system (e.g., Moore, 2003a,b). In an attempt to quantify the degree of spectral distortion introduced by smearing or jittering, a pixel-to-pixel comparison was conducted using both spectrograms and STEPs. The absolute value of the pixel-to-pixel difference in power between the spectrograms for the smeared and intact versions of a sentence was calculated. By averaging over time we arrived at the mean absolute power difference between the smeared and intact versions of a sentence as a function of frequency. We then averaged over frequency to arrive at a single value to quantify the spectral distortion introduced by smearing. Likewise spectral distortion introduced by jittering was quantified. The comparisons were repeated using STEPs instead of spectrograms.

The spectral distortion due to jittering and smearing was quantified as described above using both spectrograms and STEPs for all 400 sentences of the Revised Speech Perception in Noise test (SPIN-R; Bilger et al., 1984). The ratio of the spectral distortion for jittering vs. smearing was found to be 0.23 dB using spectrograms, and 0.02 dB using STEPs. This ratio is close to 1 and implies that the degree of spectral distortion was very similar for both the jitter and smear simulations.

1.2. Limitations of the jitter simulation of temporal asynchrony

The application of jitter to a stimulus is a simple first approximation to a loss of neural synchrony. In an abnormal listener who exhibits a loss of synchrony, we would expect the inter-spike intervals to be more variable over time both within and across fibers. When a normal listener in our simulation is presented with a jittered 500-Hz pure tone, there will be a greater degree of variability in inter-spike intervals in fibers tuned to that characteristic frequency than there would be when an unjittered 500-Hz tone is presented. In addition, in an abnormal listener who exhibits a loss of synchrony, the correlation of spikes across fibers at any given time would be lower than in a normal listener. Note that the across fiber aspect of the pathology is not captured in our simulation. Hence, the perceptual consequences of a pathological loss of synchrony in a listener might be greater than the perceptual consequences to someone with a normal auditory system listening to the jittered stimuli produced using our simulation.

2. Experiments

Our aim was to test the hypothesis that disruptions in periodicity can result in disruptions in speech intelligibility and to examine if these simulated disruptions in speech intelligibility are similar to those found previously for older adults with good audiograms. Experiment 1 was conducted to evaluate the effects of jittering on word identification over a wide range of S:N conditions and to compare the results with the results found previously for older adults with good audiograms (Pichora-Fuller et al., 1995). Experiment 2 was conducted to rule out the possibility that the reductions in word identification observed in Experiment 1 were due to spectral distortion. In Experiment 2 we tested a smear simulation that produced spectral distortion equivalent to that of the jitter simulation in the frequency range below 1.2 kHz, but with significantly less temporal distortion.

2.1. Experiment 1: Word identification for intact and low-frequency jittered sentences

In Experiment 1, SPIN-R sentences were presented in the same range of S:N conditions (+8, +4, 0, −4 dB S:N) used in a previous study of age-related differences in word identification (Pichora-Fuller et al., 1995). There were two synchrony conditions: one intact condition, and one temporally jittered condition. In the temporally jittered condition, only the components below 1.2 kHz were jittered. The parameters used to create the jitter (0.25 ms average delay and 500 Hz BW) were selected based on earlier pilot testing (Pichora-Fuller et al., 1999).

2.1.1. Method

2.1.1.1. Participants. The participants were twelve younger adult paid volunteers (mean age = 24.7 years, SD = 3.2), who spoke English as their first language and whose hearing was clinically normal (pure-tone air-conducted thresholds from 250 to 8000 Hz $\leqslant$ 20 dB HL). All participants provided informed consent and their rights as participants were protected. The protocol for the study was approved by the institutional ethics review board.

2.1.1.2. Apparatus and stimuli. For the intact condition, the sentences of the SPIN-R were used. The SPIN-R test consists of 8 standardized, recorded lists of 50 sentences on one channel and accompanying multi-talker babble on a second channel. The sentences and background babble of the SPIN-R test were digitized at a rate of 20 kHz, and presented monaurally (right ear) over TDH-39 P earphones using a Tucker-Davis Technologies System II for D-to-A conversion and to control sentence and background levels. Sentences were presented at 70 dB SPL and the babble level was adjusted to achieve the four S:N conditions of +8, +4, 0 and −4 dB.

In the jitter condition, only the low-frequency components of the SPIN-R speech sentences were jittered. To create the jittered stimuli, a Fast Fourier Transform (FFT) was used to separate the speech signal into its component frequencies. The signal was divided into two bands, one above and one below 1.2 kHz. Each band was converted back to the time domain using an Inverse Fast Fourier Transform (IFFT). The lower band was jittered (RMS = 0.25 ms and BW = 500 Hz), but the upper band was not jittered. Finally, the two bands were recombined. The average RMS amplitudes of the intact and jittered speech signals were essentially the same, differing only by approximately 0.1 dB.

2.1.1.3. Procedure. The listener was asked 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.*) Each listener was tested at all four S:N levels with both the intact and jittered speech. All participants completed the lists in a fixed order during two sessions, with each session 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 session two, 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 participant; 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.

2.1.2. Results

Word identification was significantly better in more favorable S:N conditions and it was better in high-context than in low-context sentences (Fig. 1). Importantly, word identification was significantly reduced when the low-frequency components of the speech were jittered. Furthermore, jittering had a more deleterious effect on word identification when the S:N condition was less favorable

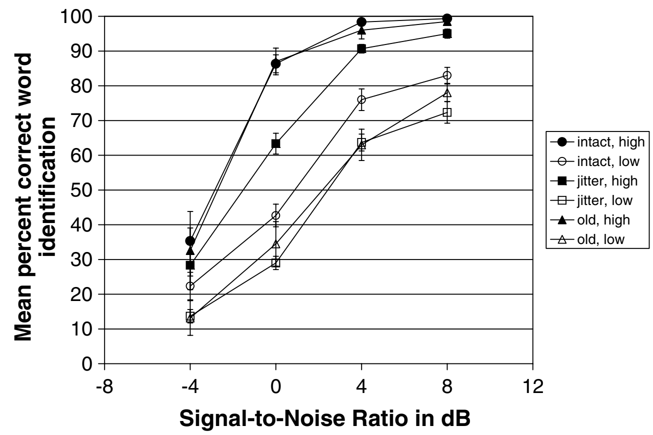

Fig. 1. Mean percent correct word identification scores for intact (circles) and low frequency jittered speech (square) at −4, 0, +4 and +8 dB S:N for high- (filled symbols) and low context (open symbols) SPIN-R sentences in younger listeners tested in the present study. For the jitter condition, in the frequency range from 0 to 1.2 kHz, the BW and RMS of the temporal jitter were 500 Hz and 0.25 ms, respectively. Also plotted are the means of the percent correct word identification scores of older adults for intact speech (triangles) in the same S:N and context conditions found in a previous study (Pichora-Fuller et al., 1995). Standard error bars are shown.

and the effect of S:N condition was more pronounced when context was low. This description was confirmed by an ANOVA with S:N (−4, 0, +4, +8), context (high, low), and jitter condition (intact, jittered) as with-subjects factors. There were significant main effects of S:N condition, $F(3,33) = 434.61$, $p < 0.001$, context, $F(1,11) = 416.48$, $p < 0.001$, and jitter condition, $F(1,11) = 295.78$, $p < 0.001$, as well as significant two-way interactions between S:N × jitter condition, $F(3,33) = 6.34$, $p < 0.001$ and between S:N condition × context, $F(3,33) = 19.74$, $p < 0.001$. The description was further confirmed using Student–Newman–Keuls (SNK) tests of multiple comparisons ($p < 0.05$).

Fig. 1 shows that jittering only the low-frequency components had a disruptive effect on word identification. Furthermore, for words in low-context sentences, the performance of younger adults in the jitter condition was essentially identical to the results obtained for older adult listeners with good audiograms in a previous study (Pichora-Fuller et al., 1995). A one-way ANOVA, ($F(2,29) = 7.44$, $p = 0.002$), and a follow-up SNK test of multiple comparisons ($p < 0.05$) confirmed that the speech reception threshold (dB S:N at which a score of 50% was achieved) for the younger listeners in the present experiment in the low-context intact conditions ($M = .83$ dB, SD = 1.13) was significantly better than their speech reception threshold in the low-context jitter conditions ($M = 2.51$, SD = .68), but that the performance of the younger listeners in the jitter conditions was not significantly different from the results obtained previously for older adults in low-context intact conditions ($M = 2.32$, SD = 1.65).

2.1.3. Discussion

These findings show that temporal jittering is deleterious to word identification. Furthermore, our jitter simulation is successful in mimicking the age-related problems in word identification in noise when sentence context is low insofar as the speech reception thresholds of the younger listeners in the jitter conditions did not differ from those of older listeners in intact conditions. Interestingly, for words in high-context sentences, the younger adults in the jitter condition perform worse than older adults when listening was effortful in the 0 and +4 dB S:N conditions. These listening conditions are like many everyday listening conditions in which older listeners would have regular need to rely on contextual cues to identify words. The superior performance of the older adults in these particular conditions may be due to better use of knowledge during word identification and/or more practice using context to compensate for difficulty perceiving the signal. This pattern of age-related differences is consistent with prior findings that older listeners benefit more than younger listeners from sentence context (e.g., Frisina and Frisina, 1997; Pichora-Fuller et al., 1995; Wingfield, 1996; but see Dubno et al., 2000).

2.2. Experiment 2: Word identification for temporally jittered or spectrally smeared low-frequencies

In Experiment 2, word identification accuracy on the SPIN-R test was measured using spectrally smeared sentences, intact sentences, and sentences jittered by the same method used in Experiment 1. If spectral distortion is responsible for the word identification difficulties found in Experiment 1 then word identification should be the same for both the jitter and smear simulations. On the other hand, if temporal distortion is responsible for the decrements in speech intelligibility then word identification should be more reduced for the jitter than for the smear simulation.

2.2.1. Method

2.2.1.1. Participants. The participants were sixteen younger adult paid volunteers (mean age = 20.7 years, SD = 3.0), who spoke English from the age of five years and whose hearing was clinically normal (pure-tone air-conduction thresholds from .25 to 8 kHz < 20 dB HL). All participants provided informed consent and their rights as participants were protected. The protocol for the study was approved by the institutional ethics review board.

2.2.1.2. Apparatus and stimuli. As in Experiment 1, the unaltered SPIN-R sentences and background babble were used in the intact condition. In Experiment 2, the babble was adjusted to achieve two of the same conditions tested in Experiment 1 (+8 and 0 dB S:N).

The jitter condition was identical to that used in Experiment 1. For the smeared condition, the SPIN-R sentences were smeared up to 1.2 kHz using a smearing factor of 6. The smeared sentences were then rescaled to have the same

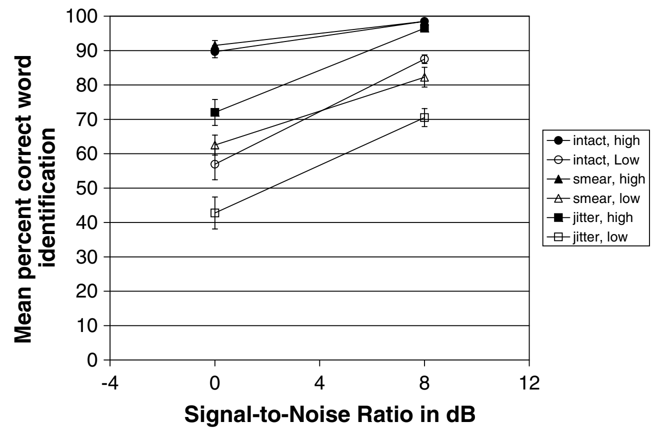

Fig. 2. Mean percent word identification scores of younger listeners in conditions of intact speech (circles), low-frequency smeared speech with a smearing factor = 6 applied from 0–1.2 kHz (squares), and low-frequency jittered speech with jitter parameters of RMS = 0.25 ms and BW = 500 Hz applied from 0–1.2 kHz (triangles), at +8 and 0 dB S:N for high-context (filled symbols) and low-context (open symbols) SPIN-R sentences. The speech signal above 1.2 kHz was not jittered or smeared. Standard error bars are shown.

average RMS amplitudes as the intact and jittered sentences.

2.2.1.3. Procedure. Each listener was tested in two S:N conditions with the intact speech and two types of speech simulation: jittered speech and smeared speech. All participants completed three sessions, with each session lasting about one hour. In session one, they heard intact sentences at +8 and then at 0 dB S:N. In session two, half of the participants heard the jittered sentences and the other half heard the smeared sentences at +8 and then at 0 dB S:N. In session three, the participants heard the other type of simulated sentences at +8 and then at 0 dB S:N. The order of SPIN-R lists differed for each participant. Sentences were presented over TDH-49 P headphones at 50 dB above the participant's babble threshold. Otherwise the procedure was the same as in Experiment 1.

2.2.2. Results

As before, word identification was significantly better in high-context than in low-context sentences, and it was better in more favorable S:N conditions (Fig. 2). Again, jittering had a more deleterious effect on word identification when the S:N condition was less favorable, and the effect of S:N condition was more pronounced when context was low. Importantly, word identification was significantly reduced when the low-frequency components of the speech were jittered, but performance was not significantly reduced when speech was spectrally smeared. This description was confirmed by an ANOVA with S:N (0, +8), context (high, low), and simulation condition (intact, jitter, smear) as with-subjects factors. There were significant main effects of context, $F(1,14) = 373.35$, $p < 0.0001$, S:N condition, $F(1,14) = 223.93$, $p < 0.0001$, and simulation condition, $F(2,28) = 21.77$, $p < 0.0001$. There were also significant two-way interactions between S:N condition × simulation condition, $F(2,28) = 4.90$, $p = 0.015$ and between S:N × context, $F(1,14) = 56.63$, $p < 0.0001$. The three-way interaction of S:N × context × simulation condition was also significant, $F(2,28) = 4.93$, $p = 0.015$. A SNK test of multiple comparisons confirmed that word identification was worse in the jitter condition than in the smear or intact conditions which did not differ significantly from each other ($p < 0.05$).

2.2.3. Discussion

By matching the degree of spectral splatter in the jitter and smear simulations, it was possible to assess the contributions to speech intelligibility of the temporal distortion introduced by the jitter simulation while controlling for the spectral splatter that was a byproduct of the simulation. When the low-frequency components of the sentences were spectrally smeared to a degree that matched the spectral distortion resulting from the jitter simulation, word identification was not reduced. Therefore, the reductions in word identification that were found when the low-frequency components of the sentences were jittered could only be attributed to the temporal distortion that was created by the jitter simulation.

3. General discussion

The results of both experiments show that a simulated decline in synchrony in younger adults resulted in a significant decrease in word identification accuracy in noise. Experiment 1 demonstrated that jittering restricted to low-frequency speech components (up to 1.2 kHz) was sufficient to reduce word identification in younger adults when the SPIN-R sentences were presented in background babble. Experiment 2 demonstrated that word identification was reduced by temporal jittering but not by spectral smearing of the low-frequency components of speech when the degree of spectral distortion was matched between the two simulations.

Previous studies have found that spectral smearing has significantly deleterious effects on speech intelligibility (Baer and Moore, 1993, 1994; ter Keurs et al., 1992, 1993); however, to our knowledge no previous study has investigated the effects of spectral smearing when it is confined to the low-frequency region of speech. It may be that the deleterious effects of smearing on speech intelligibility that were seen in the previous studies are primarily attributable to spectral distortion of the high frequencies, with smearing of low frequencies being less important. It is not unreasonable that spectral distortion of the high frequencies would have more effect on speech intelligibility than spectral distortion of the low frequencies. Conversely, it is not unreasonable that temporal distortion of the low frequencies would have a greater effect on speech intelligibility than spectral distortion of the low frequencies.

Jittering the low-frequency speech components resulted in word identification performance by younger adults that was very similar to that reported previously for older adults with good audiograms when words were presented in low-context sentences. This similarity is consistent with the hypothesis that an age-related decline in synchrony could result in a decline in speech intelligibility in noise for older adults with good audiograms. Future studies that measure synchrony and speech intelligibility in a within-subjects design are needed to shed further light on this hypothesis.

The results of the present study support the argument of Drullman (1995a,b) that the fine structure of the signal contributes to the intelligibility of speech in noise. Drullman (1995a) examined word identification across a range of S:N conditions in noise shaped to the long-term average speech spectrum. In one condition, he eliminated the noise in the peaks of the amplitude envelope of the speech signal in separate quarter-octave bands. His second condition was like the first except that the signal in the troughs of the amplitude envelopes was eliminated, leaving only noise in the troughs. The first manipulation had no effect on intelligibility, but the second manipulation shifted the speech reception threshold by 1.5 dB, suggesting that the fine structure of the signal in noise at the troughs made a significant contribution to intelligibility whereas at the peaks it did not. In the troughs, when only noise is present, the amplitude changes randomly over time. An effect of adding the signal to the noise at the trough is to regularize phase relationships in the fine temporal structure. Hence, these findings are consistent with the notion that periodicity in the speech signal contributes to intelligibility in noise.

Our approach to simulating loss of neural synchrony in aging auditory systems is very simple but it seems promising and complements simulations of cochlear hearing loss such as the smearing method of Baer and Moore (1993, 1994). Previous efforts to simulate the properties of cochlear hearing loss using masking have also been very useful to hearing researchers studying aging (e.g., Humes, 1996; Humes et al., 1988; Humes and Jestaedt, 1991). Simulations of cochlear hearing loss, however, are of limited benefit in modeling the temporal aspects of auditory processing declines in older listeners with good audiograms. Our goal was to simulate the temporal processing characteristics of auditory aging when clinically significant sensori-neural hearing loss is not a major factor. Our approach to simulating loss of synchrony based on disrupting phase-locking differs from the approach of Zeng and colleagues (Zeng et al., 1999) which involved the modification of the envelopes of specific frequency bands. Their goal was to simulate the change in modulation transfer function observed in cases of auditory neuropathy to determine if this could be responsible for their reduced speech intelligibility in quiet. The loss of synchrony exhibited in auditory neuropathy is more extreme than the mild loss of synchrony that we were attempting to model for aging adults.

A limit to our simulation is that our temporal "noise" is correlated across all fibers. In an auditory system with poor synchrony, we would expect this temporal noise to be independent across all fibers. Thus, our simulation may disrupt synchrony coding in normal hearers in a simpler manner than would be observed physiologically in a pathological hearer. While it is possible to apply jitter to multiple frequency bands independently, it does not seem possible to manipulate the acoustical signal to add temporal noise that would emulate uncorrelated jitter across individual fibers sharing the same characteristic frequency. Despite this shortcoming, our simulation was successful in mimicking the word identification performance in noise of older adults in younger adults.

Acknowledgement

This research was funded by the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research. As part of their Master's theses at the University of British Columbia, Hollis Pass conducted preliminary experiments not reported here and Sasha Brown conducted Experiment 1. Experiment 2 was conducted by Ewen MacDonald at the University of Toronto as a doctoral trainee in the Canadian Institutes of Healthy Research Strategic Research Training Program on "Communication and Social Interaction in Health Aging". We wish to thank Carol Jaeger for preparing the original between-band jitter program, Trudy Adam for assistance in data analysis and preparation of an earlier version of the manuscript, and Jennifer Tomaszczyk and Lesley Filmer for assistance in conducting Experiment 2.

Appendix A. Smearing algorithm

Using the smearing procedure of Baer and Moore (1993), a series of frames are created from a time domain signal sampled at a rate of 16 kHz, with each frame consisting of 128 samples (8 ms per frame for a 16-kHz sampling rate) multiplied by a 128 point Hamming window. There is a 50% overlap between sequential frames; i.e., the last 64 samples of the first frame are the same as the first 64 samples of the second frame. Each frame is then zero padded with 64 zeros on each side to increase the frame length to 256 samples. A 256-point FFT is performed on each frame and phase and power vectors are calculated. The power vector from each frame is multiplied by a 128-by-128 point smearing matrix. Thus, each of the 128 components of the power vector is replaced by a weighted sum of the nearby components, while the DC component remains unchanged. The smearing matrix used in Experiment 3 for 0–1187.5 Hz was the same as that used by Baer and Moore (1993). However, since we were only interested in the effects of spectral smearing on low frequencies, the power components were not smeared for frequencies from 1250 to 6785 Hz; instead, that portion of the smearing matrix was replaced by an equivalent portion of an identity matrix. Otherwise the procedure of Baer and Moore (1993) was followed. Power components higher than 6785 Hz were set to zero. The smeared power components and the original phase components were recombined and an IFFT was calculated to produce a smeared frame. These frames were then combined in an overlap-add method (Allen, 1977).

Baer and Moore (1993) calculated the smearing matrix using symmetrical and asymmetrical roex(p) filters (Patterson et al., 1982) to model auditory filters for normal and impaired ears. The bandwidth of the auditory filter increases with increased hearing impairment. When calculating the smearing matrix, a smearing factor which corresponds to the increase in auditory filter bandwidth is chosen. A smearing factor of one corresponds to a normal filter bandwidth; a smearing factor of two corresponds to a doubling in the bandwidth, etc. Using the symmetrical filter model, Baer and Moore (1993) modeled a mild-to-moderate hearing loss using a smearing factor of 3 and a moderate-to-severe hearing loss using a smearing factor of 6. In our implementation, we calculated the smearing matrix using the symmetrical auditory filter model and a smearing factor of 6.

References

Abel, S.M., Krever, E.M., Alberti, P.W., 1990. Auditory detection, discrimination and speech processing in aging, noise-sensitive and hearing-impaired listeners. Scand. Audiol. 19, 43–54.

Alain, C., McDonald, K.L., Ostroff, J.M., Schneider, B., 2001. Age-related changes in detecting a mistuned harmonic. J. Acoust. Soc. Am. 109, 2211–2216.

Allen, J.B., 1977. Short term spectral analysis, synthesis and modification by discrete Fourier transform. IEEE Trans. Acoust. Signal Process. 25, 235–238.

Backoff, P.M., Caspary, D.M., 1994. Age-related changes in auditory brainstem responses in Fischer 344 rats: effects of rate and intensity. Hear. Res. 73, 163–172.

Baer, T., Moore, B.C.J., 1993. Effects of spectral smearing on the intelligibility of sentences in the presence of noise. J. Acoust. Soc. Am. 94, 1229–1241.

Baer, T., Moore, B.C.J., 1994. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. J. Acoust. Soc. Am. 95, 2277–2280.

Bilger, R.C., Nuetzel, M.J., Rabinowitz, W.M., Rzeczkowski, C., 1984. Standardization of a test of speech perception in noise. J. Speech Hear. Res. 27, 32–48.

Boettcher, F.A., Mills, J.H., Swerdloff, J.L., Holley, B.L., 1996. Auditory evoked potentials in aged gerbils: responses elicited by noises separated by a silent gap. Hear. Res. 102, 167–178.

Divenyi, P., Simon, H., 1999. Hearing in aging: issues old and young. Curr. Opin. Otolaryngol. Head Neck Surgery 7, 282–289.

Drullman, R., 1995a. Speech intelligibility in noise: relative contribution of speech elements above and below the noise level. J. Acoust. Soc. Am. 98, 1796–1798.

Drullman, R., 1995b. Temporal envelope and fine structure cues for speech intelligibility. J. Acoust. Soc. Am. 97, 585–592.

Dubno, J.R., Ahlstrom, J.B., Horwitz, A., 2000. Use of context by young and aged adults with normal hearing. J. Acoust. Soc. Am. 107, 538–546.

Durlach, N.I., 1972. Binaural signal detection: equalization and cancellation theory. In: Tobias, J.V. (Ed.), Foundations of Modern Auditory Theory, vol. 2. Academic Press, NY, pp. 371–462.

Fitzgibbons, P.J., Gordon-Salant, S., 1996. Auditory temporal processing in elderly listeners. J. Am. Acad. Audiol. 7, 183–189.

Frisina, D.R., Frisina, R.D., 1997. Speech recognition in noise and presbycusis: relations to possible neural mechanisms. Hear. Res. 106, 95–104.

Frisina, D.R., Frisina, R.D., Snell, K.B., Burkard, R., Walton, J.P., Ison, J.R., 2001. Auditory temporal processing during aging. In: Hof, P.R., Mobbs, C.V. (Eds.), Functional Neurobiology of Aging. Academic Press, San Diego, CA, pp. 565–579.

Gates, G.A., Feeney, M.P., Higdon, R.J., 2003. Word recognition and the articulation index in older listeners with probable age-related auditory neuropathy. J. Am. Acad. Audiol. 14, 574–581.

Grose, J., 1996. Binaural performance and aging. J. Am. Acad. Audiol. 7, 168–174.

He, N.-J., Horwitz, A.R., Dubno, J., Mills, J., 1998. Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects. J. Acoust. Soc. Am. 103, 553–565.

Hellstrom, L.I., Schmiedt, R.A., 1990. Compound action potential input/output functions in young and quiet-raised gerbils. Hear. Res. 50, 163–174.

Humes, L.E., 1996. Speech understanding in the elderly. J. Am. Acad. Audiol. 7, 161–167.

Humes, L.E., Espinoza-Varas, B., Watson, C.S., 1988. Modeling sensorineural hearing loss. I. Model and retrospective evaluation. J. Acoust. Soc. Am. 83, 188–202.

Humes, L.E., Jestaedt, W., 1991. Models of the effects of threshold on loudness growth and summation. J. Acoust. Soc. Am. 90, 1933–1943.

Mills, J.H., Schmiedt, R.A., Schulte, B.A., Dubno, J.R., in press. Age-related hearing loss: a loss of voltage, not hair cells. Sem. Hear.

Moore, B.C.J., 2003a. An Introduction to the Psychology of Hearing, fifth ed. Academic Press, San Diego, CA.

Moore, B.C.J., 2003b. Temporal integration and context effects in hearing. J. Phonetics 31, 563–574.

Moore, B.C.J., Peters, R.W., 1992. Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. J. Acoust. Soc. Am. 91, 2881–2893.

Patterson, R.D., Nimmo-Smith, I., Weber, D.L., Milroy, R., 1982. The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. J. Acoust. Soc. Am. 72, 1788–1803.

Pichora-Fuller, M.K., Schneider, B.A., 1991. Masking-level differences in the elderly: a comparison of antiphasic and time-delay dichotic conditions. J. Speech Hear. Res. 34, 1410–1422.

Pichora-Fuller, M.K., Schneider, B.A., 1992. The effect of interaural delay of the masker on masking-level differences in young and old adults. J. Acoust. Soc. Am. 91, 2129–2135.

Pichora-Fuller, M.K., Schneider, B.A., 1998. Masking-level differences in the elderly: the effect of the level of the masking noise. Percept. Psychophys. 60, 1197–1205.

Pichora-Fuller, M.K., Schneider, B.A., Daneman, M., 1995. How young and old adults listen to and remember speech in noise. J. Acoust. Soc. Am. 97, 593–608.

Pichora-Fuller, M.K., Schneider, B.A., Pass, H., 1999. Temporal jitter disrupts speech intelligibility: simulations of auditory aging. J. Acoust. Soc. Am. 105, 1345.

Pichora-Fuller, M.K., Souza, P., 2003. Effects of aging on auditory processing of speech. Int. J. Audiol. 42 (Supp 2), S11–S16.

Raza, A., Milbrandt, J.C., Arneric, S.P., Caspary, D.M., 1994. Age-related changes in brainstem auditory neurotransmitters: measures of GABA and acetylcholine function. Hear. Res. 77, 221–230.

Schneider, B.A., 1997. Psychoacoustics and aging: implications for everyday listening. J. Speech–Language Pathol. Audiol. 21, 111–124.

Schneider, B.A., Pichora-Fuller, M.K., 2000. Implications of perceptual processing for cognitive aging research. In: Craik, F.I.M., Salthouse, T.A. (Eds.), The Handbook of Aging and Cognition, second ed. Lawrence Erlbaum Associates, NY, pp. 155–219.

Snyder, J.S., Alain, C., 2005. Age-related changes in neural activity associated with concurrent vowel segregation. Cognitive Brain Res. 24, 492–499.

Summers, V., Leek, M., 1998. F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. J. Speech Hear. Res. 41, 1294–1306.

ter Keurs, M., Festen, J.M., Plomp, R., 1992. Effect of spectral envelope smearing on speech reception. I. J. Acoust. Soc. Am. 91, 2872–2880.

ter Keurs, M., Festen, J.M., Plomp, R., 1993. Effect of spectral envelope smearing on speech reception. II. J. Acoust. Soc. Am. 93, 1547–1552.

Vongpaisal, T., Pichora-Fuller, M.K., in press. Effect of age on F_0 difference limen and concurrent vowel identification. J. Speech Language Hear. Res.

Wingfield, A., 1996. Cognitive factors in auditory performance: context, speed of processing, and constraints of memory. J. Am. Acad. Audiol. 7, 175–182.

Zeng, F.-G., Oba, S., Garde, S., Sininger, Y., Starr, A., 1999. Temporal and speech processing deficits in auditory neuropathy. NeuroReport 10, 3429–3435.