Does elderly speech recognition in noise benefit from spectral and visual cues?

Yatin Mahajan[1], *Jeesun Kim*[1], *Chris Davis*[1]

[1] The MARCS Institute, University of Western Sydney

y.mahajan@uws.edu.au,j.kim@uws.edu.au,chris.davis@uws.edu.au

Abstract

Previous research with young adults has shown that temporal (amplitude modulated, AM) cues are sufficient for recognizing speech in quiet but not for speech in noise. Speech perception in noise is more robust when spectral (frequency modulated, FM) cues are provided in addition to AM ones; visual cues (AV) provide an additional benefit. The elderly typically have problems recognizing speech in noise and it has recently been found that FM discrimination is worse in this group. Given this, it may be that elderly participants will not show an FM speech in noise benefit. To test the relative effectiveness of adding FM cues to AM ones for elderly versus young participants we compared auditory only (AO) speech identification of sentences, vowels and consonants in noise with AM and AM+FM presentation conditions. We also evaluated the relative effectiveness of visual cues for the elderly group. Although the elderly had poorer speech recognition performance overall, they showed a comparable visual benefit to the young group. Moreover, contrary to the prediction of a reduced benefit for FM cues, the FM benefit for the elderly was similar to that of young adults. These results were discussed in relation to speech specific auditory processing.

Index Terms: speech perception, frequency modulation, elderly

1. Introduction

Efficient speech recognition utilizes multiple speech cues; the influence of which varies depending on the listening situation [1]. At one level of description, recognition cues are composed of various time-frequency energy distributions. Indeed, the division between time-based cues and frequency-based ones is a basic one. Both temporal cues (the change in loudness over time (as represented by the amplitude envelope) and spectral cues (pitch and energy information; frequency modulation cues (FM) are fundamental to speech recognition. In addition, visual speech cues (the information gained by seeing the talker's lips, mouth, jaw, face and head movements; audio-visual cues (AV) play an important role, especially in difficult listening environments.

The use of AM, FM (and AV) cues in recognizing speech in young adults has been a topic of considerable research. For instance, it has been reported that although the AM cues alone can provide sufficient information for recognizing speech in quiet situations [2,3], recognition performance deteriorates markedly in noisy backgrounds [3,4-6]. This reduction in the efficiency of AM cues in noisy conditions can be offset by addition of FM cues to the speech signal. Indeed, it has been demonstrated that addition of even slowly varying FM cues to AM cues results in improved speech recognition abilities when listening to a competing talker and against noisy backgrounds [3,6-8]. Of course, also it has also been shown that visual speech cues have perhaps the largest beneficial effect for recognizing speech in noise. Recently it has been found that, by and large, such cues have additive effects with AM or AM+FM cues in facilitating speech recognition scores in noise [3]. The benefit of adding FM to AV cues has thus only been demonstrated in young adults. That is, it is not known that if older adults will be able to extract a similar benefit from FM cues.

Indeed, the results of several studies suggest that they may not. In general, elderly speech recognition ability in the presence of background noise is poor [9-11] and [12,13] explicitly examined FM discrimination in elderly listeners. The results of both studies showed that elderly listener's detection of frequency modulation was worse (they had higher thresholds) than younger ones. Based on this, it was concluded that FM detection is impaired in older listeners. Given these findings, it would seem a reasonable hypothesis that older adults will show less of an FM benefit for speech recognition in noise. However, it should be pointed out that these two studies specifically used a tailored FM detection tasks with non-speech stimuli. For example, a 5 Hz low pass filtered noise frequency modulated on a 1000 Hz carrier [12] and a 5 Hz sinusoidal and quasitrapezoidal frequency modulated on 500 Hz and 4000 Hz carriers [13].

To evaluate this proposition we compared the benefit to speech identification of sentences, consonants and vowels (CV) in noise that the addition of FM cues had for young and older listeners. In addition to determining the relative FM benefit for an older and young group, we also examined the extent to which AV cues would benefit speech recognition for these groups. Finding that the older group showed a weaker benefit for these rather different cue types might indicate a more general processing problem. Both sentence and consonant and vowel syllables (CV) stimuli were used in the present study because: 1) Sentences provide an opportunity to study supra-segmental cues. 2) CV syllables provide an opportunity to probe for an interaction between FM and AV cues in terms of phonetic features, such as manner and place of articulation [3].

2. Methods

2.1. Participants

Two groups of listeners (all with English as their native language) participated in the experiment. A young adult group (10 undergraduate university students, mean age: 27.5 years, range 18 - 40 years, 5 males) and an elderly group (9 individuals, mean age; 71.8 years, range 68 -76 years, 6 males). Prior to participation all were given a hearing screening (pure tone frequencies at 500, 1000, 2000 and 4000 Hz). All had normal hearing acuity: the young group (< 20 dB HL) and elderly group (< 30 dB HL). All participants had 20/20 uncorrected (or corrected) vision. No participant reported any significant psychological, medical or neurological history. All the participants scored 25 or greater on the Mini Mental Status Examination [14].

2.2. Test Materials

Two sets of stimuli were recorded for the experiments spoken by a male native Australian English speaker and recorded in a sound treated room. Set 1: Eighty sentences were selected from the 1969-revised IEEE/Harvard list of phonetically balanced sentences. IEEE sentences have a complicated sentence structure and information content that have been shown to elicit lower recognition performance than either CUNY sentences or Hearing in Noise Test sentences [6]. Set 2: Twenty-seven syllables (16 consonants in aCa context; C = /p, t, k, f, θ, s, ʃ, b, d, g, v, z, ʤ, m, n, l/ and 11 vowels in hVd context; V = /i:, I, æ, e, æ:, ɔ, a, a:, ʊ, ʉ, ɜ:/. These syllables were selected such that the different consonants and vowels covered a range of place and manner of articulation in English. The speaker was video recorded against a uniform gray background, facing the camera and the recording showed the head and shoulders. A sample from a commercially available multi-talker babble track (Auditec, St. Louis, MO) was used as competing noise stimuli. The stimuli and noise were mixed at -5dB signal to noise ratio (SNR). The sentences and the syllables once mixed with multi talker babble were processed using the FAME (frequency amplitude modulation encoding) processing algorithm [6] to extract AM and AM+FM signals. That is, the FAME algorithm partitions the speech signal (sentences, vowels and consonants) into slowly varying AM and FM signals. The onset of the noise occurred prior to the speech stimuli. Once the noise was added, the stimuli were then dubbed onto the video of the talker.

2.3. Procedure

Each participant was tested in a sound proof booth. The stimuli were presented through Sennheiser HD515 headphones. The audio and corresponding video signals were played using DMDX software [15] on a monitor with a resolution of 1024x768. In the auditory only condition a blank screen was presented.

The sentence materials which consisted of 80 IEEE sentences in 8 different conditions (10 sentences each): AO AM quiet, AO AM noise, AO AM+FM quiet, AO AM+FM noise, AV AM quiet, AV AM noise, AV AM+FM quiet, AV AM+FM noise. Conditions were presented in blocks, with the order of blocks and sentences within blocks randomized. Different versions of the sentences lists were constructed so no item was repeated in any version but all sentences and conditions were included (sentences for all conditions were fully rotated). Each participant was given randomly a version of sentence list. There were four practice items. Following the presentation of each stimulus the participant typed the responses. While scoring the data, only the content words from each sentence were scored. The mean percent word correct for each condition formed the dependent measure of speech recognition in quiet or noise consisting of four talker babble (three female talkers and one male).

All vowel and consonant items were presented in four different conditions: AO AM, AO AM+FM, AV AM and AV AM+FM. All conditions were presented in background noise of multi-talker babble. From set 2 of the stimuli, four lists of consonants and vowels were prepared containing all four conditions and presented to each participant with two lists each on either side of the sentence recognition task, with no item repeated in any condition. All participants completed all four lists. The AO and AV conditions were presented in blocks but the order in which items in each list was presented was randomized. For this task, closed set responses were taken from each participant. After every item, a response grid with all the possible options was presented on the monitor and participants selected the correct response from the grid. The participants were given adequate breaks in between the blocks to alleviate any fatigue.

3. Results

3.1. Sentences

The mean percent correct word recognition scores for sentences in each condition in quiet and noise across two groups are shown in Figures 1 & 2 (respectively). To evaluate the effect of AM, FM and AV cues on speech recognition in quiet and noise, a 2 x 2 x 2 x 2 mixed ANOVA was conducted with condition (quiet vs noise), cues (AM vs AM+FM) and modality (AO vs AV) served as with in subject factors and age (young vs older adults) as between subject factor. From Figure 1 & 2 it is clear that the identification of sentences degraded considerably when presented in presence of noise. There was a significant reduction in mean percent scores for AO AM speech in quiet versus in-noise across both the groups (from 86% correct to 10% in young adults and from 81% to 5% in older adults), $F(1,17) = 1167.23$, $p < .001$). There were two significant three way interactions between condition, cues and age ($F(1,17) = 4.59$, $p = .04$) and between condition, cues and modality ($F(1,17) = 7.42$, $p = .014$). To understand the first three-way interaction, a two-way ANOVA was performed across quiet and noise conditions with 'cues' as with in subject factor and 'age' as between subject factor. Two-way ANOVA with 'cues' and 'modality' as with in subject factors was conducted to resolve second interaction.

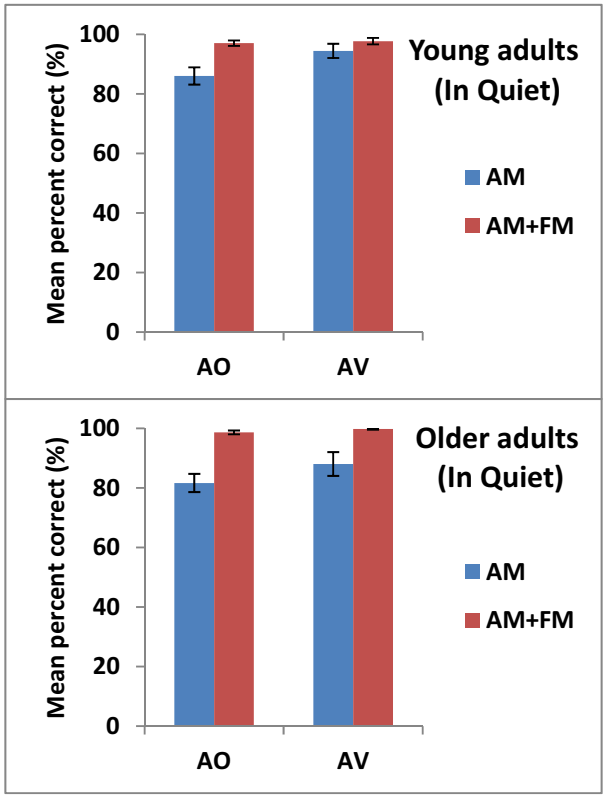

Figure 1: *Mean percent word identification scores for young (upper panel) and older adults (lower panel) in quiet across AM, AM+FM, AO and AV conditions.*

3.1.2. Word recognition in quiet

It is evident from Figures 1 & 2 that the performance of the participants was near ceiling for all the conditions, particularly for the AV conditions. Hence, the results were not pursued in detail for quiet situation. In quiet condition, additions of FM effect produced higher mean percent scores though similar FM enhancement across two groups (7% more in young adults and 15% more in older adults). There was AV (visual enhancement effect as well with marginal (4%) increase in correctly identified words in sentences.

3.1.3. Word recognition in noise

In the presence of noise, a greater number of words were correctly identified when FM cues were added to the AM signal, ($F(1,17) = 43.92$, $p < .001$) across both younger and older adults. The mean percent scores between young and older adults significantly differed ($F(1,17) = 11.77$, $p = .003$, partial $\eta^2 = .71$) with younger adults performing better than older adults. There was no interaction between the factors age and cues implying that, both young and older adults received similar FM benefit while identifying words in a sentence in the presence of noise.

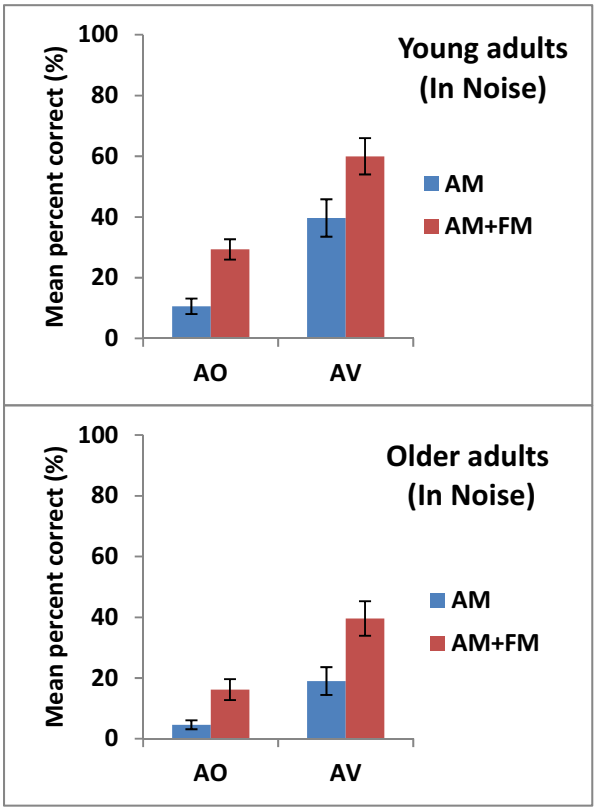

Figure 2: *Mean percent word identification scores for young (upper panel) and older adults (lower panel) in noise across AM, AM+FM, AO and AV conditions.*

Similar to speech recognition in quiet, there was a significant FM effect ($F(1,18) = 46.07$, $p < .001$) and significant AV effect ($F(1,18) = 46.25$, $p < .001$) across all the participants with approximately 18% increase in correct responses in AM+FM conditions and 25% increase in AV condition. The FM and AV effect was comparable across young and older adults. The lack of interaction between cues and modalities restricted the evaluation of benefits across AM, AM+FM and AO and AV conditions.

3.2. Consonants and Vowels

The mean percent correct scores in noise for consonants and vowels across the two groups are shown in Figures 3 & 4 (respectively). A mixed ANOVA with 2 (cues: AM vs AM+FM) x 2 (modality: AO vs AV) x 2 (type of syllable: consonants vs vowels) x 2 (age: young vs older adults), with 'cues', 'modality' and 'type of syllable' as within subject and age as between subject factor was conducted to evaluate the effect of FM and AV cues on perception of consonants and vowels in noise. The results revealed poorer identification scores for both consonants and vowels by older adults. The younger adults identified more consonants than vowels correctly ($F(1,17) = 7.23$, $p = .015$) whereas the older adults did not show any difference in correct identification between consonants and vowels.

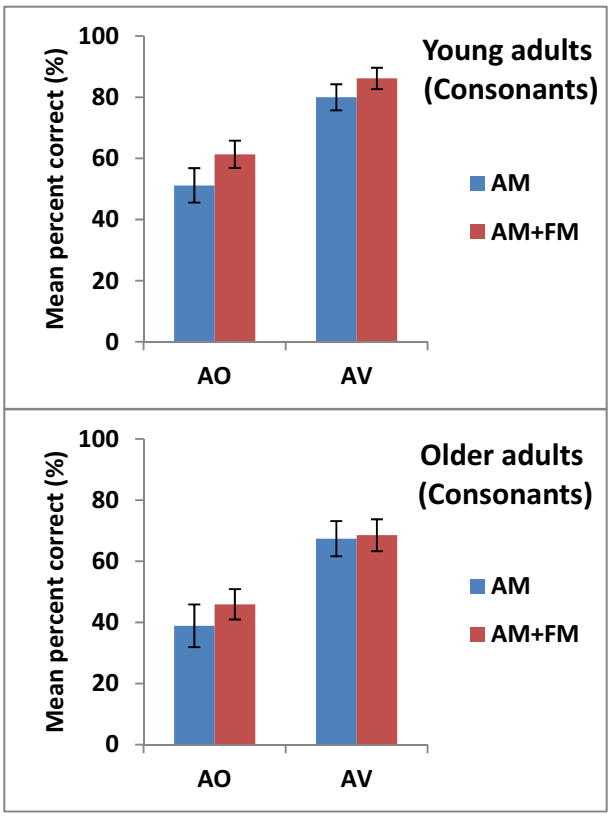

Figure 3: *Mean percent correct consonants scores in noise for young (upper panel) and older adults (lower panel) across AM, AM+FM, AO and AV conditions.*

Across groups, there was a significant visual enhancement effect (AV effect) with more number of correct consonants ($F(1,18) = 130.53$, $p < .001$) and vowels ($F(1,18) = 15.37$, $p = .001$) identified in AV condition than AO with comparable performance in young and older adults. In addition, there was significant FM effect as well on syllable identification with a greater number of syllables correctly identified in AM+FM condition in both AO ($F(1,17) = 6.23$, $p = .023$) and AV conditions ($F(1,17) = 5.49$ $p = .035$) with equal FM effect in both younger and older adults.

In summary, speech recognition in noise was poor in older adults when compared to younger adults. For both sentences and syllable identification, the FM benefit and visual enhancement (AV) were similar across young and older adults.

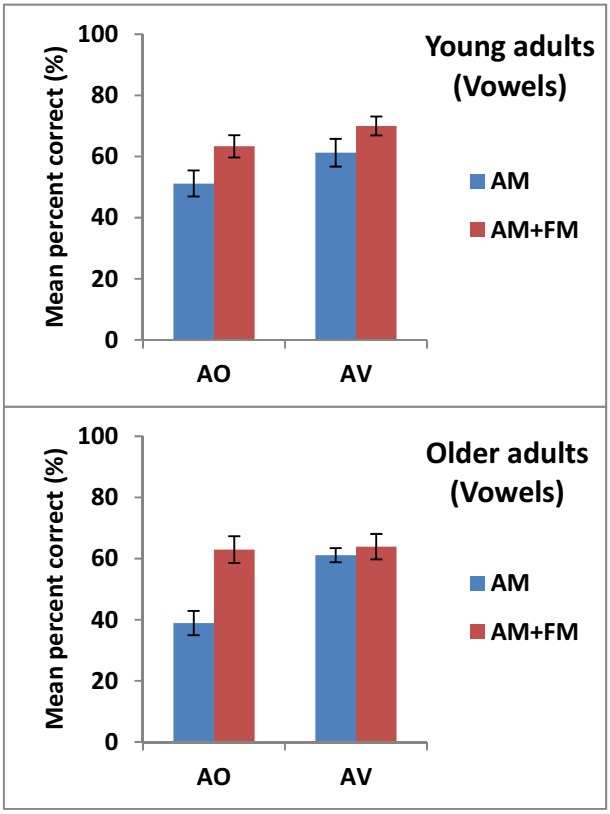

Figure 4: *Mean percent correct vowels scores in noise for young (upper panel) and older adults (lower panel) across AM, AM+FM, AO and AV conditions.*

4. Discussion

The present experiment examined whether elderly and young listeners would exhibit a similar benefit for speech perception in noise from the addition FM speech cues. A further test was made to determine the relative size of the visual speech benefit for these groups. Two speech identification in noise tasks were carried out, one examining the identification of sentence material and other, the identification of aCa and hVd syllables. The results showed that although, overall speech recognition performance was poor in the older adults, the benefit they accrued from the addition of FM cues was comparable to the young adults. Likewise, the benefit of AV cues was similar in the older and younger groups.

4.1 FM benefit for speech recognition in noise in elderly

The FM benefit was similar in older adults when compared to young adults for speech recognition in noise. This result was contrary to our hypothesis of reduced FM benefit in older adults. Two possibilities may explain this outcome. The first concerns stimulus related differences between previous studies. In [12,13], meaningless and abstract non-speech stimuli were used. These stimuli are useful for the examination of elementary auditory processing at sensory peripheral levels. On the other hand, the present study used the speech stimuli which are processed by memory backed, higher level auditory processing. It is possible that the impairment in FM discrimination as shown with non-speech stimuli in the elderly is not represented in speech recognition in noise, a feature of higher order of sensory perception that has access to stored perceptual representations. The second is that the current elderly group simply does not have an FM discrimination problem (and that is why they could effectively utilize the FM speech cues). In this regard, a follow-up experiment has been planned which includes the joint testing of FM discrimination abilities in older adults along with speech recognition in noise to determine if FM benefits exists in the elderly in case there is an impaired FM discrimination.

The speech identification scores for sentences in noise were enhanced in both younger and older adults when FM cues were added in AO condition. The speech stimuli processed by FAME processing algorithm delivers both temporal envelope and temporal fine structure cues in turn transmitting more acoustic information for speech processing noise [3]. The temporal fine structure cues are preserved by FAME which are responsible for F0 and harmonic properties of speech stimuli along with intonation patterns and may segregate signal and noise into separate perceptual streams resulting in better speech identification in noise. The FM cues transmitted by FAME also helps in perceiving consonant-vowel transitions essential for consonant and vowel identification in the presence of background noise leading to more number of consonants and vowels identified in noise in AM+FM condition compared to AM condition.

4.2 AV benefit for speech recognition in noise for elderly

A clear additive benefit of AV cues when added either to AM or FM cues was evident for both young and older adults. Along with the extra acoustic information provided by the FM cues, the visual cues provide additional facilitation in terms of identity of the target speech e.g. for consonant identification in the form of place of articulation transmitted by FM/AM cues [3]. This explanation is supported only in the better speech identification scores for consonants than vowels in case of young adults. The contribution of AV cues was similar for consonants and vowels in older adults. Facilitation from FM cues has been reported to be better for vowels than consonants [3]. The elderly might have been able to integrate auditory cues (manner of articulation) provided by AM and FM signals better with AV cues (place of articulation) resulting in comparable recognition of vowels and consonants.

5. Conclusions

The present study determined if the relative benefit of FM cues reduces in older adults compared to young adults, as result of poor FM discrimination skills reported among older adults. The results indicated that FM cues facilitate speech perception in noise similarly in both young and older adults. A strong additive effect of AV cues was also indicated in the results which assisted better perception of speech.

6. References

[1] McMurray, B., and Jongman, A., "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations", Psychol. Rev., 118:219-246, 2011.

[2] Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M., "Speech recognition with primarily temporal cues", Science, 270:303–304, 1995.

[3] Kim, J., Davis, C., "Speech identification in noise: contribution of temporal, spectral, and visual speech cues", J. Acoust. Soc. Am., 126:3246-3257, 2009.

[4] Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z., "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels", J. Acoust. Soc. Am., 104: 3583–3585, 1998.

[5] Fu, Q. J., and Shannon, R. V., "Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant", Ear Hear., 20:332–344, 1999.

[6] Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K., "Speech recognition with Amplitude and Frequency Modulations", Proc. Natl. Acad. Sci. U.S.A., 102:2293–2298, 2005.

[7] Nie, K., Stickney, G., and Zeng, F. G., "Encoding frequency modulation to improve cochlea implant performance in noise", IEEE Trans. Biomed. Eng., 52:64–73, 2005.

[8] Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G., "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences", J. Acoust. Soc. Am., 122:1069–1078, 2007.

[9] Pichora-Fuller, K., "Effects of aging on auditory processing of speech", Int. J. Audiol., 42:2S11-2S16, 2003.

[10] Helfer, K. S., and Freyman, R. L., "Aging and speech-on-speech masking", Ear Hear., 29:87-98, 2008.

[11] Dubno, J. R., Lee, F., Matthew, L. J., Ahlstrom, J. B., Horwitz, A. R., and Mills, J. H., "Longitudinal changes in speech recognition in older persons", J. Acoust. Soc. Am., 123:462-475, 2008.

[12] Sheft, S., Shafiro, V., Lorenzi, C., McMuellen, R., and Farrell, C., "Effects of age and hearing loss on relationship between discrimination of stochastic frequency modulation and speech perception", Ear Hear., 33:709-720, 2012.

[13] He, N., Mills., J. H., and Dubno, J. R., "Frequency modulation detection: Effects of age, psychophysical method, and modulation waveform", J. Acoust. Soc. Am., 122:467-477, 2008.

[14] Folstein, M. F., Folstein, S. E., and McHugh, P. R., "Mini-mental state, A practical method for grading the cognitive state of patients for the clinician", J. Psychiatr. Res., 12:189-198, 1975.

[15] Forster, K. I., and Forster, J. C., "DMDX: A windows display program with millisecond accuracy", Behav. Res. Methods Intrum. Comput., 35:116-124, 2003.