Individual differences in top-down restoration of interrupted speech

Benard, Michel Ruben; Mensink, Jorien Susanne; Başkent , Deniz

Published in:
Journal of the Acoustical Society of America

DOI:
10.1121/1.4862879

Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities

Michel Ruben Benard[a] **and Jorien Susanne Mensink**
Pento Audiology Center Zwolle, Oosterlaan 20, 8011 GC Zwolle, The Netherlands
m.r.benard@umcg.nl, j.mensink@pento.nl

Deniz Başkent[b]
University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology/Head and Neck Surgery, Postbus 30.001, 9700 RB Groningen, The Netherlands
d.baskent@umcg.nl

Abstract: Top-down restoration mechanisms can enhance perception of degraded speech. Even in normal hearing, however, a large variability has been observed in how effectively individuals can benefit from these mechanisms. To investigate if this variability is partially caused by individuals' linguistic and cognitive skills, normal-hearing participants of varying ages were assessed for receptive vocabulary (Peabody Picture Vocabulary Test; PPVT-III-NL), for full-scale intelligence (Wechsler Adult Intelligence Scale; WAIS-IV-NL), and for top-down restoration of interrupted speech (with silent or noise-filled gaps). Receptive vocabulary was significantly correlated with the other measures, suggesting linguistic skills to be highly involved in restoration of degraded speech.

PACS numbers: 43.71.An, 43.71.Rt, 43.71.Es, 43.72.Dv [DOS]
Date Received: October 9, 2013 **Date Accepted:** December 11, 2013

1. Introduction

Listeners use top-down restoration mechanisms in complex acoustical environments to enhance perception of degraded speech (Sivonen *et al.*, 2006; Warren, 1983). The brain is able to use audible segments of interrupted speech for its perceptual restoration. However, while this effect is robust and highly repeatable on average, even in well-controlled young normal-hearing populations, a large variability has been observed in how well individuals can make use of these mechanisms (Başkent, 2010).

The influences of linguistic and cognitive skills on the perception of interrupted speech has been a topic of debate (Başkent, 2012; Benard and Başkent, 2013; Sivonen *et al.*, 2006). Previous research has indirectly implied that, in such top-down restoration, linguistic and cognitive factors seem to play an important role. For example, restoration seems to be easier with speech segments that are linked by context and linguistic rules (Sivonen *et al.*, 2006; Verschuure and Brocaar, 1983; Wang and Humes, 2010). Linguistic skills, vocabulary, and verbal comprehension were indirectly suggested to be of importance in the perception of interrupted speech (Bashford *et al.*, 1992; Saija *et al.*, 2013). Training improves the intelligibility of interrupted speech, and aging can have an effect on it, both implying an involvement of the cognitive system

[a]Author to whom correspondence should be addressed. Also at: Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands.
[b]Also at: University of Groningen, Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), The Netherlands.

(Benard and Başkent, 2013; Saija *et al.*, 2013). Despite these suggestions, however, to date, no systematic study has been performed to directly and fully establish if and to what degree such association exists between linguistic and cognitive factors and the top-down restoration of interrupted speech.

In this study, we hypothesized that (1) linguistic skills, such as receptive vocabulary, and (2) cognitive skills, such as overall intelligence, would account for the individual differences in the use of top-down restoration mechanisms. To test the two hypotheses, standardized and validated tests were used to characterize linguistic and cognitive abilities of normal-hearing individuals with varying ages, and the correlations of these measures with the ability to perceptually restore interrupted speech were investigated.

2. Materials and methods

2.1 Listeners

Twelve native speakers of Dutch, aged between 21 and 63 yrs [mean age = 40.4 yrs, standard deviation (SD) = 15.3 yrs, 9 women], participated in the study. Normal hearing was confirmed with hearing thresholds less than or equal to 20 dB hearing level (HL) when averaged across 500, 1000, 2000, and 4000 Hz, applied to the better hearing ear, was confirmed. Speech and language problems were ruled out via a questionnaire.

2.2 Speech stimuli

Dutch high context sentences, representing conversational speech, spoken by a male speaker and digitally recorded at a sampling rate of 44.1 kHz, were used in the study (Versfeld *et al.*, 2000). This corpus consists of 39 sets, with 13 unique sentences per set. Each sentence consists of 4 to 9 words and each set contains 74 to 88 words in total.

Perceptual restoration of interrupted speech was measured with various speech manipulations so that individual variability could be fully exploited. Some of these manipulations involved interrupting speech stimuli with periodic silent intervals, where the listener had to reconstruct the speech stream from the remaining samples (e.g., Chatterjee *et al.*, 2010), and some others involved filling the silent intervals with loud noise bursts, which induced continuity illusion and phonemic restoration, quantified by an increase in intelligibility of interrupted speech (e.g., Benard and Başkent, 2013). Further, to capture cognitive effects fully, such as potential effects of cognitive slowing down due to aging, slow speech rates were also included (Saija *et al.*, 2013). As a result, a number of settings were used to measure top-down restoration, with two interruption rates (1.25, 2.50 Hz) at two speech rates (slow, normal), presented with or without filler noise, producing eight testing conditions.

During signal processing, first the speech rate was changed. For slowing down, sentences were lengthened to twice the original duration, preserving the original pitch using PRAAT software (Saija *et al.*, 2013). The normal and slow-rate sentences were interrupted with the two interruption rates, with a duty cycle of 50%, and with 10-ms cosine-ramping on and off transitions. The filler noise bursts were produced with the same gating function, but with an inverse phase, from the steady speech-shaped noise matching the long-term average speech spectra of the sentences in the corpus (Versfeld *et al.*, 2000). In total, 104 unique sentences (8 conditions $\times$ 13 sentences per sentence set) were used for training and 104 unique sentences were used for testing. The training and the test procedures took 1 h each, including a break. All manipulations were processed online using MATLAB on a Macintosh computer.

The sentences were calibrated to a presentation level of 60 dB sound pressure level (SPL) at the approximate position of the participant's ear, and the filler noise to 70 dB SPL.

2.3 Peabody Picture Vocabulary Test

We used the Peabody Picture Vocabulary Test (PPVT-III-NL) to measure receptive Dutch vocabulary and verbal intelligence (Bell *et al.*, 2001). The experimenter presented the test verbally, at the same time showing a series of four numbered pictures to the participant. The participants' task was to say the number of the picture that described the word stated by the experimenter. Each series of pictures was part of a set of 12 series. The first presented set depended on the age of the participant, after more than three incorrect responses in this set the experimenter presented an easier set. The starting set was the set in which less than four incorrect responses were given. After completing one set the experimenter presented the next and more difficult set, ending with the set in which nine or more incorrect responses were given. The test procedure took around 15 min.

The raw scores of this test are age-independent, and can be translated into an age-dependent receptive vocabulary quotient. A score of 100 points is defined as the median of the norming sample; the SD is 15 points. The Dutch standardization sample contains 1746 children (age 2:3 up to 15:11 yrs) and 1164 adults (age 17:0 up to 90:0 yrs), controlled for gender, age, education, social status, and geographic spread (Schlichting, 2005).

2.4 Wechsler Adult Intelligence Scale

We used the Dutch version of Wechsler Adult Intelligence Scale (WAIS-IV-NL) to measure the Full Scale Intelligence Quotient (FSIQ) (Wechsler, 2012), a composite score composed of ten subtests that measure major components of intelligence, namely, verbal comprehension, perceptual reasoning, working memory, and processing speed. The experimenter verbally instructed the participants prior to each subtest. The participants had to arrange blocks according to a pattern, describe similarities between two concepts, repeat digit spans, solve nonverbal abstract problems, describe meaning of words (vocabulary), perform arithmetic tasks, search or substitute symbols on a response form, arrange visual puzzles, and perform a general knowledge subtest (Benson *et al.*, 2010). The experimenter scored each response and proceeded to a subsequent subtest when the stop criteria, based on incorrect responses or time limits, were met. The test procedure took around 2 h, including a short break.

Different than the earlier version WAIS-III-NL, the WAIS-IV-NL does not provide separate verbal and performance intelligence quotient scores, but quotient scores for verbal comprehension, perceptual reasoning, working memory, and processing speed. A score of 100 points is defined as the median of the norming sample; the SD is 10 points. The norms are divided into 10 age groups, from 16:0 to 84:11 yrs. The sample of standardization is based on 1000 participants in The Netherlands and 500 Dutch-speaking participants in Belgium.

2.5 Experimental procedure

The participants were assessed in random order with the PPVT-III-NL and the WAIS-IV-NL, and they were trained and tested with the normal and slow-rate interrupted sentences with or without filler noise (based on Benard and Başkent, 2013). A clinical psychologist conducted the linguistic and cognitive tests according to the standard clinical procedures, as explained above. A clinical audiologist measured the intelligibility of interrupted speech. The speech stimuli were directed via an AudioFire 4 external soundcard of Echo Digital Audio Corporation (Santa Barbara, CA). The participants were seated together with the experimenter in a sound proof booth and the participant listened to the audio stimulus via a Sennheiser HD 600 headphone (Old Lyme, CT). The experimenter was listening to the participant's response and scored online using a customized MATLAB graphical user interface (method based on Benard and Başkent, 2013) with the sentence text presented on the screen. All words of the sentences were used for scoring; the MATLAB program calculated the ratio between the total number of correctly repeated words and the total number of words within the set presented.

3. Results

Figure 1 shows the percent correct scores for intelligibility of interrupted speech plotted against the quotient scores of the PPVT-III-NL (upper panels) and the WAIS-IV-NL (lower panels). The left and right panels show the results of the normal and slow speech, respectively. The individual scores are overlaid with fitted linear regression lines and shown separately for different interruption conditions. The statistical analysis, by means of linear regression, shows that most speech conditions (5 out of 8) correlate significantly ($p < 0.05$) with the PPVT-III-NL, only 1 out of 8 conditions correlates significantly and 1 out of 8 conditions approaches significance ($p = 0.051$) with the FSIQ scores of the WAIS-IV-NL.

Linear regression was conducted to explore the correlation between the two tests, as the WAIS-IV-NL has subcomponents on linguistic skills (Bell *et al.*, 2001; Benson *et al.*, 2010). This analysis showed that the PPVT-III-NL is indeed highly correlated with the FSIQ ($F(1,10) = 12.93$, $p = 0.005$, $r = 0.751$).

Linear regressions aiming on the correlations between the speech conditions and the major components of intelligence (verbal comprehension, perceptual reasoning, working memory, and processing speed) of the WAIS-IV-NL showed significant correlations with verbal comprehension in only three out of eight conditions ($r > 0.65$, $p < 0.05$). Working memory, perceptual reasoning, and processing speed did not correlate with the speech scores. However, note that these correlations were also underpowered ($r < 0.58$, $p > 0.05$).

To compare the present findings with previous literature on phonemic restoration benefit, i.e., the increase in performance as a result of the addition of the filler noise (N), the data were also inspected for this aspect. Figure 2 shows the increase in intelligibility when the silent intervals are filled with noise (N scores minus S scores), plotted against the scores of the interrupted sentences without filler noise (S scores) for

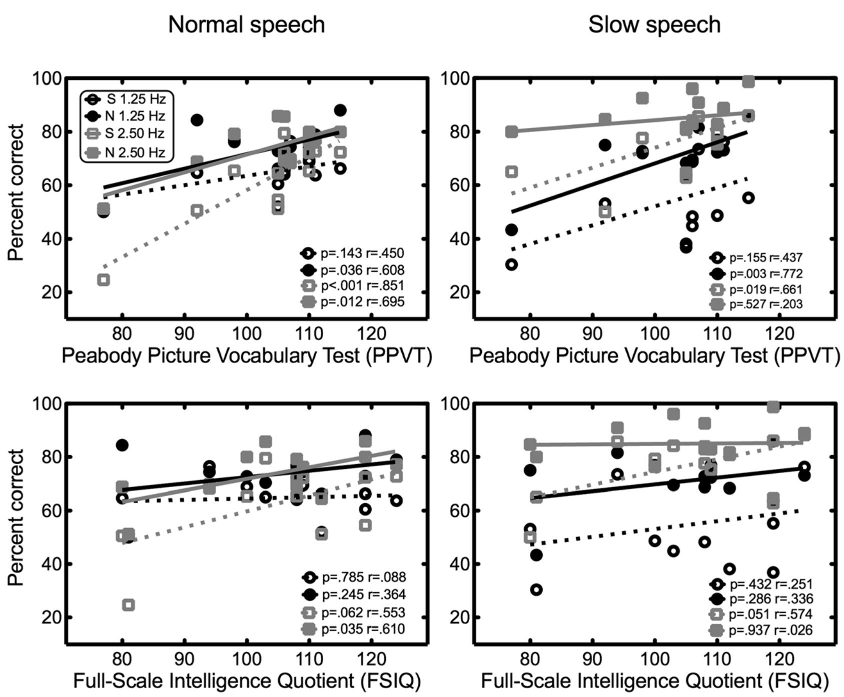

Fig. 1. The percent correct scores for intelligibility of interrupted speech shown as a function of the quotient scores of the PPVT-III-NL (upper panels) and the full-scale intelligence quotient scores of the WAIS-IV-NL (lower panels). The columns show results with normal and slow-rate sentences. The open and filled symbols represent the scores without (S conditions) and with filler noise (N conditions), respectively. The black and gray symbols represent the scores with the slow (1.25 Hz) and fast (2.50 Hz) interruptions. The black and gray trend lines represent the best linear regression, by means of the linear least squares method, through the black and gray data points (continuous for N conditions, and dashed for S conditions), respectively. The correlation analyses are indicated in each panel for individual regression lines.

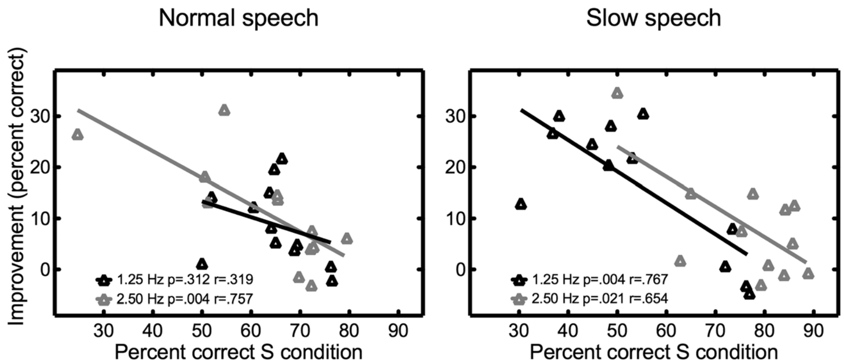

Fig. 2. Phonemic restoration benefit shown as a function of the intelligibility scores in the silent conditions (S) for the normal speech (left panel) and slow speech (right panel). The black and gray triangles show the results for the slow (1.25 Hz) and fast (2.50 Hz) interruptions, respectively, and the regression lines are shown with corresponding colors.

normal (left panel) and slow speech (right panel), along with best-fit regression lines. Figures 1 and 2 together confirm the benefit from adding filler noise (better N scores than corresponding S scores in Fig. 1 and statistical significant improvement directly shown in Fig. 2). A three-factor repeated measure analysis of variance with the within-subject factors the speed of the sentences (normal and slow), the interruption rate (1.25 and 2.50 Hz) and the addition of filler noise, showed significant differences between the S and N scores ($F(1,11) = 32.86$, $p < 0.001$, power = 0.999), and between the slow and fast interruption rates ($F(1,11) = 16.04$, $p = 0.002$, power = 0.957). No significant difference was observed between the normal and slow speech conditions ($F(1,11) = 2.74$, $p = 0.126$, power = 0.328). There was a significant interaction between the factors speed of the sentences and the interruption rates ($F(1,11) = 32.55$, $p < 0.001$, power = 0.999). Further, similar to previous studies (Başkent, 2010; Verschuure and Brocaar, 1983), the improvement depends on the baseline scores (S conditions, see Fig. 2); more benefit is observed at lower S scores.

4. Discussion

The intelligibility of interrupted sentences, interrupted at two different rates and played at normal and slow speed, was measured with and without filler noise in the silent intervals. Similar to previous studies, a phonemic restoration benefit was observed, in a magnitude and trend as was expected from literature (Bashford *et al.*, 1992; Başkent, 2010, 2012; Benard and Başkent, 2013; Verschuure and Brocaar, 1983). Hence, despite the relatively small number of participants in this study, we can assume that the intelligibility scores of the present study were a good representation of top-down restoration of interrupted speech.

In the present study, using the individual differences for understanding interrupted speech, we explored the involvement of linguistic and cognitive factors on top-down restoration of speech. Our first hypothesis was about linguistic skills, and to test this hypothesis, we assessed the receptive vocabulary and verbal intelligence, by means of the PPVT-III-NL (see Fig. 1). Confirming our first hypothesis, the PPVT-III-NL scores showed in general significant correlations with the intelligibility scores of normal and slow speech, with or without filler noise in the silent intervals. Given these correlations, we conclude that the restoration mechanisms of interrupted speech make use of knowledge of receptive vocabulary and verbal intelligence.

A second hypothesis was on cognitive skills, and we used a measure for the full-scale intelligence, namely the WAIS-IV-NL. This hypothesis was not supported by our data. This may mean that linguistic factors are more important for top-down restoration. An alternative explanation is that a full-scale intelligence score is not sufficiently sensitive to capture effects from specific cognitive components, such as working memory (Akeroyd, 2008), because of the composite nature of the WAIS-IV-NL. The full-scale intelligence quotient covers verbal comprehension and working memory, assessed with mostly verbal subtests, perhaps the cause of high correlation of the WAIS scores with PPVT scores. However, besides this, perceptual reasoning and processing speed are assessed with mostly non-verbal or visual subtests (Benson *et al.*, 2010). Hence, future research with more statistical power would be needed to determine which major components of intelligence are involved in the comprehension of interrupted speech.

All results combined, our data showed significant correlations between receptive vocabulary and verbal intelligence (PPVT-III-NL) and perception of interrupted speech, implying an involvement of these factors in top-down restoration mechanisms. This knowledge can be useful in developing future training paradigms for older or hearing-impaired people. These can, for example, focus on linguistically relevant training methods (such as crossword puzzles), and may be easier to implement compared to cognitive training methods.

Acknowledgments

The authors would like to thank Grietje Bos for valuable feedback on the study design, Jefta Saija for help with the PRAAT speech manipulations, Wydineke Boels for help with collecting data, and Anke van der Meijde, Age Hoekstra, Peter van Hengel, and Bas Franck for their valuable feedback on earlier versions of this manuscript. This study was supported by Pento Audiology Center Zwolle, VIDI Grant No. 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), Heinsius Houbolt Foundation, Rosalind Franklin Fellowship from University Medical Center Groningen, and is part of the research program of our department: Healthy Aging and Communication.

References and links

Akeroyd, M. A. (**2008**). "Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults," Int. J. Audiol. **47** Suppl. 2, S53–S71.

Bashford, J. A., Riener, K. R., and Warren, R. M. (**1992**). "Increasing the intelligibility of speech through multiple phonemic restorations," Percept. Psychophys. **51**, 211–217.

Başkent, D. (**2010**). "Phonemic restoration in sensorineural hearing loss does not depend on baseline speech perception scores," J. Acoust. Soc. Am. **128**, EL169–EL174.

Başkent, D. (**2012**). "Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation," J. Assoc. Res. Otolaryngol. **13**, 683–692.

Bell, N. L., Lassiter, K. S., Matthews, T. D., and Hutchinson, M. B. (**2001**). "Comparison of the Peabody Picture Vocabulary Test-Third Edition and Wechsler Adult Intelligence Scale-Third Edition with university students," J. Clin. Psychol. **57**, 417–422.

Benard, M. R., and Başkent, D. (**2013**). "Perceptual learning of interrupted speech," PLoS One **8**, e58149.

Benson, N., Hulac, D. M., and Kranzler, J. H. (**2010**). "Independent examination of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV): What does the WAIS-IV measure?," Psychol. Assess. **22**, 121–130.

Chatterjee, M., Peredo, F., Nelson, D., and Başkent, D. (**2010**). "Recognition of interrupted sentences under conditions of spectral degradation," J. Acoust. Soc. Am. **127**, EL37–EL41.

Saija, J. D., Akyürek, E. G., Andringa, T. C., and Başkent, D. (**2013**). "Perceptual restoration of degraded speech is preserved with advancing age," J. Assoc. Res. Otolaryngol. (e-pub November 7, 2013).

Schlichting, L. (**2005**). "Peabody Picture Vocabulary Test, 3rd ed. (PPVT-III-NL) Dutch version."

Sivonen, P., Maess, B., Lattner, S., and Friederici, A. D. (**2006**). "Phonemic restoration in a sentence context: Evidence from early and late ERP effects," Brain Res. **1121**, 177–189.

Verschuure, J., and Brocaar, M. P. (**1983**). "Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise," Percept. Psych. **33**, 232–240.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Wang, X., and Humes, L. E. (**2010**). "Factors influencing recognition of interrupted speech," J. Acoust. Soc. Am. **128**, 2100–2111.

Warren, R. (**1983**). "Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception," J. Audio Eng. Soc. **31**, 623–629.

Wechsler, D. (**2012**). "Wechsler Adult Intelligence Scale, 4th ed. (WAIS-IV-NL) Dutch version."