

Reversed Speech Comprehension Depends on the Auditory Efferent System Functionality

Grataloup C.¹, Hoen M.¹, Pellegrino F.¹, Veuillet E.², Collet L.², & Meunier F.¹

¹Laboratoire Dynamique Du Langage. UMR 5596 CNRS & Université Lumière Lyon2.
Institut des Sciences de l'Homme. Lyon – France.

Claire.Grataloup@etu.univ-lyon2.fr; Fanny.Meunier@univ-lyon2.fr

²Neurosciences & Systèmes Sensoriels. UMR 5020 CNRS & Université Claude Bernard Lyon1.
Lyon – France.

Abstract

In the present study we explore the implication of high and low level mechanisms in degraded (time-reversed) speech comprehension in normal hearing subjects. In experiment 1 we compared the loss of intelligibility due to the increasing size of reversion windows in both words and pseudowords. Results showed that words are generally reconstructed better than pseudowords, suggesting the existence of a lexical benefit in degraded speech restoration. Moreover, there was greater variability between individuals when reconstructing pseudowords than words. In experiment 2, we demonstrated that this interindividual variability correlated with the subjects' medial olivocochlear bundle functionality, as measured by contralateral suppression of otoacoustic emissions (OAEs). Together these experiments highlight the importance of low-level auditory mechanisms in degraded speech restoration. Moreover they put forward the existence of major interindividual variability in the capacity to reconstruct degraded speech, which correlates with the physiological properties of the auditory system (low-level property). In addition, our results also suggest the existence of multiple higher-level strategies that can compensate on-line for the lack of information caused by speech degradation.

1. Introduction

Understanding speech is fast and automatic; this process is a daily task achieved without any difficulty although it involves cognitive functions which are both numerous and complex. The intelligibility of speech depends both on the quality of the emitted signal and on the ability of the cognitive system to process this signal. To understand spoken messages requires the complementary engagement of low-level (auditory) and high-level mechanisms (e.g., lexical knowledge, contextual integration).

At the sensory level, the human auditory system may be described through two types of pathway: “ascending” and “descending” auditory pathways. Ascending pathways carry the auditory message from receptive hair cells in the cochlea to the primary auditory cortex. Descending pathways extending from the auditory cortex to the periphery through the superior olivary complex constitute a sort of inhibitory filter [1] which may play a role in degraded speech comprehension. The existence of a central influence on the sensory level by means of efferent feedback pathways has been highlighted with the investigation of the medial olivocochlear bundle (MOCB). It is possible to explore the MOCB in a simple and non invasive way through contralateral suppression of otoacoustic emissions (OAEs), sounds

produced by the outer hair cells of the cochlea [2]. Important interindividual variability in normal hearing subjects was also observed by the same authors when assessing the functionality of the MOCB. Though the precise functional role of the MOCB remains unclear, several studies have suggested its involvement in speech intelligibility in degraded conditions. A correlation between the functionality of MOCB and the detection threshold of pure tones present in noise was first reported by Khalfa and Collet [1]. Giraud and al. [3] also reported that MOCB activation via contralateral noise stimulation improved speech-in-noise intelligibility in normal hearing subjects. More recently Kumar and Vanaja [4] suggested a correlation between the contralateral suppression of evoked OAEs and speech identification scores at certain signal/noise ratios.

However speech comprehension is likely to benefit during auditory sentence comprehension from higher level knowledge such as lexical knowledge stored in the mental lexicon or contextual information. Interactive theories of speech perception postulate that several sources interact during perceptual analysis [5]. For example Ganong [6] and Warren [7] have established that phonemes identification is influenced by feedback from lexical and semantic levels. In their experiments listeners heard an ambiguous stimulus like “#eel” where the first phoneme is degraded. This item can be perceived and interpreted as the words *meel*, *wheel*, *heel* or *peel*. When the context was “The #eel was on the table” they subjects would preferably say that they heard “meel”. In such cognitive theories the perceptual analysis of speech is influenced by top-down feed-back from higher levels of processing. However the debate continues regarding the involvement of both cognitive and sensory processes during speech comprehension.

The aim of our study was to characterize involvement of high and low-level processes during the comprehension of speech in degraded conditions. To study these mechanisms we manipulated the intelligibility of speech signals by applying temporal reversions on different-sized windows (i.e., flipping the signal on its horizontal axis). Reversed speech has the particularity that it maintains the physical characteristics of speech such as the distribution frequency of sounds, their global amplitude and, to some extent, their temporal and rhythmic characteristics. The main difference between speech and reversed speech lies in the coarticulations which are totally distorted in the reversed signal. Thus, reversed speech is unpronounceable though to some extent it remains understandable. Saberri and Perrott [8] demonstrated that it is possible to reconstruct locally reversed English speech (for comparable results see Greenberg and Arai [9]). Data from our group first replicated these behavioural observations for

French [10]. All these studies applied time reversions according to arbitrarily augmenting time-windows (20 ms, 40 ms, 60 ms...). For the present work, time reversions were based on the syllable unit as it seems to be the privileged unit during lexical access in French [11]. In a first experiment, we measured the accuracy in restoring single words and pseudowords to quantify the lexical effects during this process. In a second experiment auditory measurements were made to test the implication of auditory pathways in the reconstruction phenomenon.

2. Experiment 1: behavioural measures

2.1. Materials and methods

2.1.1. Stimuli

Word stimuli consisted of 120 French nouns from the *Lexique* database [12]. All were disyllabic words and belonged to a common vocabulary. They were selected according to two main criteria: their frequency and their number of phonological neighbours. For example the word /ballon/ (balloon) is a very frequent noun that has many phonological neighbours (e.g., /vallon/, /baron/, /baton/...). We crossed Frequency and Neighbour factors in order to create 4 conditions of 30 experimental nouns each. For each condition the Table 1 gives an example followed by its frequency and the number of phonological neighbours.

Table 1: The four conditions of experimental words

	High Frequency	Low Frequency
Much Neighbours	<i>ballon</i> 21 (20)	<i>boulon</i> 20.5 (20)
Few Neighbours	<i>bijou</i> 20.3 (2)	<i>grelot</i> 2.8 (0)

We also constructed 120 experimental disyllabic pseudowords using the same syllables contained in the 120 experimental words. These pseudowords resulted from syllable switching and thus satisfied the constraints observed in French phonotactics. All 240 items were recorded in a soundproof room by a native French speaker and saved as Windows PCM files (22 kHz, mono, 16 bits). We applied five kinds of reversion to the items:

- Condition R_0 : no reversion applied.
- Condition $R_{0.5}$: first half syllable reversed.
- Condition R_1 : first syllable reversed.
- Condition $R_{1.5}$: first syllable and a half reversed.
- Condition R_2 : whole item reversed.

The half syllable was defined as the half in terms of the duration of a syllable.

2.1.2. Experimental procedure

The experiment took place in a quiet room and lasted for about 50 minutes. Participants faced a computer screen (PC type) and heard the stimuli delivered binaurally via headphones (Beyerdynamic DT 48, 200 Ω). The presentation order of the 120 word stimuli was randomized across subjects as presentation order of the 120 pseudowords stimuli. The participants had to type on a computer keyboard what they had understood. They were told whether they had to reconstruct words or pseudowords though two different types of instructions.

2.2. Participants

50 students (22f, 28m) native French speakers aged between 18 and 25 participated in this study. None had any auditory

problems. They were all volunteers and were paid 7.5 € for their participation.

2.3. Results

Results were obtained by comparing the subjects' transcriptions to the original words or pseudowords. Taken as a whole, the percentage of intelligibility decreased with increasing distortion. The average reconstruction score was 97.3% for condition R_0 ; 79.65% for $R_{0.5}$; 51.15% for R_1 ; 2.85% for $R_{1.5}$ and 1.81% for R_2 . Figure 1 shows intelligibility rates for words and pseudowords as a function of the reversion size.

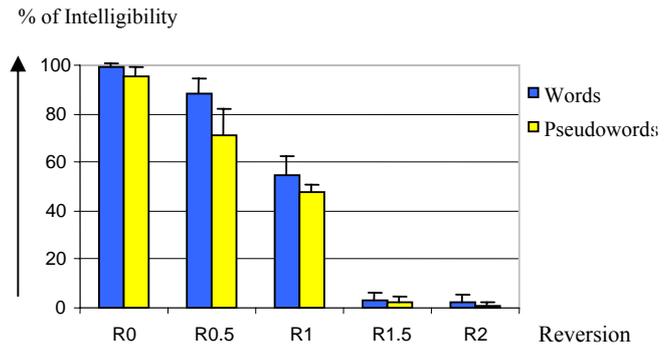


Figure 1: Rate of intelligibility for words and pseudowords plotted against the size of the reversion.

The subjects performed better overall with words than with pseudowords ($F(1,49)=158.5$; $p<.0001$). For both categories of stimuli the degradation of the first half syllable ($R_{0.5}$) gave high scores of restoration (average 79.65%). When a little part of the first syllable is degraded the cognitive system is able to reconstruct the signal. When only the first syllable is degraded scores turned around 50% as the second syllable was well identified. However, the degradation of one syllable and a half ($R_{1.5}$) gave disastrous reconstruction scores (below 3%). When the distortion is longer than the syllable the system seems to be unable to reconstruct neither the first syllable nor the second one which is only partly damaged. This tends to confirm the main role of the syllable as a unit in French comprehension processes.

Comparing words and pseudowords in condition R_1 (first syllable degraded) we observed that the first syllable is sometimes reconstructed only in words (so the whole word is found): 54.8% of words are reconstructed. In pseudowords the first syllable is almost never reconstructed (47.5% of reconstruction) moreover the degradation disturbs the well understanding of the second one (though it is intact).

In addition we observed that for pseudowords restitution participants tend to answer a real word phonological similar to the target instead of the right pseudoword. In average participants answer by a word for 3.65% ($SD= 3.76$) of the 120 pseudowords. For example the word "parfum" was frequently given as an answer of the target stimulus "rafin". This result shows the robustness of lexical effects in a non lexical task such as pseudowords restoration.

Furthermore we observed both a frequency effect ($F(1,49)=17.12$; $p=.0001$) and a neighbour effect ($F(1,49)=20.15$; $p<.0001$) on word reconstruction. The most frequent words were better reconstructed whatever the inversion degree was and words with fewer neighbours were also better reconstructed. The interaction between Frequency and Neighbour was significant only for condition R_1 . This interaction indicates that when the first syllable is distorted the

number of neighbours is the first criterion that could modify the reconstruction process. A high frequency could thus help the reconstruction but only for words with few phonological competitors.

The fact that pseudowords were more difficult to reconstruct than words seems coherent because pseudowords do not have any representation stored in the mental lexicon. Consequently, they do not benefit as much from lexical help. Overall performances for pseudowords were very homogenous over participants for most conditions (see Figure 2). However the condition $R_{0.5}$ showed a large interindividual variability. In this condition, some subjects were undisturbed by the inversion whereas others were deeply perturbed and failed the reconstruction task: performances ranged from 47.9% to 91.7% of correct reconstruction (mean=71.2%, SD=10.4).

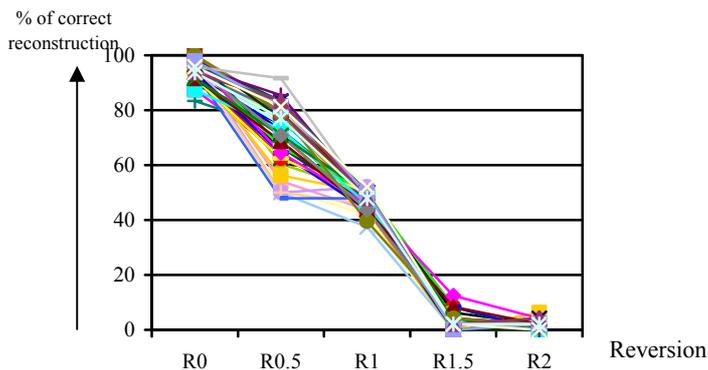


Figure 2: Reversed pseudowords reconstruction. Each curve represents the results of one subject.

Pseudowords are considered as speech and thus general phonotactics apply (giving cues to the identity of the contiguous phonemes) but they do not have any representation stored in the mental lexicon. Consequently, reversed pseudowords reconstruction is primarily based on auditory information. So why did the subjects – all young and without any auditory problems – present such varied performances? Given that descending auditory pathways could interfere in speech perception and that their contribution is unequal among subjects, we can formulate the hypothesis that the MOCB functionality may be responsible for the behavioural results we observed. In order to answer this question we compared the results obtained for pseudoword reconstruction and auditory performance in each participant.

3. Experiment 2: Auditory measurements

Two groups of 10 persons each were then formed from the subjects in experiment 1 according to their performance for pseudowords reconstruction in condition $R_{0.5}$ (i.e., the condition showing the largest variability). The HP group (High Performance) was composed of the 10 persons who showed the best performances (m=83.3%). The LP group (Low Performance) was composed of the 10 subjects who showed the lowest performances (m=50%). We measured the functionality of the MOCB of these selected subjects (uncrossed pathway) in a soundproof room. OAEs were recorded according to the method of Bray and Kemp using the Otodynamics ILO88 measuring device. A miniaturized microphone placed in the external ear canal delivered acoustic stimulations (clicks) and recorded responses. For each participant we recorded OAEs for 20 ms after clicks delivery with and without a broadband noise (30 dBSL) applied in the contralateral ear. We calculated the contralateral suppression

of OAEs in each ear and the corresponding lateralization (for details see Veuille et al. [13]). The tests lasted one hour.

3.1. Participants

The 20 participants selected from experiment 1 were all volunteers. They all had normal peripheral hearing (thresholds better than 20 dB HL between 250 Hz and 8000 Hz) and a normal tympanometry (stapedian reflex greater than 65 dB HL). The subjects had no history of otological or neurological disorders and all were right handed (more than 80% according to the Edinburgh test). They were paid 20 € for their participation.

3.2. Results and Discussion

The results are clear cut: the HP group showed better contralateral suppressions in both ears than did the LP group. Table 2 shows the mean and the standard deviation of contralateral suppressions of OAEs in the right and left ears respectively, for each group of subjects. The last column indicates the lateralization (Contralateral suppression of OAEs on right ear minus left ear).

Table 2: Contralateral suppressions of OAEs

	Right ear	Left ear	Lateralization
HP Group	-4.2 (1.1)	-2.7 (1.8)	-1.5 (2)
LP Group	-1.4 (2)	-1.8 (1.3)	0.4 (1.6)

The ANOVA analysis ran on lateralisation showed a significant effect of the Group factor ($F(1,9)=7.64$; $p=.022$ see Figure 3). According to these results subjects of the HP group were more lateralized on the right ear than subjects of the LP group.

Lateralization
(Right ear –left ear)

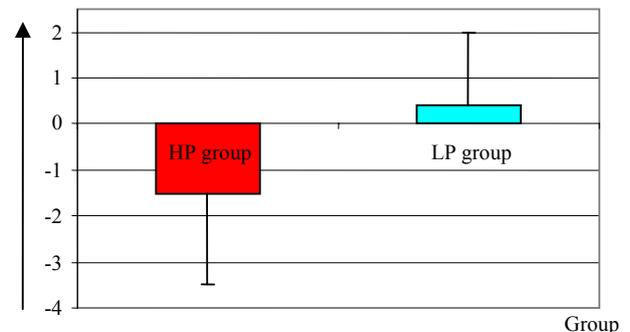


Figure 3: Lateralization for the two groups of subjects.

Indeed auditory areas are asymmetric and the peripheral auditory system reflects this asymmetry. An efficient MOCB is characterised by a large inhibitory power and by the lateralization (for right handed people, the more negative the lateralization the more efficient the MOCB).

Auditory measurements demonstrated clearly that the HP group had a better MOCB and was more right lateralized than the LP group. We found a correlation between behavioural performances and the lateralization of the subjects ($r=0.7$; $p<.001$). It suggests that the stronger the asymmetry of the auditory system the higher the behavioural performances. This confirms the participation of MOCB in cognitive processes of reversed speech restoration. The role of the MOCB would be to filter the damaged signal in order to highlight perception of those elements which are pertinent. The MOCB would be a

sort of adapting mechanism of the ear to situations in which the signal is perceived. Together, experiments 1 and 2 throw light on the participation of high and low-level mechanisms in reversed speech restoration.

4. General discussion

The aim of the study was to understand the respective influence of high and low-level mechanisms that underlie the cognitive reconstruction of degraded speech. Our results suggest that descending auditory pathways may intervene significantly during this process: more precisely, activity in MOCB may be one of the neurophysiological mechanisms participating in reversed speech restoration. In addition, high-level knowledge is activated to help the reconstruction. Speech comprehension is the result of an interaction between low-level mechanisms (phonological reconstruction) and elaborated knowledge stored in the mental lexicon. In experiment 1 we saw that the system was able to reconstruct the speech signal to some extent. Reversed words and pseudowords are not reconstructed in the same way. For words both high and low-level mechanisms intervene, often resulting in successful comprehension. Both phonological neighbours and the frequency of noun effects suggest the implication of high-level mechanisms (lexical search) in word reconstruction. These mechanisms may either improve the signal reconstruction in a top-down scheme or more probably compensate for the lack of information if the low-level mechanisms fail to reconstruct the signal.

With the pseudowords, high-level knowledge is of no help as the target pseudowords are not listed in the mental lexicon. However, high-level mechanisms are still involved since subjects tend to reconstitute a real word phonologically similar to the target pseudoword. In addition, our results showed a syllable effect. As long as the damaged information is shorter or equal to one syllable the cognitive system still understands, even if the target item exhibits numerous phonological neighbours. When the distortion gets larger than one syllable, the reconstruction becomes more difficult (at least for disyllabic words in French). These results are consistent with the hypothesis that the syllable could be a perceptual unit in French. To extend our knowledge it would be interesting to test the syllable effect for disyllabic words in a language such as English in which rhythm is not based on the syllable.

Another interesting finding is the large variability observed between participants in pseudoword reconstruction. For the reversion of the first half syllable some of the participants had no trouble in reconstructing pseudowords whereas others largely failed. This variability occurred between normal hearing persons. Moreover, behavioural results and MOCB functionality of the subjects correlate significantly (experiment 2). This suggests implication of the efferent auditory pathways in reversed speech reconstruction. Reversed speech intelligibility seems to be linked with functioning MOCBs. The MOCB seems to modify the properties of auditory fibres in order to improve the ability of the auditory system to focus on pertinent information included in the percept. However, we did not find any correlation between auditory measurements and word reconstruction. The variability may fade during word reconstruction because of the intervention of lexical strategies which leads us to speculate on the importance of lexical strategies as a consequence of poor functionality of the MOCB. In a further study it would also be interesting to test whether lexical strategies are more important to people with hearing impairments.

5. Conclusions

Speech reconstruction is the result of rapid, efficient activation of several complex mechanisms. Our experiments show that the respective involvements of the high and low-level mechanisms depend on the nature of the stimuli: pseudoword stimuli highlight the major role of the MOCB while word stimuli emphasize lexical strategies. Further experiments will be necessary to precisely determine the role of each mechanism, and in the short term the extension of this experiment to hearing impaired patients is planned since it could lead to the development of new hearing aids.

6. Acknowledgments

This study was supported by an ACI (n° 67068) of the French *Ministère de l'Enseignement Supérieur et de la Recherche* attributed to Fanny Meunier.

7. References

- [1] Khalfa, S. and Collet, L. (1996). Functional asymmetry of medial olivocochlear system in humans. Towards a peripheral auditory lateralization. *NeuroReport* **7**, 993-996.
- [2] Collet, L., Kemp, D. T., Veuillet, E., Duclaux, R., Moulin, A., et Morgnon, A. (1990). Effect of contralateral auditory stimuli on active cochlear micro-mechanical properties in human subjects. *Hearing Research* **43**, 251-262.
- [3] Giraud A.L., Garnier S, Micheyl C., Lina-Granade G, Chays A., Chery-Croze S.(1997). Medial Olivocochlear Efferent Involved In Speech In Noise Intelligibility. *Neuroreport*, **8**, pp. 1779-1783.
- [4] Kumar U. A. and Vanaja C. S. (2004). Functioning of olivocochlear bundle and speech perception in noise. *Ear and Hearing*. **25**(2):142-6.
- [5] McClelland, J.L. & Elman, J.L. (1986) The TRACE model of speech perception. *Cognitive Psychology* **18**:1-86.
- [6] Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology*, **6**, 110-125.
- [7] Warren, R.M. (1970). Restoration of missing speech sounds. *Science*, **167**, 392-393.
- [8] Saberi, K. and Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, **398**, 760.
- [9] Greenberg, S. & Arai, T. (2001). The relation between speech intelligibility and the complex modulation spectrum, Proceedings of the 7th *Eurospeech Conference on Speech Communication and Technology (Eurospeech-2001)*, pp. 473-476.
- [10] Meunier, F., Cenier, T., Barkat, M. & Magrin-Chagnolleau, I., (2002). Mesure d'intelligibilité de segments de parole à l'envers en français, *Actes des 24ème Journées d'étude sur la parole (JEP)*, p.117-120, Nancy, France.
- [11] Mehler, J., Segui, J., Frauenfelder, U. (1981). The role of the syllable in language acquisition and perception. In T.F. Myers, J. Laver, J. Anderson (Eds), *The Cognitive Representation of the Speech*, Amsterdam, North Holland.
- [12] New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur Internet : LEXIQUE. *L'Année Psychologique*, **101**, 447-462.
- [13] Veuillet E, Collet L, Bazin F. Objective evidence of peripheral auditory disorders in learning-impaired children. *Journal of Audiological Medicine*, 1999, **8**(1): 18-29.