Auditory language comprehension of temporally reversed speech signals in native and non-native speakers

Miklos Kiss[1], Tamara Cristescu[2], Martina Fink[1], and Marc Wittmann[1,3]*

[1]Generation Research Program, Ludwig-Maximilian University Munich, Bad Tölz, Germany; [2]Institute of Medical Psychology, Ludwig-Maximilian University Munich, München, Germany; [3]Department of Psychiatry, University of California San Diego, La Jolla, CA ,USA, *Email: wittmann@ucsd.edu

Neuropsychological studies in brain-injured patients with aphasia and children with specific language-learning deficits have shown the dependence of language comprehension on auditory processing abilities, i.e. the detection of temporal order. An impairment of temporal-order perception can be simulated by time reversing segments of the speech signal. In our study, we investigated how different lengths of time-reversed segments in speech influenced comprehension in ten native German speakers and ten participants who had acquired German as a second language. Results show that native speakers were still able to understand the distorted speech at segment lengths of 50 ms, whereas non-native speakers only could identify sentences with reversed intervals of 32 ms duration. These differences in performance can be interpreted by different levels of semantic and lexical proficiency. Our method of temporally-distorted speech offers a new approach to assess language skills that indirectly taps into lexical and semantic competence of non-native speakers.

Key words: Temporal processing, bilingualism, first and second language processing

INTRODUCTION

Our experience of time is based on the two main concepts of succession and duration (Fraisse 1984, Wittmann 1999). Whereas perception of succession refers to the sequential characteristics of events, that is, their temporal order (Fink et al. 2005), perception of duration refers to the time interval between two events or the persistence of an event over time (Wackermann and Späti 2006). For humans it has been shown that the perception of the temporal order of acoustic events is only possible when the events are separated by inter-stimulus intervals (ISI) of about 20–60 ms (e.g. Hirsh 1959, Pastore and Farrington 1996, Lotze et al. 1999, Kanabus et al. 2002, Fink et al. 2005). Temporal order thresholds have been assessed over different sensory modalities, with varied stimuli (clicks, tones, noises), varied psychophysical procedures, different presenta-tion modes (monaural, binaural), and varied participant characteristics (trained vs. naive participants, age of the participants). The studies show that auditory temporal-order thresholds converge at the mentioned values of some tens of milliseconds (Hirsh and Sherrick 1961, Mills and Rollman 1980, Fink et al 2006a). These converging results have been discussed as point-ing to a central-timing mechanism responsible for temporal-order judgement (Pöppel 1997). It has been assumed that this mechanism creates basic temporal units of perception and provides discontinuous pro-cessing of information based on neuronal oscillations in the 40-Hz range (Pöppel 1970, Joliot et al. 1994). A neuronal oscillation is initiated every time a stimulus is processed in a specific modality. These oscillations have a period of about 30 ms and each of them repre-sents one processing unit. It is assumed that within these processing units, incoming events are treated as co-temporal (Pöppel et al. 1990), meaning that the temporal order of sensory events cannot be indicated when the two events occur during one period of such an oscillation.

Correspondence should be addressed to M. Wittmann, Email: wittmann@ucsd.edu

The speech signal contains information on different time scales (Rosen 1992) and, specifically, the processing of temporal order has been discussed as underlying speech and language processing (Tallal et al. 1998, Wittmann and Fink 2004, Wittmann et al. 2004). For example, information about the place of articulation in stop consonants is contained in the time domain of about 20–40 ms. Formant transitions, characterized as short sound waveforms that change frequency across a time interval of ca. 40ms vary according to the place of articulation. These spectral changes at the release of the closure depend on the articulatory structure that is used to form the constriction (labial /b/, /p/; alveolar /d/, /t/; velar /g/, /k/) (Stevens 1998). Moreover, voiced stop consonants (/b/, /d/, /g/) are distinguished from voiceless stop consonants (/p/, /t/, /k/) by a difference in duration of about 20 ms in the time between the burst and the onset of laryngeal pulsing, defined as the voice-onset time (VOT) (Stevens 1998). Neurophysiological studies have shown the critical role of the length of the voice-onset time in the approximate time range of the order threshold (Simos et al. 1998, Ackermann et al. 1999). In studies using implanted electrodes in awake monkeys, the syllable /da/ with VOTs of 0 and 20 ms elicited one response in the primary auditory cortex; with the syllable /ta/ (VOTs of 40 and 60 ms), two such responses were recorded (Steinschneider et al. 1995). The authors concluded that the temporal-response pattern represented a physiological correlate for categorical phonetic perception that relied exclusively on the interval between the onset of the plosive and the onset of the periodic portion of the syllable. Numerous neuropsychological studies with brain-injured, aphasic patients and children with language-learning deficits have emphasized the association of language comprehension with auditory temporal-processing abilities (Swisher and Hirsh 1972, Tallal and Piercy 1973, Farmer and Klein 1995, Steinbüchel et al. 1999, Wittmann et al. 2004, Fink et al. 2006b). Subjects with language impairments showed significantly elevated mean temporal-order thresholds and problems in discriminating phonemes that are critical in the time domain.

Since the stimulus properties of speech stimuli seem to be segmented into sequential processing units of 20 to 40 ms, temporal information within a segment of a speech sound may not be relevant for decoding. This hypothesis was tested by manipulating speech signals that where presented to subjects (Steffen and Werani 1994, Steffen et al. 1997). Recorded speech signals were cut into different segment sizes of 20 to 70 ms at multiples of 10 ms and then time reversed within an interval without changing the order of the segments. The temporally-distorted speech stimuli were then presented to participants, starting with the longest segment of 70 ms and working gradually down to 20 ms. Understanding of the sentences improved step by step with shorter time-reversed segments. With reversed segments of 20 ms, comprehension approached the level of normally presented speech. The authors interpreted this result as a further indication that temporal information in the speech signal below units of 20 ms does not contribute to the understanding of spoken sentences. A further investigation confirmed that time-reversed speech signals are sensitive to segment length (Saberi and Perrott 1999). Intelligibility of spoken sentences appeared to be a function of the interval length of the reversed speech segments. Up to segment durations of 50 ms, comprehension was almost 100%. Partial intelligibility, however, was still preserved up to segment durations of 130 ms. This finding is unexpected, because such a long reversed interval should normally distort comprehension almost completely. This latter result confirms the notion that speech comprehension is not only dependent on the physical properties of the signal, as would be suggested by a purely bottom-up model. Bottom-up processing is complemented by top-down processes, such as by the activation of lexical and semantic information. In order to select the appropriate meaning from a variety of possibilities, listeners have to integrate the environmental context with their general knowledge. Therefore, in the experiments with the reversed speech paradigm, the remaining phonologic cues in the distorted speech signal can be complemented by semantic knowledge.

In pilot tests related to earlier experiments investigating the reversed-speech paradigm (Steinbüchel et al. 2000, Kiss 2002), we noticed that participants who had acquired German as second language performed worse in understanding the temporally distorted German sentences than native German speaking participants. This finding was challenging for two reasons: the subjects did not have a temporal-processing deficit, and their comprehension of the non-reversed sentences was 100%. When not distorted, these same sentences were easily understood by the subjects with German as second-language and good German language proficiency. Presented stimuli such as "Turn your head to the right"

were short and consisted of frequently employed words and simple grammar. In the study presented here, we intended to systematically test whether participants who had acquired German as a second language, thus mastering the grammatical, lexical, and semantic knowledge of simple spoken German sentences, nevertheless experience more difficulties in understanding temporally distorted speech signals than native German speakers.

METHODS

Based on earlier findings by Steffen and Werani (1994) suggesting a new method for inverting speech, natural speech was modified by the reversal of its temporal structure within segments. The temporal information of the speech signal was thus destroyed, although the spectral information was still available.

Table I

Sentences

Semantically coherent sentences

1. *Lege den gelben Stein auf den roten*
(Put the yellow stone on the red one)
2. *Drehe den Kopf nach rechts*
(Turn your head to the right)
3. *Lege die Kugel auf das Tuch*
(Put the ball on the towel)
4. *Greif dir an die Nase*
(Touch your nose)
5. *Schlage das Buch auf.*
(Open the book)

Semantically incoherent sentences

6. *Der Keller auf dem Dachboden friert*
(The cellar on the roof is freezing)
7. *Es grünt so grau der Morgenmond*
(The morning moon greens so grey)

Pseudo-homophonic sentences

8. Pimafen rotel mipeln gausend
9. Schnorpel blackt mulig

German sentences used in the study. English translations are shown in brackets.

Material

German semantically coherent, German semantically incoherent, and German pseudo-homophonic sentences were spoken by a male speaker and recorded digitally (see Table I). Semantically coherent sentences were instructions like "Turn your head to the right". In these sentences lexical and contextual cues are included. Semantically incoherent sentences were sentences with real words that made no sense like "The cellar on the roof is freezing". In this category of sentences only lexical cues are presented. Contextual cues, however, are not included. Pseudo-homophonic sentences were phonologically correct
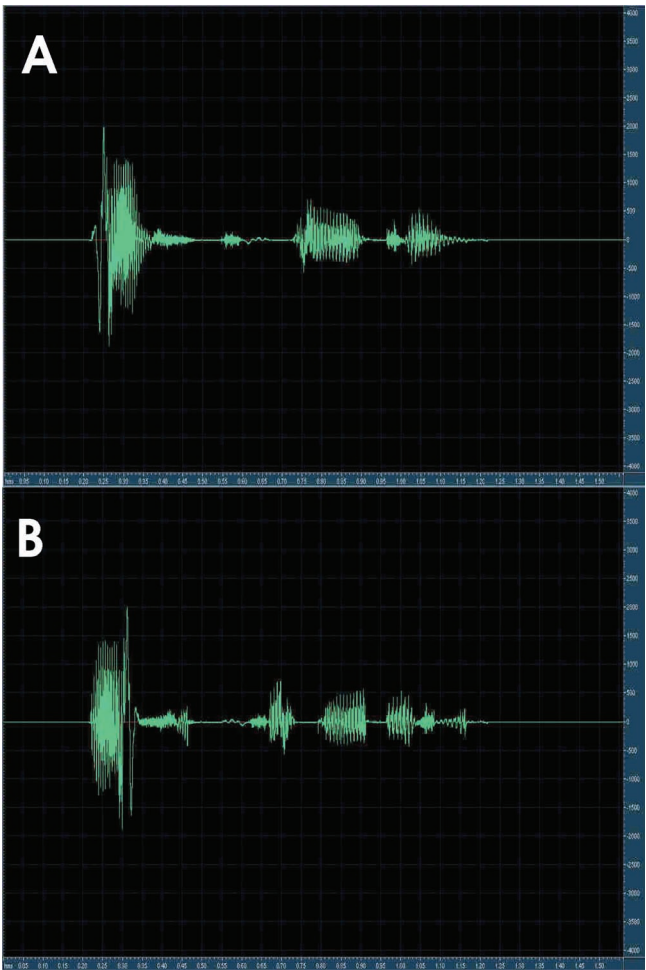

Fig. 1. (A) The power line of the German word /*Postbote*/ as a natural speech signal. The x-axis is the time axis; the y-axis shows the sound pressure. (B) The power line of the German word /*Postbote*/ as a temporally inverted speech signal (inverted segment = 122 ms). The x-axis is the time axis; the y-axis shows the sound pressure.

according to the phonotactic rules of German, but without meaning. In contrast to the other two sentence types no lexical or contextual cues were embedded in these sentences. Natural speech signals were digitally recorded in a sound-proof room. The recording was set at 11-MHz sampling rate, which was up-sampled to a 44-MHz sampling rate for audio CD recording. The material was segmented into equal intervals of certain lengths for each sentence with the software "Cool Edit Pro". The speech signal within each interval was then turned around. After this procedure, the former end of a segment formed the beginning of the new segment. Nine different sentences were inverted with eight different levels of distortion per sentence. The intervals were: 122 ms, 98 ms, 74 ms, 50 ms, 38 ms, 26 ms, 14 ms, and 0 ms. These specific interval lengths appeared to produce the lowest number of artifacts when cutting and inverting the speech signals. The difficulty of the tasks decreased with decreasing interval length. An interval of 0 ms means no distortion of the signal at all. Taken together, nine sentences were inverted with eight different levels of distortion, resulting in 72 stimuli. To get an undisturbed signal with as few click artifacts as possible, we allowed a jitter of $\pm$ 2 ms in each interval to be able to cut the signal in the zero crossing of the wave-line. With the example of the German word "*Postbote*" (postman), one can see

the original signal of the word (Fig. 1A) and the wave-line of the word after inverting with intervals of 122 ms (Fig. 1B). In this way, spectral information is kept intact. The spectral signal is just moved to a different position in the word.

An effect of our procedure is that certain physical properties of the speech signal are separated by twice the interval length. Other physical characteristics are not separated at all, as illustrated in Fig. 2. This important aspect was overlooked in the aforementioned studies (Steffen and Werani 1994, Saberi and Perrott 1999). An interval of 14 ms, for example, has a temporal distortion of 0 ms at the center of each interval and a distortion of 28 ms at the outside margins of each interval (Fig. 2).

Participants

Two groups of subjects were tested. One group were native German speakers (n=10; 6 females), and the other group consisted of participants that had acquired German as a second language during adolescence and adulthood (n=10; 8 females). The native languages of the latter group belonged to the most important Indo-European language families: Germanic, Romanic and Slavic. The most common other foreign languages spoken (other than German) were French and English. The duration that participants had been
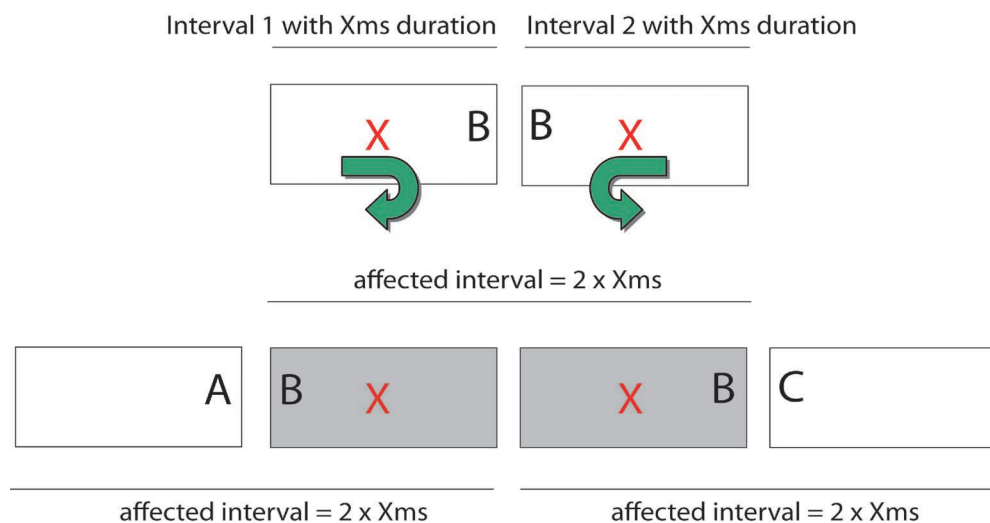
The effect of reversing the speech segments

Fig. 2. "Turning around" two intervals of X ms. Through this inversion of a segment, the speech signal is distorted in a time range between 0 times X ms and 2 times X ms. The border "B" moves 2 times X ms, but the center "X" does not move at all.

learning German ranged between 2 and 18 years with an average of 6.5 years and a standard deviation of 4.9 years. Subjects were only included if they had no known hearing problems and no history of hearing problems.

The groups were matched as closely as possible according to chronological age and educational level. The native German speakers ranged in age from 22 to 45 years (mean = 25.9 and standard deviation = 7.29). The non-native speakers ranged between 21 and 48 years of age with an average of 25.6 years (standard deviation: 8.17 years). All subjects were healthy volunteers who were not paid for their participation. The non-native speakers all lived in Germany. For more details about the participants, see Table II.

Procedure

The sentences were played to the participants *via* a SONY CFD-V10 CD player. The sound pressure level was held constant for each subject during the study. Each stimulus was presented twice with three short beeps at the start to focus the subjects' attention. Between the first and the second presentation of the same sentence there was a 15-second pause for the subjects to write down what they had heard. After the second presentation of the sentence there was a 20-second pause for writing and preparing for the next stimulus. If they had only partially understood the sentence, they were asked to only write down the parts they had understood. To accurately identify speech-reversed language material, three types of stimuli were presented: 5 semantically coherent sentences, 2 semantically incoherent sentences, and 2 pseudo-homophonic sentences. All sentences were presented with reversed segments starting with the longest interval having a duration of 122 ms. Then the following segment durations were presented in the descending order of 94 ms, 74 ms, 50 ms, 38 ms, 26 ms, 14 ms, and 0 ms over all sentence types. The sentences of one sentence type appeared in random order for each interval.

Statistical analysis

The general performance between groups was compared using the Mann-Whitney *U*-Test applying separate comparisons for the semantically coherent, semantically incoherent, and pseudo-homophonic sentences. Analyses within each group for differences between the sentence types were carried out using the Wilcoxon test. Two types of measures are compared: (1) Percent of words correctly understood in a given sentence and (2) the duration of the first segment interval at which the sentence was completely understood by an individual subject.

Table II

Characteristics of the non-native-speakers group

No.	Age	Native Language	No. of years speaking German	Other Foreign Language Proficiency
1	21	Italian	2	English
2	22	Italian	2	English
3	23	Czech	4	English, French
4	21	Slovak	6	English, Spanish
5	24	Russian	5	English
6	23	Croat	4	
7	28	Spanish	4	English
8	25	Norwegian	10	English
9	21	English	10	
10	48	English	18	Spanish

The table shows age, native language, and the number of years individuals were learning German separately for each participant in the non-native-speaker group.

RESULTS

Figures 3 to 5 show non-native and native speakers' performance in speech comprehension for all three types of sentences. For the coherent and incoherent sentence types, the performance (percentage of correctly detected words in a given sentence) of the German native speakers approached 100% comprehension earlier (already at longer intervals) than of the non-native group (see statistics below). When collapsing the performance over all reversed interval lengths, native language subjects identify a higher percentage of the words in the coherent sentences ($P<0.001$), incoherent sentences ($P<0.012$), and the pseudo-homophonic sentences ($P<0.023$). Group differences for individual inversion intervals in the coherent sentences were significant for durations of 38 ms (0.001), 50 ms (0.001), 74 ms (0.003), as well as 122 ms ($P<0.03$) (see Fig. 3). Group differences in the incoherent sentences were significant for durations of 38 ms (0.024) and 50 ms (0.001) (Fig. 4). In the pseudo-homophonic sentences only the 38-ms interval ($P<0.023$) showed a significant difference between groups (Fig. 5).
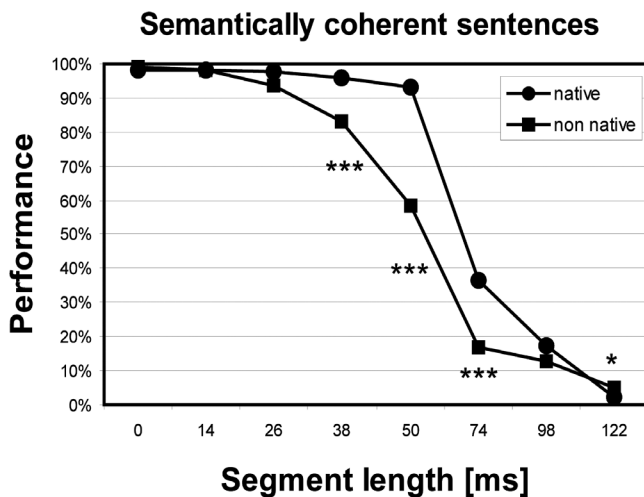
Semantically incoherent sentences

Fig. 4. Performance of speech comprehension for reversed speech material with semantically incoherent sentences. The x-axis shows the inverted segment length, and the y-axis indicates the performance in percent of words understood correctly. The performance of the non-native speakers improves at shorter segment lengths as compared to the native-language speakers. Performance differences between groups are indicated separately for each inversion interval by asterisks (***$P<0.001$, *$P<0.05$).

Semantically coherent sentences

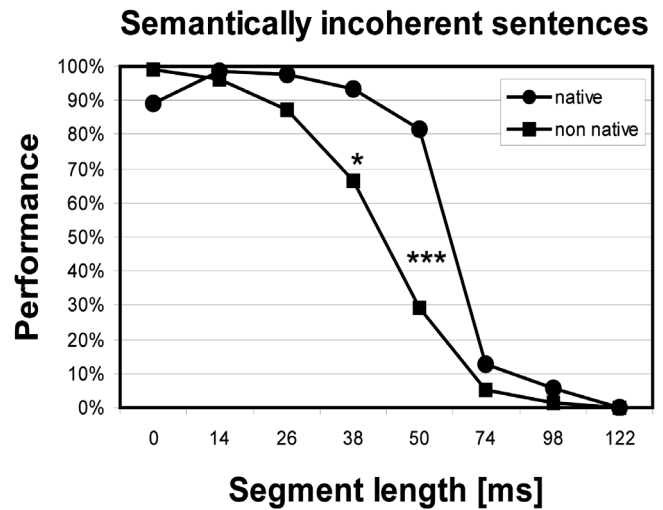

Fig. 3. Performance of speech comprehension for reversed speech material with semantically coherent sentences. The x-axis shows the inverted segment length, and the y-axis shows the performance in percent of words understood correctly. The performance of the non-native speakers improves at shorter segment lengths as compared to the native-language speakers. Performance differences between groups are indicated separately for each inversion interval by asterisks (***$P<0.001$, *$P<0.05$).
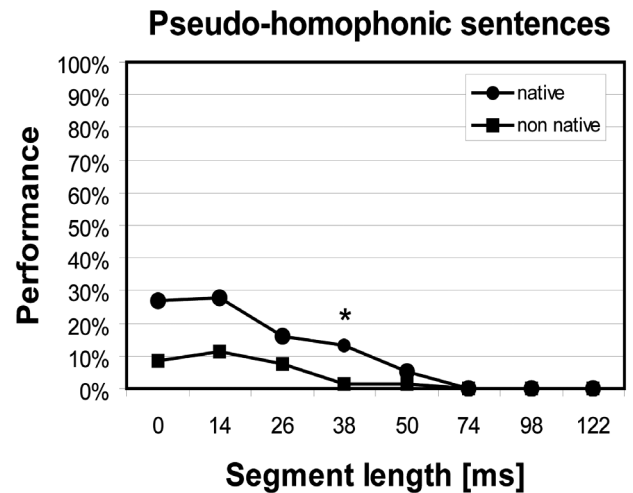
Pseudo-homophonic sentences

Fig. 5. Performance of speech comprehension for reversed speech material with pseudo-homophonic sentences. The x-axis shows the inverted segment length, and the y-axis indicates the performance in percent of words understood correctly. Native speakers only achieved a 30%-correct performance level in the non-disturbed signal (0 ms inversion) condition. Performance differences between groups are indicated separately for each inversion interval by asterisks (*$P<0.05$).
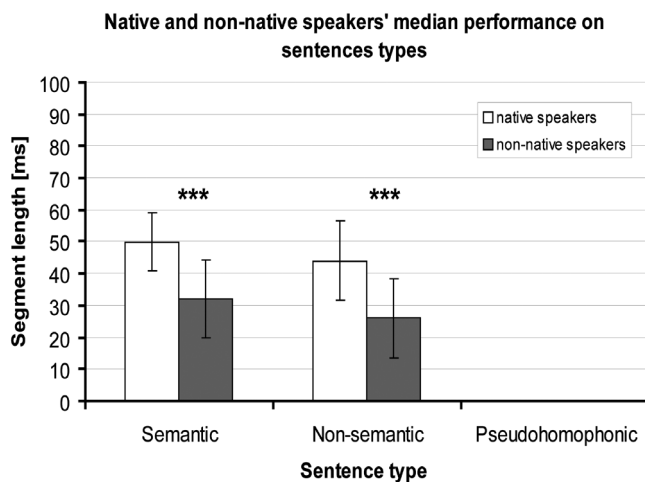
Native and non-native speakers' median performance on sentences types

Fig. 6. Significant performance differences (***P<0.001) between the two groups for the different types of sentences. Bars show the median duration of the group for the first segment interval at which the sentence was understood completely.

Figure 6 shows the performance of the two groups for each sentence type when taking as a variable the duration of the first interval length at which a sentence was completely understood. The median duration of the reversed interval at which native speakers could detect 100% of the semantically coherent sentences was 50 ms (as compared to 32 ms for non-native speakers). Native speakers completely understood semantically incoherent sentences with median reversion intervals of 44 ms (as compared to 26 ms for non-native speakers). The two groups differed significantly in the semantically coherent sentences (P<0.001) as well as in the semantically incoherent sentences (P<0.001). In the pseudo-homophonic sentences no participants in either group were able to understand 100% of the speech signal correctly, even with the non-distorted original sound (0 ms).

Within-group comparisons between the understanding of semantically coherent and semantically incoherent sentences also revealed a significant difference. For the non-native group, there was a difference between sentence type (P<0.047), revealing better comprehension for semantically coherent sentences. In the group of native speakers, there were no difference between the sentence types (P<0.661).

DISCUSSION

The performance between native and non-native speakers to understand temporally reversed speech signals differs to a varying degree for all the used sentence types. The performance of subjects who acquired German as second language started to improve at shorter inversion intervals as compared to the German native language group, where individuals could understand sentences at longer inversion intervals. Moreover, when taking as variable the duration of the first interval length at which a sentence was completely understood, the within group comparisons showed that non-native speakers were better at understanding semantically coherent sentences than incoherent sentences. In this variable, the group of native speakers showed no difference between these two sentence types.

Speech processing is thought to be strongly dependent on the temporal resolution of the auditory system (Pastore and Farrington 1996, Pöppel 1997, Berwanger et al. 2004, Szelag et al. 2004). If temporal-processing abilities in the range of some tens of milliseconds were important for speech comprehension on the bottom-up level, then the detrimental effect of inverting intervals of the speech signal in our experiment should already be found at intervals of 15 ms (in these segments certain features are separated through the inversion of the interval by 30 ms – see Fig. 2). However, our results show that native speakers can reliably identify semantically coherent sentences even at 50-ms intervals (where elements are distorted up to 100 ms). In semantically incoherent sentences, participants have complete understanding with distorted intervals up to 44 ms. Our German subjects tolerated much longer inversion intervals than expected, a result that complements earlier findings (Saberi and Perrott 1999). This effect can be explained in terms of top-down processing mechanisms. Studies have shown the influence of lexical and contextual information on phoneme categorization (e.g. Ganong 1980). The so-called "lexical effect" demonstrated in studies, means that the identification of a phoneme is affected by the lexical status of the spoken word in which the phoneme occurs. For example, an ambiguous item that is neither a /d/ nor a /t/ will be perceived more often as a /d/ when it is followed by /uke/ resulting in a real word (/duke/) in comparison to a nonword, like /tuke/ (Burton et al. 1989). But not only lexical information can be used to identify phonemes. Studies have shown that the sentence context influences phoneme identification as well (e.g., Borsky et al. 1998). Therefore, it can be assumed that the native

speakers in our study used their lexical and contextual knowledge as native language speakers to decode the speech signal, knowledge that to this extent is naturally not available to non-native speakers. Top-down mechanisms, thus, complement the analysis of acoustic-phonetic features of the speech signal and, therefore, also sentences with inversion intervals above 15 ms can be identified. A similar effect can be seen when acoustic information is distorted such as by noise (Warren 1970). In a noisy environment listeners can automatically add the missing element for identification of the speech signal.

Results of the pseudo-homophonic sentence type showed that no one in either group could correctly identify 100% of the words, even at 0 ms inversion (the non-distorted speech signal). In order to interpret this result it is helpful to look at the logogen model of word recognition (Morton 1969). In this model of word recognition, it is assumed that each word has a exical entry (Morton 1969). While recognizing a word, the perceptual input has to be matched with the phonological specification of the lexical entry. If there is a match, the lexical entry gets activated. According to the logogen model, a non-semantic route could also be used for pseudo-homophonic word recognition. This process starts with an acoustic analysis and a following acoustic to phonological conversion. The phonological form is then stored in a response buffer and translated into a graphemic form using phoneme-grapheme correspondence rules. This form is then stored in the graphemic output buffer before it is written down. Deficient processing can occur at several stages of this processing route and lead to an incorrect understanding. One complicating factor in understanding pseudo-homophonic sentences is that in spoken language word boundaries are not as unambiguous as in written language, and therefore, segmentation of the speech stream into words is more difficult. However, results of a previous study indicate that lexical access is also activated in pseudo-word recognition (Specht et al. 2003). This supports the assumption that in our study phonologically similar words were activated while listening to the pseudo-homophonic words used in this study. It is assumed that an activation of phonologically similar words took place in both subject groups. This would explain why also native speakers were not able to 100% identify the sentences even at 0 ms inversion.

The non-native speakers could only tolerate temporal distortions up to an interval of 32 ms (semantically coherent sentences) and 26 ms (semantically incoherent sentences). The respective values for native speakers lie at 50 ms and 44 ms. The difference in understanding abilities between native and non-native speakers can be explained by differences in the language processing of the first and the acquired, second language. In word recognition, a process of multiple, simultaneous activation of word candidates is assumed (Marslen-Wilson and Welsh 1978). Regarding this multiple activation, studies have shown that recognition of spoken words by non-native speakers is aggravated by multiple vocabulary activation (Doctor and Klein 1992, Dijkstra et al. 1999). This means that when listening to a non-native language, phonologically similar words in the native language are additionally activated during word recognition. Another factor that makes word recognition in the second language more difficult is that the representations of phoneme categories are not as unambiguous as in the native language, and, therefore, phoneme perception is less accurate (Strange 1995). These examples clearly show that understanding of a second language is complicated by several factors. In the present investigation, the aggravating factor (inversion of the speech signal) that was added revealed differences between native and non-native speakers which otherwise (with the non-distorted language material of 0 ms inversion intervals) would not have been detected. An explanation for these differences is that different levels of proficiency in lexical and semantic processing influence phoneme detection.

CONCLUSIONS

Our study investigated the understanding of three different types of temporally reversed sentences in native and non-native speakers. In semantically coherent sentences, lexical and contextual cues are available to understand the distorted speech signal. In the category of semantically incoherent sentences, only lexical cues can be used to decode the auditory input. In contrast, no lexical or contextual cues are embedded in the pseudo-homophonic sentence type. A purely bottom-up analysis of the phonetic-acoustic structure is therefore needed to understand the pseudo-words. In this context of an increasing amount of linguistic cues, which are present in the three types of sentences (pseudo-homophonic sentences: no cues, semantically incoherent sentences: lexical cues,

semantically coherent sentences: lexical and contextual cues), the differences in speech comprehension between the two groups can be understood. Native language speakers were better at detecting the semantically coherent and incoherent sentences because of their higher proficiency in lexical and semantic knowledge. In the pseudo-homophonic sentences these cues were not available, thus leading to a similar degree of low comprehension in both groups. We believe that this test with temporally-distorted speech could offer a new approach to assess language skills that indirectly taps into lexical and semantic competence of non-native speakers. Complementing the employment of complex language tasks, our test on distorted speech can be used to assess lexical and semantic processing abilities of various subject groups such as foreign language learners or – after some further modifications – patients with language disorders.

REFERENCES

Ackermann H, Lutzenberger W, Hertrich I (1999) Hemispheric lateralization of the neural encoding of temporal speech features: a whole-head magnetencephalography study. Cog Brain Res 7: 511–518.

Berwanger D, Wittmann M, von Steinbüchel N, von Suchodoletz W (2004) Measurement of temporal-order judgment in children. Acta Neurobiol Exp (Wars) 64: 387–394.

Borsky S, Tuller B, Shapiro LP (1998) "How to milk a coat". The effects of semantic and acoustic information on phoneme categorization. J Acoustic Soc Am 103: 2670–2676.

Burton MW, Baum SR, Blumstein SE (1989) Lexical effects on the phonetic categorization of speech: The role of acoustic structure. J Exp Psychol Hum Percept Perform 15: 567–575.

Dijkstra A, Grainger J, van Heuven W (1999) Recognition of cognates and interlingual homographs: The neglected role of phonology. J Mem Lang 41: 496–518.

Doctor E, Klein D (1992) Phonological processing in bilingual word recognition. In: Cognitive Processing in Bilinguals (Harris R, Ed.). Elsevier, Amsterdam, p. 237–252.

Farmer ME, Klein RM (1995) The evidence for a temporal processing deficit linked to dyslexia: A review. Psychon Bull Rev 2: 460–493.

Fink M, Churan J, Wittmann M (2005) Assessment of auditory temporal-order thresholds – a comparison of different measurement procedures and the influences of age and gender. Restor Neurol Neurosci 23: 281–296.

Fink M, Ulbrich P, Churan J, Wittmann M (2006a) Stimulus-dependent processing of temporal order. Behav Processes 71: 344–352.

Fink M, Churan J, Wittmann M (2006b) Temporal processing and context dependency of phoneme discrimination in patients with aphasia. Brain Lang 98: 1–11.

Fraisse P (1984) Perception and estimation of time. Annu Rev Psychol 35: 1–36.

Ganong WF (1980) Phonetic categorization in auditory word perception. J Exp Psychol Hum Percept Perform 6: 110–125.

Hirsh I (1959) Auditory perception of temporal order. J Acoust Soc Am 31: 759–767.

Hirsh I, Sherrick C (1961) Perceived order in different sense modalities. J Exp Psychol 62: 423–432.

Joliot M, Ribary U, Llinas R (1994) Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. Proc Natl Acad Sci U S A 88: 4966–4970.

Kanabus M, Szelag E, Rojek E, Pöppel E (2002) Temporal order judgment for auditory and visual stimuli. Acta Neurobiol Exp (Wars) 62: 263–270.

Kiss M (2002) The Relationship Between Temporal Processing and Language in Patients with Fluent Aphasia (in German). Verlag Dr. Hut, Munich.

Lotze M, Wittmann M, von Steinbüchel N, Pöppel E, Roenneberg T (1999) Daily rhythm of temporal resolution in the auditory system. Cortex 35: 89–100.

Marslen-Wilson WD, Welsh A (1978) Processing interactions and lexical access during word-recognition in continuous speech. Cognit Psychol 10: 29–63.

Mills L, Rollman G (1980) Hemisphere asymmetry for auditory perception of temporal order. Neuropsychologia 18: 41–47.

Morton J (1969) Interaction of information in word recognition. Psychol Rev 76: 165–178.

Pastore R, Farrington S (1996) Measuring the difference limen for identification of order of onset for complex auditory stimuli. Percept Psychophys 58: 510–526.

Pöppel E (1970) Excitability cycles in central intermittency. Psychol Forsch 34: 1–9.

Pöppel E, Schill K, von Steinbüchel N (1990) Sensory integration within temporally neutral system states: a hypothesis. Naturwissenschaften 77: 89–91.

Pöppel E (1997) A hierarchical model of temporal perception. Trends Cogn Sci 1: 56–61.

Rosen S (1992) Temporal information is speech: Acoustic, auditory, and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336: 367–373.

Saberi K, Perrot D (1999) Cognitive restoration of reversed speech. Nature 398: 760.

Simos P, Diehl R, Breier J, Molis M, Zouridakis G, Papanicolaou A (1998) MEG correlates of categorical perception of a voice onset time continuum in humans. Brain Res Cogn Brain Res 7: 215–219.

Specht K, Holtel C, Zahn R, Herzog H, Krause B, Mottaghy M, Radermacher I, Schmidt D, Tellmann L, Weis S, Willmes K, Huber W (2003) Lexical decision of non-words and pseudowords in humans: a positron emission tomography study. Neurosci Lett 345: 177–181.

Steffen A, Werani A (1994) An experiment on temporal processing in language perception (In German). In: Sprechwissenschaft und Psycholinguistik 6 (Kegel G, Arnhold T, Dahlmeier K, Schmid G, Tischer B, Eds.). Westdeutscher Verlag, Opladen, p. 189–205.

Steffen A, Werani A, Kegel G (1997) Temporal sequencing in speech perception. Abstract. Exp Brain Res 117 (Suppl.): 63.

Steinbüchel von N, Wittmann M, Szelag E (1999) Temporal constraints of perceiving, generating, and integrating information: Clinical indications. Restor Neurol Neurosci 14: 167–182.

Steinbüchel von N., Wolff C, Heel S, Kiss M, Wittmann M, Schwender D (2000) Electrophysiological indicators of auditory order identification. Soc Neurosci Abstracts 26: 1248.

Steinschneider M, Schroeder C, Arezzo J, Vaughan H (1995) Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey. Brain Lang 48: 326–340.

Stevens K (1998) Acoustic Phonetics. MIT Press, Cambridge, MA.

Strange W (Ed.) (1995) Speech Perception and Linguistic Experience: Issues in Cross-Language Research. New York Press, Baltimore, MD.

Swisher L, Hirsh I (1972) Brain damage and the ordering of two temporally successive stimuli. Neuropsychologia 10: 137–152.

Szelag E, Kanabus M, Kolodziejczyk I, Kowalska J, Szuchnik J (2004) Individual differences in temporal information processing in humans. Acta Neurobiol Exp (Wars) 64: 349–366.

Tallal P, Piercy M (1973) Defects of non-verbal auditory perception in children with developmental aphasia. Nature 241: 468–469.

Tallal P, Merzenich M, Miller S, Jenkins W (1998) Language learning impairment. Integration research and remediation. Scand J Psych 39: 197–199.

Wackermann J, Späti J (2006) Asymmetry of the discrimination function for temporal durations in human subjects. Acta Neurobiol Exp (Wars) 66: 245–254.

Warren RM (1970) Perceptual restoration of missing speech sounds. Science 267: 392–393.

Wittmann M (1999) Time perception and temporal processing levels of the brain. Chronobiol Int 16: 17–32.

Wittmann M, Fink M (2004) Time and language – critical remarks on diagnosis and training methods of temporal-order judgement. Acta Neurobiol Exp (Wars) 64: 341–348.

Wittmann M, Burtscher A, Fries W, Steinbüchel von N (2004) Effects of brain-lesion size and location on temporal-order judgment. NeuroReport 15: 2401–2405.