Clinical Linguistics & Phonetics, December 2010; 24(12): 980–996

informa
healthcare

The influence of phonetic dimensions on aphasic speech perception

DÖRTE HESSLER, ROEL JONKERS, & ROELIEN BASTIAANSE

Neurolinguistics, Center for Language and Cognition Groningen (CLCG), University of Groningen, The Netherlands

(Received 8 March 2010; Accepted 6 July 2010)

Abstract
Individuals with aphasia have more problems detecting small differences between speech sounds than larger ones. This paper reports how phonemic processing is impaired and how this is influenced by speechreading. A non-word discrimination task was carried out with 'audiovisual', 'auditory only' and 'visual only' stimulus display. Subjects had to decide whether two presented stimuli were the same. Six aphasic subjects with speech sound processing difficulties and 14 non-brain-damaged control subjects participated in this study. It was found that the aphasic subjects have difficulties in discriminating pairs of non-words, which are more profound for small differences. Differences in 'voicing' were least often detected and therefore seem most difficult to perceive. This implies different processing of the phonetic dimensions in speech sound perception. Performance improved when speechreading was possible. As this improvement is not based on differences in place of articulation only, theories of audiovisual processing need to be revised.

Keywords: *aphasia, speech reading, speech sound processing, phonetic dimensions*

Introduction

In this paper we will report the influence that different phonetic dimensions have on speech sound processing in Dutch aphasic subjects with a disorder in speech sound processing and how speechreading[1] can aid processing. First we will, however, give a background on the phonetic dimensions, the influences of speechreading on comprehension, and how that can be accounted for in a speech processing model. Deficits in speech sound discrimination will also be discussed.

Phonetic dimensions

Phonemes are considered to consist of phonetic features (Chomsky and Halle, 1968). These features can be categorized into the three (distinctive) phonetic dimensions 'place of articulation', 'manner of articulation' and 'voicing'. A combination of these dimensions uniquely identifies each phoneme. Changes in one phonetic dimension will lead to a different phoneme: changing, for example, the 'place of articulation' from bilabial to alveolar would

Correspondence: Dörte Hessler, Neurolinguistics, Faculty of Arts, University of Groningen, PO Box 716, 9700 AS Groningen, The Netherlands. Tel: +31-50-363 5970. Fax: +31-50-363 6855. E-mail: d.a.hessler@rug.nl

ISSN 0269-9206 print/ISSN 1464-5076 online © 2010 Informa UK Ltd.
DOI: 10.3109/02699206.2010.507297

transform a /p/ into a /t/. However, also broader contrasts between two phonemes (changes in two or three of the phonetic dimensions) are possible: /p/ and /z/, for example, are distinguished by 'voicing', 'place of articulation' and 'manner of articulation'. Therefore, it is not sufficient to classify two words or syllables as different in one phoneme, as this distinction can be formed by differences in one, two or all three distinctive dimensions. Rather the number and type of the phonetic dimensions differing should be mentioned as well.

The phonetic characteristics of a particular phonetic dimension are manifested differently across languages: The distinction 'voiced' vs. 'voiceless' is made by differences in voice onset time (VOT). According to Lisker and Abramson (1964), the voicing distinction in English is achieved by contrasting onset of voicing at the release of the lips (for /b/) with an onset of voicing up to 100 ms later (/p/). Therefore, the distinction in English is rather one between voiceless and voiceless-aspirated. For other languages such as Dutch and Hungarian, however, Lisker and Abramson (1964) found that the voiceless /p/ was produced by having the onset of voicing aligned with the release of the lips, while the voiced counterpart /b/ was produced with a voice onset at least 50 ms prior to the lip-release.

A distinction in 'voicing' therefore refers to different phonetic distinctions in different languages, which makes a comparison of, for example, Dutch and English data difficult. In Dutch and Hungarian, however, the distinction between 'voiced' and 'voiceless' is accomplished by similar VOT patterns.

Speechreading

The extraction of phonetic information out of the speech stream is an early component of language comprehension. Language comprehension is, however, a multimodal process. Not only auditory but also visual information (seen speech) is employed in perception (Rosenblum, 2008). It has been demonstrated that seeing the speaker facilitates comprehension in a noisy environment (Sumby & Pollack, 1954) or with cognitively demanding contents under good listening conditions (Reisberg, McLean, & Goldfield, 1987).

More evidence for the fact that speechreading is automatically integrated into speech perception is provided by experiments carried out by McGurk and MacDonald (1976). In their study, subjects watched dubbed videos in which auditory and visual information did not match, and they were asked to report what they perceived. Instead of answering with the auditory (/ba/) or the visual (/ga/) component of the video, they usually reported a fusion of both (/da/). This even occurred when the subjects were aware of the dubbing. This so-called 'McGurk' effect actually shows the influence of speechreading. It is also a demonstration of the integration of both modalities by producing a percept which is a fusion of seen and heard speech. This proves that the information a listener gains from the lip-movements of the speaker cannot be ignored and is automatically taken into account in generating a percept. Therefore, speechreading should not be understood as a substitute mechanism that only mediates when needed, but as one that supports auditory comprehension.

Campbell (1988; 1990) suggested a model explaining speech perception with multimodal (auditory and visual) input. Her model is based on the TRACE model of language processing (McClelland & Elman, 1986). In this interactive activation model of speech processing several levels of processing are assumed: a phonetic, a phonological and an abstract phonemic level. The phonetic level consists of acoustic as well as lip-read features. These are connected to the phonological level, where phonemes are represented. The third level, consisting of abstract phonemic units, is necessary to explain why some people actually 'hear' phonemes which are not articulated, but only written, such as the /b/ in 'comb'. The three levels are fully interconnected. Units within one level are connected via lateral inhibition, that means that units can inhibit each other. Across levels the connections are excitatory. This excitation applies both bottom-up and top-down. The input is temporally organized to assure the correct order of phonemes in a word. This model incorporates a delay window, called TRACE, to account for context and mispronunciation effects (where incorrectly pronounced words are still recognized). In the TRACE all patterns of activation are stored that correspond to a stimulus that has not yet been identified.

To integrate the information from speechreading, Campbell (1988; 1990) introduced two new features at the level of phonetic units into the model: 'mouth opening' and 'lip-shape'. Together with the acoustically perceived features these visual features form the phonetic units which interact with each other and with the phonological units. A schematic overview of the 1990 version of the model is shown in Figure 1. The features 'mouth opening' and 'lip-shape' mainly convey information necessary to decode the 'place of articulation'. This model therefore predicts 'place of articulation' to be influenced by speechreading, but it is not clear whether there is also influence on the dimensions 'manner of articulation' and 'voicing'.

Impairments of processing

Brain damage can lead to an impairment in processing phonemes. This disorder was first described by Kussmaul (1877), who called it 'pure word deafness', because the patient he described did not suffer from other aphasic symptoms. Terminology was not consistent and now a common term is word-sound deafness (Franklin, 1989). Word-sound deafness can be diagnosed with auditory discrimination tasks, in which subjects have to report whether two auditory (phonologically related) stimuli (words or non-words) are the same or different (e.g.

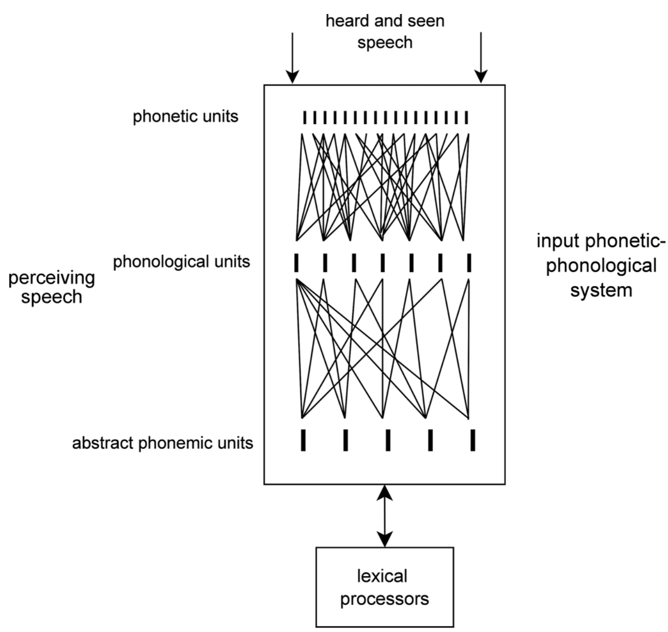

Figure 1. Schematic overview of a model of audiovisual processing, based on Campbell (1990).

'house' and 'mouse'). In word-sound deafness, the problems are restricted to linguistic material, while there are no problems in discriminating or identifying non-linguistic auditory stimuli.

Aphasic subjects with a disorder in analysing speech sounds will have problems in discrimination of items differing in fewer phonetic dimensions. The more dimensions are different, the easier the discrimination task becomes (Blumstein, Baker, & Goodglass, 1977). Factors that have a beneficial influence are the use of context or slowed speech and, most importantly for the present study, the possibility to see the speaker, thus the possibility to gain speechreading information (Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986; Shindo, Kaga, & Tanaka, 1991). The use of speechreading information has also been successfully utilized in treatment studies (Gieliewski, 1989; Grayson, Hilton, & Franklin, 1997; Hessler & Stadie, 2008; Morris, Franklin, Ellis, Turner, & Bailey, 1996).

Hessler and Stadie (2008) evaluated the effects of a systematic treatment of the auditory analysis of speech in a patient with aphasia. The treatment was based on the beneficial influence of speechreading. During treatment six different tasks were carried out: auditory discrimination of syllables, auditory discrimination of phonemes, word-picture matching, word-picture verification, heard word-written word matching and heard word-written word verification. In all tasks the distractors were phonologically related to the targets. Treatment started with items with broad distinctions (three phonetic dimensions) and speechreading possible. After mastery of this condition more difficult conditions were presented (less dimensions different, no speechreading possible). The efficiency of treatment was measured by comparing pre- and post-treatment performance in the treatment tasks on a set used during treatment and a matched, non-trained set of stimuli. Apart from showing general improvement in both the trained and the untrained set, the authors also analysed the performance on individual phonetic dimensions. The aphasic patient improved in discrimination of 'place of articulation' contrasts as well as 'manner of articulation' contrasts (also for untrained stimuli). These results cannot be explained by the model by Campbell (1988; 1990) described above. Influence on other dimensions than 'place of articulation' is not predicted by this model. The results of Hessler and Stadie (2008) indicate that speechreading can be beneficial for perceiving the other dimensions as well. It is therefore not clear which phonetic dimensions ('place of articulation', 'manner of articulation' and/or 'voicing') make use of the additional information from seen speech.

Processing of which phonetic dimensions is most impaired in aphasic comprehension disorders has been investigated previously: Blumstein et al. (1977) compared the processing of the dimensions 'place of articulation' and 'voicing' in English-speaking aphasic subjects and found that they have most problems with 'place of articulation'. They did not include the dimension 'manner of articulation'. Saffran, Marin, and Yeni-Komshian (1976) and Caplan and Aydelott-Utman (1994), however, found (also for English aphasic subjects) that 'voicing' actually is more difficult than 'place of articulation'. Similar results have been found for Hungarian by Csépe, Osman-Sági, Molnár, and Gósy (2001) for two aphasic subjects with unilateral left-hemisphere lesions (opposed to the bilateral cases also investigated). Klitsch (2008) used two sub-tests of the Dutch version of the PALPA (Bastiaanse, Bosje, & Visch-Brink, 1995) to investigate whether aphasic subjects showed differences in detecting distinctions in the dimensions 'place of articulation', 'manner of articulation' and 'voicing'. In the two discrimination tasks carried out, words as well as non-words were investigated. Generally the performance of the aphasic subjects was better when word pairs had to be distinguished. The detection of differences in 'place of articulation' was worse than 'manner of articulation' but only for non-words. The comparison of either 'place of articulation' or 'manner of articulation' with 'voicing' was more difficult, as 'voicing' distinctions were realised in initial positions, the other contrasts however in final (less salient) position or metathesis. For comparison with 'voicing' only the performance in metathesis distinctions in 'place of articulation' and 'manner of articulation' were taken into account. Compared like this, performance on 'voicing' distinctions was worse than on 'manner of articulation' distinctions, while there was no difference with 'place of articulation'. Klitsch (2008), therefore, came to the cautious conclusion that 'place of articulation' was affected most, but also noted that the dimension 'voicing' could not be compared reliably to the other dimensions, because they were not occurring in the same position within the stimuli.

In the current study all three dimensions will be compared again, but manipulated in the same position (initially). In order to investigate the influence speechreading has on the discrimination performance, the task will be presented in three conditions: with only auditory input, with audio-visual input and with only visual input (a video of lip-movements, serving as a control condition).

Based on the former studies we expect that aphasic subjects with a disorder in speech sound processing benefit from information derived from speechreading in discriminating between similar phonemes. Therefore, their overall performance in the 'audiovisual' condition will be better than in the 'auditory only' condition. This beneficial influence of speechreading will be manifested in the phonetic dimension 'place of articulation'. Based on Campbell's (1988; 1990) model no beneficial influence is predicted on the dimensions 'manner of articulation' and 'voicing'.

It is also expected that the degree of difference has an influence on the performance: The more dimensions differ between two items, the easier discrimination becomes for the individuals with aphasia. They will have least difficulties with distinguishing stimuli when all three phonetic dimensions differ, while differences in only one phonetic dimension are expected to cause most problems.

Additionally it will be investigated whether all three phonetic dimensions are equally difficult for individuals with aphasia or whether one of them is particularly difficult and, if so, which one. For the 'audiovisual' condition it is predicted that the dimension 'voicing' will be most difficult, as within this dimension it is not possible to make use of visual information. This prediction, however, does not hold for the 'auditor only' condition, as there is no visual information available.

Methods

Subjects

Six subjects with aphasia (three female) and 14 non-brain-damaged control subjects (seven female) participated in this study. All subjects were native speakers of Dutch, right-handed, and reported normal hearing. The hearing was also judged as within functional limits by their speech-therapists. Vision was normal or corrected to normal. The aphasic subjects were between 47–64 years old (mean age: 52.33). The subjects in the control group were matched with the aphasic subjects for age (mean age 56.29; range 49–67), gender and region of origin. They had never experienced neurological problems and had no (history of) language disorders.

All aphasic subjects were at least 3 months post-onset. None of them had demonstrated any language disorders prior to the CVA. They did not suffer from any neuropsychological problems influencing the testing (such as severe attention disorders). The subjects were selected on the basis of their results in the PALPA non-word discrimination task (Bastiaanse et al., 1995). In this task subjects hear two non-words which are either the same or differ in one phonetic dimension from each other. They have to judge whether both heard stimuli are the same. A failure to do so has been attributed to an impairment of the auditory analysis of speech. This task was administered using a recording of the stimuli in order to maximize the comparability between subjects. As the normative data of the PALPA (Bastiaanse et al., 1995) are collected with direct speech, Klitsch (2008) collected normative data for the recorded version.

We compared the results of our subjects to her normative data. Performance more than 2 SD below the mean of Klitsch's group were considered as impaired. Only aphasic subjects with impaired performance on the discrimination task were included in this study. Thus, all aphasic subjects in this study had a deficit in the auditory analysis of speech. Furthermore, all but one participant had been diagnosed with a standardized battery, the Akense Afasietest (AAT) (Graetz, De Bleser, & Willmes, 1992). The performance in two sub-parts, the Token Test and the comprehension part, gives an indication of the comprehension abilities of the subjects. In the Token Test, subjects have to follow commands such as 'Touch the green rectangle' or 'Put the red square under the red circle'. These commands, of which 50 are presented, vary in length and complexity. The results are reported as error scores. A score of '0' would therefore mean 'no errors', while '50' represents the fact that no command at all could be executed correctly. The comprehension part of the AAT consists of word and sentence comprehension tasks. A word or sentence is presented and the subject is asked to choose between four pictures, one depicting the target and the other three distractors (one or two of which are related to the target). The maximum score that can be reached is 120. An overview of the personal data of the aphasic subjects and their results on these two AAT tasks (Graetz et al., 1992) and the PALPA non-word discrimination task (Bastiaanse et al., 1995) are given in Table I .

Materials

The materials consisted of one-syllable non-words with CVC(C) structure. They were spoken by a male native speaker of Dutch, who was recorded in a quiet room with daylight. Additionally a light diffuser was used to avoid shading on the recorded material in order to ensure optimal visual information. The recorded frame included the lower part of the speaker's face (from the bottom of the nose), the neck and the upper chest. For the recording, a video camera and separate cardioid microphone were used. The video was then digitized into avi-files at a sampling rate of 48 kHz with 32-bit-stereo quantization. All stimuli were edited with Adobe Premiere to form video files with a duration of 3 seconds each. Recording was done with 25 frames per second (thus 40 ms per frame). Therefore, each file consisted of 75 frames. The video showed the speaker in rest (with a closed mouth) for 12 frames (480 ms) at the end of each video. The resting phase in the beginning was varied slightly to ensure equal length of all videos. To provide equal length of rest, the last or first frame of the video was artificially prolonged, where necessary. As the experiment was carried out in three conditions ('audiovisual', 'auditory only' and 'visual only' presentation) the audiovisual video files were then further edited to create the stimuli for the other conditions. For the 'auditory only' condition the picture was deleted, leaving the sound and a blank screen. In the 'visual only' condition, the audio trace was removed resulting in a video without sound. Finally the video

Table I. Overview of aphasic subjects.

Initials	Age	Gender	Type of aphasia	Months post-onset	AAT Token Test	AAT Comprehension	PALPA Non-word discrimination
WB	57	male	Wernicke	148	–	–	56/72
BB	64	male	Global	5	50	67	53/72
EK	48	male	Anomia	16	11	88	58/72
TB	47	female	Global	8	33	53	68/72
JH	51	female	Mixed	44	36	89	66/72
MB	47	female	Global	4	50	68	64/72

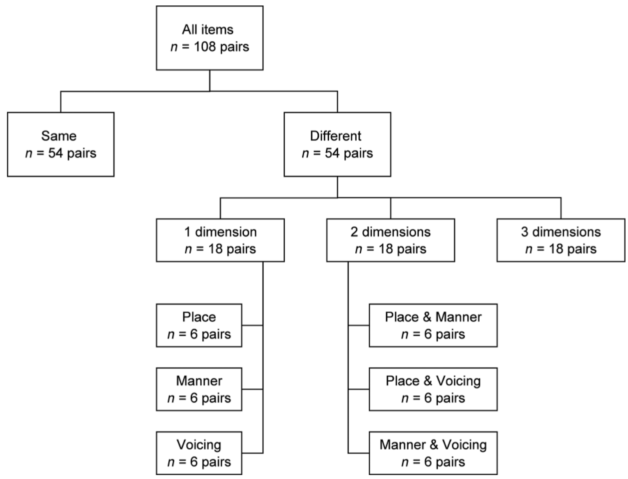

Figure 2. Structure of materials used.

files were converted into Windows Media files (.wmv), reducing file size in order to guarantee smooth running of the experiment without long delays for loading the files.

As said above, the material consisted of pairs of non-words. These were presented either ‘auditorily only’, ‘audiovisually’ or ‘visually only’, depending on the sub-condition of the experiment. Distribution of the material can be seen in Figure 2. A complete overview of the used stimuli can be found in Appendix A.

Procedure

Each participant was tested in three sessions: In the first session the PALPA non-word discrimination task (Bastiaanse et al., 1995) was carried out. The experimental task was administered in two further sessions. The task used in this study was a discrimination task, asking the subjects to state whether two heard and/or seen syllables are the same. It was carried out in three different conditions: (1) ‘auditory only’ (2) ‘audiovisual’ and (3)‘visual only’. The last condition was introduced as a control condition, indicating that the presumed better performance in the ‘audiovisual condition’ was not solely due to the visual information. For all three conditions the same stimuli were used. The items of each condition were split into two blocks, so that only half of the items were presented in one session. The order of presentation of the blocks was balanced between subjects.

The materials were presented to the participant on a laptop equipped with headphones and a response box using E-Prime 2.0 (Psychology Software Tools). For each condition there were five practise trials before the experiment started. On those items feedback was provided. The practise trials were repeated if the participant requested it or if it seemed necessary to explain the procedure again. The experimental task was only started once the subjects responded correctly to at least 80% of the trials in the ‘auditory only’ and ‘audiovisual’ conditions without help. Each condition was presented separately, not mixing the conditions. Items were randomized to prevent learning effects across conditions. The order in which the conditions were presented was varied between subjects, so that a possible learning effect would not favour a certain condition.

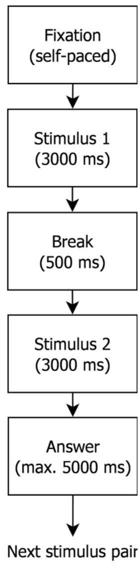

Figure 3. Flowchart of procedure.

A schematic outline of the procedure per trial is shown in Figure 3. Item presentation occurred self-paced. Prior to each pair of non-words an asterisk was shown on the screen. The video only started when requested by the participant by pushing a button. After both non-words were presented, the participant had 5 seconds to respond with 'yes' or 'no'. This response was also given by pushing a button (a green button for 'yes' and a red one for 'no'). If no response had been recorded after 5 seconds, the non-word pair was presented again with another 5 seconds to decide. If there was no answer after the second presentation, the next stimulus pair was presented. Each condition of the experiment consisted of 108 stimulus pairs; however, as mentioned above, only half of those were presented within one session.

Results

Overall performance

The non-brain-damaged control subjects scored at ceiling for the 'auditory only' and 'audio-visual' conditions. In the 'visual only' condition they performed worse, failing mainly in contrasts involving only 'voicing' or 'manner of articulation' or the combination of both. The aphasic subjects scored significantly lower than control subjects on all conditions (2-tailed Mann–Whitney U tests: 'auditory only': 99–87% correct, $Z = -3.521$, $p < .001$; 'audiovisual': 99–90% correct, $Z = -3.545$, $p < .001$; 'visual only': 83–63% correct, $Z = -3.387$, $p < .001$). Because the non-brain-damaged control subjects performed at ceiling the following analyses were only conducted within the group of aphasic subjects.

Influence of speechreading

Based on previous studies it was expected that aphasic subjects with speech sound processing disorders benefit from speechreading. It was investigated whether the performance of the aphasic subjects also improved with speechreading. The results in the three conditions 'auditory only', 'audiovisual' and 'visual only' (control condition) differed significantly (Friedman Anova $\chi^2 = 12$, $df = 2$, $p = .002$). Post-hoc Wilcoxon tests revealed that the 'audiovisual' condition was significantly easier than both the 'auditory only' ($Z = -2.207$, $p = .027$) and the 'visual only' condition ($Z = -2.201$, $p = .028$). There was also a significant advantage for the 'auditory only' over the 'visual only' condition ($Z = -2.207$, $p = .027$). This also holds on an individual basis: The performance in the 'audiovisual' condition was better than in the 'auditory only' condition for five out of the six aphasic subjects.

Analysis by phonetic dimension. According to Campbell's (1988; 1990) model, it was expected that performance concerning the dimension 'place of articulation' would improve with the addition of speechreading cues. No improvement was predicted for the other two dimensions. In order to investigate the influence of speechreading, a comparison of the 'audiovisual' and the 'auditory only' conditions separately for each phonetic dimension was carried out. It revealed no significant differences for the dimensions 'place of articulation' (2-tailed Wilcoxon test: $Z = -.816$, $p = .414$) and 'voicing' (2-tailed Wilcoxon test: $Z = -.674$, $p = .500$). For the dimension 'manner of articulation' a trend for better performance on 'audiovisual' stimuli could be found (2-tailed Wilcoxon test: $Z = -1.826$, $p = .068$). The individual results in Appendix B.2 show that this trend was not caused by single subjects, but was found for four of the six aphasic subjects, while the other two showed no difference between both conditions.

Number of distinguishing dimensions

Based on the results of Blumstein et al. (1977), it was predicted that the number of phonetic dimensions differing would influence the performance of the aphasic subjects, such that the fewer phonetic dimensions differ the worse the performance becomes.

Analyses revealed that the number of dimensions differing within the pair played a role for aphasic subjects in the 'auditory only' (Friedman Anova: $\chi^2 = 8.667$, $df = 2$, $p = .013$) and 'audiovisual' (Friedman Anova: $\chi^2 = 11.143$, $df = 2$, $p = .004$) conditions. In the auditory condition it was found that differences in one distinctive dimension were significantly less likely to be detected than differences in two (2-tailed Wilcoxon test: $Z = -2.023$, $p = .043$) or three (2-tailed Wilcoxon test: $Z = -2.207$, $p = .027$) dimensions. There was, however, no significant difference between distinctions in two and three dimensions (2-tailed Wilcoxon test: $Z = .000$, $p = 1.000$). Similar results have been found for the audiovisual condition: distinctions in two and three dimensions were not significantly different from each other (2-tailed Wilcoxon test: $Z = -1.604$, $p = .109$), while both were easier to perceive than distinctions in one dimension (2-tailed Wilcoxon test: $Z = -2.201$, $p = .028$ for both comparisons). The individual data of the aphasic subjects (Appendix B.1) show that these findings were not caused by the performance of single subjects, but hold for all aphasic subjects in the 'auditory only' condition and all but one in the 'audiovisual' condition.

Type of distinguishing dimension

Previous studies have found contradictory results concerning the question which phonetic dimension is most impaired in aphasic perception. Therefore, no prediction was made for the 'auditory only' condition. For the 'audiovisual condition' it was expected that differences in the dimension 'place of articulation' would be the easiest to perceive, as, following Campbell's (1988; 1990) model, beneficial influence of speechreading is assumed for this dimension, but not for the other two dimensions.

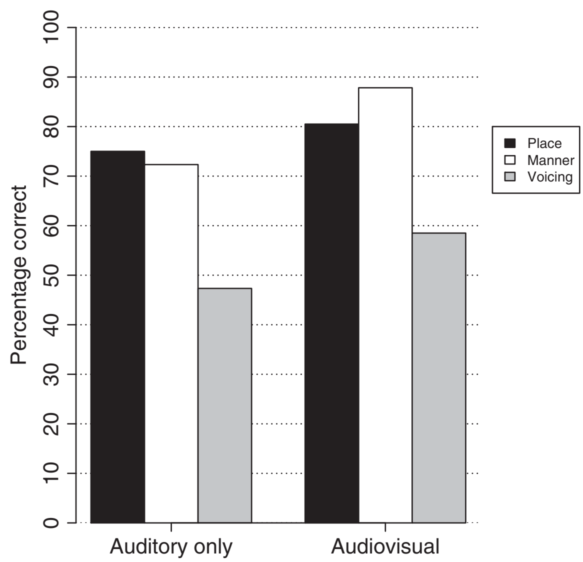

Figure 4. Percentage of correct aphasic responses to different dimensions in auditory only and audiovisual conditions.

An analysis of the influence of type of dimension ('place of articulation' vs. 'manner of articulation' vs. 'voicing') was carried out, showing significant results for the 'auditory only' condition (Friedman Anova: $\chi^2 = 6.7$, $df = 2$, $p = .035$) and marginally significant results for the 'audiovisual' condition (Friedman Anova: $\chi^2 = 4.727$, $df = 2$, $p = .094$). For both conditions it appears that 'voicing' was the most difficult to distinguish, followed by 'place of articulation' and 'manner of articulation' (see Figure 4). In both the 'auditory only' and the 'audiovisual' condition, five out of six aphasic subjects showed the same pattern as the group, with 'voicing' being most difficult (see Appendix B.2). The group result, therefore, reflects a vast majority of the subjects' performances, rather than being caused by extremes in the data distribution.

Answer bias

Results of 'yes–no-paradigms' can be influenced by an answer bias of the subjects. This can be corrected by using methods from signal-detection research. Within this paradigm the hit-rate and the false-alarm-rate are used to calculate a measure of discriminability, d-prime. The calculation of d-prime is a parametric procedure. As the current data do not fulfil the demands for parametric testing, d-prime could not be calculated. Instead a non-parametric variant, a-prime (A'), was calculated to correct for a response bias. A'-scores vary between '0' (no discriminability) and '1' (perfect discriminability), with '0.5' being chance-level. In the current study we applied the algorithms from Snodgrass, Levy-Berger, and Haydon (1985) to calculate A'. All statistical analyses have been repeated using the bias-corrected A'-scores. Also using these scores (rather than the non-corrected ones) it becomes evident that the aphasic subjects score significantly worse than the non-brain-damaged controls in all three conditions ('audiovisual', 'auditory only' and 'visual only'). Regarding the analyses within the aphasic group the results resembled those of the non-corrected scores, with two exceptions: The overall difference between the 'audiovisual' and the 'auditory only' condition does not yield significance, but forms a trend when based on A'-scores. The same is true for the difference between two and one dimension distinctions found in the 'auditory only' condition: Using the corrected A'-scores a trend, rather than a significant difference, can be found, indicating that two dimensions cause more difficulties than one. The individual A'-scores are mentioned in Appendix A and the results of the statistics using A'-scores are provided in Appendix C.

Discussion

The aim of the current study was to investigate how perception of phonetic dimensions is impaired in speech processing by individuals with aphasia, and how that processing is influenced by speechreading. A discrimination task was carried out in three conditions: 'auditory only', 'audiovisual' and 'visual only' stimulus presentation. A group of 14 non-brain-damaged control subjects and six aphasic subjects participated in this study. The aphasic subjects were diagnosed with different syndromes, but shared a deficit in processing speech sounds. The small number of aphasic subjects does not allow for general conclusions, rather all conclusions drawn refer only to the group tested.

It was found (repeating numerous previous studies) that discriminating pairs of non-words is more difficult for individuals with an aphasic disorder in speech sound processing than for non-brain-damaged control subjects. When analyses use the bias-corrected A'-scores this observation also holds.

Generally the aphasic subjects showed a very homogeneous pattern. For all analyses reported in this paper, a broad majority of the aphasic subjects showed performance in the same direction as the group. The group analyses were, therefore, based on a consistent pattern within the group rather than on extreme performances of single subjects. The possibility that hearing problems influenced the results is ruled out by the fact that the aphasic subjects had only slight problems with differences in three dimensions in the 'auditory only' condition. If the underlying problems were in hearing, this condition should have been affected as well.

Overall there was a trend to better performance of individuals with aphasia in the 'audiovisual' condition than in the 'auditory only' condition, indicating that the additional visual information gained from speechreading facilitates their discrimination abilities. The performance in the control condition with 'visual only' stimulus presentation was worse than in both other conditions, indicating that the superiority of the 'audiovisual' condition is not due to pure visual information, but rather the combination of auditory and visual input. For the non-brain-damaged control group no difference between the 'auditory only' and the 'audiovisual' conditions were found, as they performed at ceiling in both. No further analyses were carried out for the non-brain-damaged control group.

For the aphasic subjects it was further tested whether the general advantage of the 'audiovisual' over the 'auditory only' condition was due to improvement on one of the phonetic dimension in particular. Therefore, the difference between the 'audiovisual' and the 'auditory only' conditions was analysed individually for each of the three phonetic dimensions 'place of articulation' 'manner of articulation', and 'voicing'. According to the model of Campbell (1988; 1990), improvement, particularly in the dimension 'place of articulation' was expected when additional speechreading is possible. However, we did not find significant differences between 'audiovisual' and 'auditory only' presentation for any of the dimensions individually. It is, hence, not possible to say whether there was more improvement for one of the dimensions than for another. The general improvement is, therefore, not due to one dimension in particular, but rather to a summation of improvement on all of them.

These findings are not in line with Campbell's (1988; 1990) model, as only improvement for distinctions in 'place of articulation' was predicted. This prediction was previously questioned in the treatment study by Hessler and Stadie (2008). They found improvement for 'manner of articulation' after a treatment based on utilizing speechreading. Therefore, it seems that Campbell's (1988; 1990) model needs to be extended to account for influences from visual features other than 'mouth opening' and 'lip-shape'.

The difficulties individuals with aphasia experience when discriminating stimuli were more profound for smaller distinctions. As previously reported by Blumstein and Cooper (1972) and Blumstein et al. (1977) for English, we also found for Dutch that differences are less likely to be detected if the items within the pair differ in one phonetic dimension (rather than in two or three). This holds for the 'auditory only' as well as for the 'audiovisual' condition. Even though performance is generally better in the 'audiovisual' condition it is still impaired, especially regarding the small differences. Speechreading adds information that enhances speech sound processing for small as well as larger differences. Therefore, the distinction between one and two or three dimensions can also be found for the 'audiovisual' condition. In their explanation of why smaller differences are more difficult to perceive, Blumstein and Cooper (1972) note that in a discrimination task it is not necessary to analyse the auditory information into its linguistic components. A mere comparison of the phonetic properties of the two stimuli is sufficient. Therefore, they argue, the worse performance for the small differences can be explained by the fact that they are perceptually closer together. This argument is supported by the fact that the dimensions are phonetically conveyed differently: 'voicing' is based on temporal cues, while 'place of articulation' and 'manner of articulation' rely mainly on spectral cues. When 'voicing' and at least one of the spectral dimensions differ, both types of cues are involved, while in differences in one dimension only either temporal or spectral cues are altered. Therefore, one type of cue is the same in the stimuli, making a distinction more difficult.

The question of which dimension is most difficult to perceive for individuals with aphasia has been addressed previously with ambiguous results. Blumstein et al. (1977), for English, and Klitsch (2008), for Dutch, found 'place of articulation' to cause most difficulties. Saffran et al. (1976) and Caplan and Aydelott-Utman (1994), for English, and Csépe et al. (2001), for Hungarian, however, found 'voicing' to be most impaired, as in the current study. As mentioned above, Dutch and English, though both Germanic languages, differ in their phonetic realization of 'voicing' of plosives: While Dutch contrasts voiced (voice onset before lip-release) and voiceless-unaspirated (voice onset during lip-release) sounds, English shows a differentiation between voiceless-unaspirated (voice onset during lip-release) and voiceless-aspirated (voice onset after lip-release) (Jansen, 2004; Lisker & Abramson, 1964). This difference cannot explain the ambiguous results within the English data. It, however, makes it difficult to compare the English and Dutch data. A comparison of the Dutch and Hungarian data, on the other hand, is possible as both languages have a similar phonetic realization of 'voicing' (Jansen, 2004; Lisker & Abramson, 1964).

The difference in performance between 'voicing' on the one hand and the two other dimensions on the other hand could, for the 'audiovisual' condition, be explained by the fact that 'voicing' cues are considered to be not visible. As we, however, found the same pattern in the 'auditory' condition, thus without visual information, we suggest a different analysis: The difference between the phonetic dimensions can be explained by the different phonetic cues encoding them. As explained above, 'voicing' is phonetically conveyed by temporal cues, while 'place of articulation' and 'manner of articulation' are based on spectral cues. As distinctions in 'voicing' are most difficult for the aphasic subjects, they seem to have predominantly an impairment in processing the temporal cues necessary to perceive the difference between 'voiced' and 'voiceless'.

In conclusion, the current study shows that additional visual information (gained from speechreading) positively influences the discrimination abilities of aphasic subjects with a speech sound processing disorder. To what degree the individual phonetic dimensions are influenced could not be conclusively answered. Contrasts between items are more easily detected if they result from wider distinctions (more differing phonetic dimensions). Furthermore, the type of dimension differentiating items is of importance, indicating that differences in ‘voicing’ are most difficult to perceive for Dutch individuals with an aphasic disorder of speech sound processing.

Clinical implications

In the current study it was shown that different phonetic dimensions can be affected to different degrees in speech sound processing disorders. It is yet to be determined how this relates to more general comprehension tasks as lexical decision and word-picture matching or real-life comprehension. As lexical retrieval is dependent on the correct phonetic input, the auditory analysis of speech sounds is an important part of the comprehension process: It is the first step to accurate word processing. The actual influences of different phonetic dimensions on higher-level tasks and real-life comprehension, however, still need to be established in follow-up research. Only then can the next step, improving treatment, be made. However, Hessler and Stadie (2008) have shown that taking into account the phonetic structure of stimuli while utilizing speechreading is beneficial. The results of the current study give more information about the characteristics that need to be considered in developing treatment, such as the fact that distinctions in ‘voicing’ were more difficult to detect for the current group of aphasic subjects than those in ‘place of articulation’ or ‘manner of articulation’. If these differences actually effect higher processes and real-life comprehension as well, it should be investigated for all patients prior to treatment, which dimensions are especially problematic for them. Treatment for patients as the ones described in the current study should then include a focus on the timing cues necessary to perceive distinctions in ‘voicing’. The current study therefore provides not only the theoretical conclusions described above, but also preliminary clinical implications can be drawn.

Acknowledgements

This project was funded by an Ubbo Emmius Scholarship of the University of Groningen to the first author. Research of the second and third author was funded by the University of Groningen. The authors gratefully acknowledge Callista Jippes for help with the preparation of the material and students in Speech and Language Pathology for helping collecting the data of the control subjects. Furthermore, they would like to thank the aphasic and control subjects, speech therapists and rehabilitation centres for participating in this study.

***Declaration of interest*:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Note

1. We are aware that terminology is not consistent in the literature and both ‘speechreading’ and ‘lipreading’ have been used. In this paper we use the term ‘speechreading’ because the visual input received is not restricted to the lips, but rather covers the lower face, neck, and upper chest. This terminology has also been suggested by Campbell, Dodd, and Burnham (1998) in order to clearly state that more than just lip information is taken into account and to stress that what is read is indeed natural speech.

References

Bastiaanse, R., Bosje, M., & Visch-Brink, E. (1995). *Psycholinguistic Assessment of Language Processing in Aphasia: The Dutch version*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Blumstein, S.E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, *15*, 19–30.

Blumstein, S.E., & Cooper, W. (1972). Identification versus discrimination of distinctive features in speech perception. *Quarterly Journal of Experimental Psychology*, *24*, 207–214.

Buchman, A., Garron, D., Trost-Cardamone, J., Wichter, M., & Schwartz, M. (1986). Word deafness: One hundred years later. *Journal of Neurology, Neurosurgery and Psychiatry*, *49*, 489–499.

Campbell, R. (1988). Tracing lip movements: Making speech visible. *Visible Language*, *22*, 32–57.

Campbell, R. (1990). Lipreading, neuropsychology, and immediate memory. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 268–286). Cambridge: Cambridge University Press.

Campbell, R., Dodd, B., & Burnham, D. (Eds.). (1998). *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. ix–xiv). Hove: Psychology Press.

Caplan, D., & Aydelott-Utman, J. (1994). Selective acoustic phonetic impairment and lexical access in an aphasic patient. *The Journal of the Acoustical Society of America*, *95*, 512–517.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Csépe, V., Osman-Sági, J., Molnár, M., & Gósy, M. (2001). Impaired speech perception in aphasic patients: event-related potential and neuropsychological assessment. *Neuropsychologia*, *39*, 1194–1208.

Franklin, S. (1989). Dissociations in auditory word comprehension: evidence from nine fluent aphasic patients. *Aphasiology*, *3*, 189–207.

Gieliewski, E. (1989). Acoustic analysis and auditory retraining in the remediation of sensory aphasia. In C. Code & D. J. Muller (Eds.), *Aphasia therapy* (pp. 138–145). London: Whurr.

Graetz, P., De Bleser, R., & Willmes, K. (1992). *Akense afasietest (AAT)*. Amsterdam: Hogrefe.

Grayson, E., Hilton, R., & Franklin, S. (1997). Early intervention in a case of jargon aphasia: Efficacy of language comprehension therapy. *European Journal of Disorders of Communication*, *32*, 257–276.

Hessler, D., & Stadie, N. (2008). War die störungsspezifische Therapie der auditiven Analyse effektiv? Eine Einzelfallstudie bei Aphasie. In M. Wahl, J. Heide, & S. Hanne (Eds.), *Spektrum Patholinguistik 1*. Potsdam: Universitätsverlag Potsdam, 131–134.

Jansen, W. (2004). Laryngeal contrast and phonetic voicing (Doctoral dissertation, University of Groningen). *Groningen Dissertations in Linguistics (GRODIL)*, *47*.

Klitsch, J. (2008). Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia (Doctoral dissertation, University of Groningen). *Groningen Dissertations in Linguistics (GRODIL)*, *67*.

Kussmaul, A. (1877). *Die Störungen der Sprache*. Leipzig: Vogel.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Morris, J., Franklin, S., Ellis, A., Turner, J., & Bailey, P. (1996). Remediating a speech perception deficit in an aphasic patient. *Aphasiology*, *10*, 137–158.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 97–114). London: Lawrence Erlbaum.

Rosenblum, L.D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409.

Saffran, E., Marin, O., & Yeni-Komshian, G. (1976). An analysis of speech perception in word deafness. *Brain and Language*, *3*, 209–228.

Shindo, M., Kaga, K., & Tanaka, Y. (1991). Speech discrimination and lipreading in patients with word deafness or auditory agnosia. *Brain and Language*, *40*, 153–161.

Snodgrass, J.G., Levy-Berger, G., & Haydon, M. (1985). *Human experimental psychology*. New York: Oxford University Press.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*, 212–215.

Appendix

A: Stimuli

Table A.1: Pairs with identical stimuli

Stimuli			
/ba:f/ - /ba:f/	/bo:f/ - /bo:f/	/bœ͡yp/ - /bœ͡yp/	/du:p/ - /du:p/
/dɛ͡if/ - /dɛ͡if/	/dɔ͡uk/ - /dɔ͡uk/	/dɔ͡um/ - /dɔ͡um/	/dœ͡ys/ - /dœ͡ys/
/fe:t/ - /fe:t/	/fu:p/ - /fu:p/	/fɛ͡ip/ - /fɛ͡ip/	/fø:l/ - /fø:l/
/fø:p/ - /fø:p/	/fœ͡yp/ - /fœ͡yp/	/ke:m/ - /ke:m/	/kɔ͡uk/ - /kɔ͡uk/
/ky:m/ - /ky:m/	/la:p/ - /la:p/	/ly:m/ - /ly:m/	/lø:l/ - /lø:l/
/lø:p/ - /lø:p/	/ma:f/ - /ma:f/	/ me:m/ - /me:m/	/my:k/ - /my:k/
/mø:l/ - /mø:l/	/mœ͡yp/ - /mœ͡yp/	/ni:x/ - /ni:x/	/no:k/ - /no:k/
/pa:f/ - /pa:f/	/pi:x/ - /pi:x/	/py:k/ - /py:k/	/pɛ͡if/ - /pɛ͡if/
/pøm/ - /pøm/	/pœ͡yp/ - /pœ͡yp/	/sa:f/ - /sa:f/	/si:x/ -/si:x/
/sy:n/ - /sy:n/	/sø:m/ - /sø:m/	/sø:p/ - /sø:p/	/sɛ͡if/ - /sɛ͡if/
/ta:f/ - /ta:f/	/ti:x/ - /ti:x/	/tɔ͡uf/ - /tɔ͡uf/	/tɔ͡un/ - /tɔ͡un/
/vœ͡ys/ - /vœ͡ys/	/vɔ͡ul/ - /vɔ͡ul/	/xi:m/ - /xi:m/	/xø:p/ - /xø:p/
/xœ͡ys/ - /xœ͡ys/	/xɔ͡uk/ - /xɔ͡uk/	/za:f/ - /za:f/	/za:p/ - /za:p/
/zi:m/ - /zi:m/	/zi:x/ - /zi:x/		

Table A.2: Pairs with different stimuli

Condition (difference in)		Stimuli		
1 Dimension	Place	/fe:t/ - /se:t/	/fø:p/ - /sø:p/	/py:m/ - /ty:m/
		/pɛ͡if/ - /tɛ͡if/	/si:x/ - /fi:x/	/ti:x/ - /pi:x/
	Manner	/du:p/ - /nu:p/	/kø:p/ - /xø:p/	/kɔ͡uk/ - /xɔ͡uk/
		/sa:f/ - /ta:f/	/tɛ͡if/ - /sɛ͡if/	/ti:x/ - /si:x/
	Voicing	/ba:f/ - /pa:f/	/bø:m/ - /pø:m/	/dɛ͡if/ - /tɛ͡if/
		/dɔ͡uf/ - /tɔ͡uf/	/pœ͡yp/ - /bœ͡yp/	/tɔ͡um/ - /dɔ͡um/
2 Dimensions	Place & Manner	/fø:l/ - /tø:l/	/kɛ͡ip/ - /fɛ͡ip/	/pø:m/ - /sø:m/
		/tɔ͡uf/ - /xɔ͡uf/	/vœ͡ys/ - /dœ͡ys/	/xø:p/ - /tø:p/
	Place & Voicing	/bo:f/ - /to:f/	/dɔ͡um/ - /pɔ͡um/	/fi:x/ - /zi:x/
		/pɛ͡if/ - /dɛ͡if/	/za:p/ - /fa:p/	/zi:m/ - /xi:m/
	Manner & Voicing	/di:x/ - /si:x/	/dy:n/ - /sy:n/	/mœ͡yp/ - /pœ͡yp/
		/py:k/ - my:k/	/sɛ͡if/ - /dɛ͡if/	/zi:x/ - /ti:x/
3 Dimensions		/ba:f/ - /sa:f/	/dœ͡ys/ - /xœ͡ys/	/dɔ͡uk/ - /xɔ͡uk/
		/fœ͡yp/ - /dœ͡yp/	/fo:k/ - /no:k/	/fœ:l/ - /lœ:l/
		/ke:m/ - /me:m/	/kɔ͡ul/ - /vɔ͡ul/	/ky:m/ - /ly:m/
		/la:p/ -/fa:p/	/lø:p/ - /xø:p/	/ni:x/ - /fi:x/
		/nu:p/ - /fu:p/	/pi:x/ - zi:x/	/sø:m/ - /bø:m/
		/ta:f/ - /ma:f/	/tø:l/ - /mø:l/	/za:f/ - /pa:f/

B: Individual data

Table B.1. Individual results for the differences in 1, 2 and 3 dimensions in all three conditions.*

	Audiovisual			Auditory only			Visual only		
Initials	1 dim.	2 dim.	3 dim.	1 dim.	2 dim.	3 dim.	1 dim.	2 dim.	3 dim.
WB (W)	83%	100%	100%	83%	94%	94%	61%	72%	44%
(A')	(95%)	(99%)	(99%)	(94%)	(97%)	(98%)	(79%)	(84%)	(71%)
BB (G)	47%	63%	71%	39%	72%	61%	22%	44%	44%
(A')	(84%)	(89%)	(91%)	(77%)	(89%)	(85%)	(63%)	(67%)	(50%)
EK (A)	67%	94%	100%	72%	78%	94%	33%	44%	56%
(A')	(88%)	(97%)	(98%)	(92%)	(97%)	(98%)	(60%)	(68%)	(75%)
TB (G)	94%	100%	100%	72%	100%	100%	50%	50%	56%
(A')	(97%)	(99%)	(99%)	(88%)	(88%)	(97%)	(70%)	(70%)	(73%)
JH (M)	89%	94%	100%	83%	83%	100%	50%	0%	86%
(A')	(97%)	(99%)	(100%)	(96%)	(96%)	(100%)	(84%)	24%)	(95%)
MB (G)	72%	78%	78%	39%	83%	78%	33%	67%	33%
(A')	(87%)	(89%)	(89%)	(76%)	(92%)	(90%)	(52%)	(76%)	(52%)
Mean	75%	88%	91%	65%	87%	88%	42%	46%	53%
(A')	(91%)	(95%)	(96%)	(87%)	(95%)	(95%)	(68%)	(67%)	(74%)

* W = Wernicke's Aphasia, G = Global Aphasia, A = Anomia, M = Mixed Aphasia; A' = A'-Scores calculated according to Snodgrass et al. (1985).

Table B.2. Individual results for the different dimensions in all three conditions.*

	Audiovisual			Auditory only			Visual only		
Initials	Place	Manner	Voicing	Place	Manner	Voicing	Place	Manner	Voicing
WB (W)	83%	100%	67%	100%	100%	50%	67%	67%	50%
(A')	(95%)	(99%)	(90%)	(99%)	(99%)	(85%)	(82%)	(82%)	(74%)
BB (G)	67%	60%	17%	50%	50%	17%	33%	17%	17%
(A')	(90%)	(88%)	(72%)	(82%)	(82%)	(63%)	(86%)	(57%)	(57%)
EK (A)	83%	100%	17%	83%	67%	67%	33%	17%	50%
(A')	(93%)	(98%)	(66%)	(95%)	(91%)	(91%)	(60%)	(36%)	(71%)
TB (G)	100%	100%	83%	67%	100%	50%	50%	67%	33%
(A')	(99%)	(99%)	(94%)	(86%)	(97%)	(80%)	(70%)	(79%)	(58%)
JH (M)	100%	100%	67%	100%	67%	83%	100%	0%	33%
(A')	(100%)	(100%)	(92%)	(100%)	(92%)	(96%)	(99%)	(24%)	(78%)
MB (G)	50%	67%	100%	50%	50%	17%	50%	17%	33%
(A')	(78%)	(85%)	(96%)	(80%)	(80%)	(60%)	(66%)	(22%)	(52%)
Mean	81%	88%	58%	75%	72%	47%	56%	31%	36%
(A')	(92%)	(95%)	(85%)	(90%)	(90%)	(79%)	(75%)	(50%)	(65%)

* W = Wernicke's Aphasia, G = Global Aphasia, A = Anomia, M = Mixed Aphasia; A' = A'-Scores calculated according to Snodgrass et al. (1985).

C: Statistics with A'-scores

Table C.1. Comparison of conditions for the aphasic group using A'-scores.

Comparison	Test used	Test statistics	*df*	*p*
AV vs. AO vs. VO	Friedman Anova	$\chi^2 = 10.333$	2	.006
AV vs. AO	Wilcoxon test	$Z = -1.725$		.084
AV vs. VO	Wilcoxon test	$Z = -2.201$		.028
AO vs. VO	Wilcoxon test	$Z = -2.201$		.028
AV vs. AO (place)	Wilcoxon test	$Z = -.406$		.684
AV vs. AO (manner)	Wilcoxon test	$Z = -2.023$		.043
AV vs. AO (voicing)	Wilcoxon test	$Z = -.943$		.345

Table C.2. Comparison of dimensions for the aphasic group using A'-scores.

Comparison	Test used	Test statistics	*df*	*p*
Auditory only				
1 vs. 2 vs. 3 dimensions	Friedman Anova	$\chi^2 = 8.667$	2	.013
2 vs. 1 dimensions	Wilcoxon test	$Z = -1.826$		.068
3 vs. 1 dimensions	Wilcoxon test	$Z = -2.201$		.028
3 vs. 2 dimensions	Wilcoxon test	$Z = -1.089$		.276
place vs. manner vs voicing	Friedman Anova	$\chi^2 = 6.700$	2	.035
manner vs. place	Wilcoxon test	$Z = .000$		1.000
voicing vs. place	Wilcoxon test	$Z = -2.207$		.027
voicing vs. manner	Wilcoxon test	$Z = -1.753$		.080
Audiovisual				
1 vs. 2 vs. 3 dimensions	Friedman Anova	$\chi^2 = 11.143$	2	.004
2 vs. 1 dimensions	Wilcoxon test	$Z = -2.226$		.026
3 vs. 1 dimension	Wilcoxon test	$Z = -2.207$		.027
3 vs. 2 dimensions	Wilcoxon test	$Z = -1.633$		.102
place vs. manner vs. voicing	Friedman Anova	$\chi^2 = 4.727$	2	.094