Speech with Gaps: Effects of Periodic Interruptions on Speech Intelligibility

Alexander Schnotz Frank Digeser Ulrich Hoppe

Department of Audiology, ENT Clinic, University of Erlangen-Nuremberg, Erlangen, Germany

Key Words

Speech recognition · Intermittent speech · Redundancy · Auditory processing · Temporal cues · Glimpses

Abstract

Objective: This study examined the role of central auditory completion in speech understanding. The perception of periodically interrupted speech was investigated. For this purpose, gaps were inserted into speech signals by silencing out defined intervals. The main hypothesis was that word recognition increases with shorter gaps and is less influenced by the total amount of gaps. **Patients and Methods:** Seventeen normal-hearing young adults took part in this study. Phrases from the German HSM speech recognition test were used as speech material. The examination comprised 220 modulated sentences presented binaurally at 65 dB. Intervals with durations ranging from 50 to 700 ms were taken to silence out 50, 65 and 80% of each sentence, respectively. **Results:** Mean speech perception values were in the range of 65–92%, 35–92% and 35–95% correct answers for gap ratios of 50, 65 and 80%, respectively. When comparing the same interval duration, word recognition was better for smaller gap ratios. Both, gap ratio and gap duration had a significant influence on identification performance. **Conclusions:** Speech can be understood even when very large proportions are blanked out. A significant decrease in perception can be observed when gap duration exceeds more than half of the syllable length.

Copyright © 2009 S. Karger AG, Basel

Introduction

Speech is a highly redundant signal. This redundancy leads to the effect that speech which is partially drowned out can be reconstructed from its remaining parts. Bad wireless broadcasting or dishes clattering are situations in which listeners ‘fill in’ the missing speech information [1].

Auditory perception of speech is usually carried out by a three-component system consisting of a peripheral auditory, central auditory and cognitive component. This conception is based on underlying anatomy and physiology. It was originally brought up by a working group of the National Research Council’s Committee on Hearing and Bioacoustics and Biomechanics in 1991. A fair amount of overlap between the ‘central auditory’ and the ‘cognitive’ components occurs within the framework [2]. Besides an awareness of speech defined by the sensitivity of the auditory system listeners need auditory processing abilities that consist of background and lexical knowledge [3], semantic knowledge [4], temporal resolution ability [5], prosodic cues [6], memory and reasoning [7].

Though there are several concurrent theories for speech perception, most of them assume that the acoustic input (e.g. a syllable) evokes activity in a set of similar sounding acoustic-phonetic patterns in memory. The degree of activation is determined by the degree of similarity between stored mind patterns and perceived input. Abstract models for the process of speech understanding

usually include excitatory and inhibitory pathways with subsequent specific word decision units.

Speech perception in a quiet environment can be sufficiently predicted for hearing-impaired listeners based upon their pure-tone hearing thresholds [8]. However, results scatter widely when speech perception in noise is measured, especially when fluctuating noises are used [9]. Listeners are usually able to interpolate a substantial amount of nonperceived speech segments and take advantage of 'glimpses' [9]. Stuart and Phillips [10] performed experiments on word recognition in interrupted broadband noise. They experimented with gated maskers, i.e. noise bursts whose duration varied randomly between 5 and 95 ms. Nelson et al. [1] performed a similar investigation with background noise periodically gated at frequencies between 1 and 32 Hz. These studies have shown that speech reception is substantially improved with background maskers changing in amplitude compared to stationary ones. Improvement was higher in younger subject groups than in groups of elderly people with normal hearing thresholds [10]. The possibility of detecting important speech cues in the unmasked parts of sentences results in higher recognition scores than expected from the audible fraction [9]. Powers and Wilcox [11] and recently Shinn-Cunningham and Wang [12] have investigated recognition of interrupted speech at different signal-to-noise ratios with and without intervening noise. These experiments showed that intelligibility is improved when the interruptions are filled with broadband noise compared to silent gaps. This effect is known as 'continuity effect' or 'auditory induction' and may be explained either by differences caused by the influence of the noise on peripheral auditory processing or by differences in central speech detection [13]. Bashford et al. [14] showed that interruptions in the speech signal were clearly detectable for gaps above 100 ms. If noise was inserted into the interruptions, the threshold for clear identification of the gaps increased to above 280 ms. It was also shown that the detection threshold for gaps in spoken language is also dependent on the number of spoken syllables per time unit. In summary, these limits of 'illusory continuity' correspond to a fixed amount of verbal information, and not to a fixed temporal value [14].

All investigations mentioned above have in common that speech and gaps were fractionated 1:1 (= 50% gap ratio, gr). The aim of this article was to describe the influence of gaps in speech signals on speech understanding for different amounts of interruptions (i.e. gap ratios) and fragmental lengths (gap durations). In particular, gap ratios of 50, 65 and 80% were investigated.

Materials and Methods

Participants

Seventeen young healthy adults took part in the study. Prior to the investigation, hearing thresholds were determined using a standard clinical audiometer (Auritec AT900 with a Sennheiser HDA200 headphone) in a sound-treated room. Only subjects with hearing thresholds below 20 dB HL at octave frequencies between 500 and 8,000 Hz were included. Furthermore, all subjects showed type A tympanograms and regular transient evoked otoacoustic emission responses. Age of the investigated subjects (6 men, 11 women) ranged between 19 and 42 years with a mean age of 23.6 ± 2.4 years. Subjects were students and staff at the Medical School of Erlangen.

Speech Material

The Hochmair-Schulz-Moser test was used as test material. It is delivered digitally on a compact disc (44,100 Hz sample rate, 16 bit resolution) and consists of 30 lists containing meaningful sentences. Each list contains 106 words forming 20 sentences of 4–8 words each with a mean of 5.3 words per sentence. The speech material is spoken by a male speaker at a syllable frequency of 3.7 ± 0.66 Hz, which corresponds to a medium speech rate. The difficulty level corresponds to sentences used in everyday life. The Hochmair-Schulz-Moser test is characterized by low variation in the perceptibility of the different test lists [15].

The speech material was processed using Matlab™ (release 14). The sentences were divided into several intervals of equal length by applying a modulation function. This function is +1 during the speech interval and zero for the gap. A Gaussian-shaped gated signal with rise and fall times of 10 ms between 10 and 90% of the signal amplitude was introduced for the transition [16]. The original length of the sentence remained unaffected. The phase of the modulation function was randomized in order to avoid a systematic influence on the speech material.

A schematic representation of the modulation pattern used is given in figure 1. Sound pressure versus time is shown for a complete sentence along with the same sentence processed under four different experimental conditions. In the example given modulation was performed with a 50% gap ratio. From top to bottom, 50% was erased from every section of 50, 100, 200 and 400 ms (i.e. interval durations), while the gap durations were 25, 50, 100, and 200 ms. In order to keep measurement time below 1 h, for one presentation setting (i.e. interval duration and gap ratio) 10 sentences with 53 words in total were used.

In this study, three different gap ratios, namely 50, 65 and 80%, were investigated. Appropriate interval duration values were determined in a few preliminary experiments. Interval durations varied between 50 and 700 ms. Details of the investigated settings are given in table 1. According to the score, perception rates of almost 100% were assumed for short interval durations and perception rates of nearly 100% – gr for the longest interval durations.

Test Procedure

Experiments were performed in one session for each subject with a complete duration of about 90 min. Subjects were seated in a comfortable armchair in an sound-treated room. At first, subjects were familiarized with the test procedure and the periodically interrupted speech. They were instructed to repeat sentence by sentence immediately after the end of a sentence. Subjects were

Fig. 1. Sound pressure as a function of time for a sample sentence. The top panel shows the unprocessed complete phrase ('Wir möchten ein Zimmer mit Blick aufs Meer.'). The other panels show modified versions of the same sentence for interval durations of 50–400 ms. In all plots, the gap ratio is 50%.

asked to guess whenever parts of the sentence were not perceived. For later verification the answers of the patients were recorded with a digital recorder. Each correctly repeated word was scored. Speech perception rate was defined as the number of correctly repeated words divided by the total number of presented words (i.e. 53). Speech material was presented using a clinical audiometer (AT900, Auritec) and headphones (Sennheiser HDA200) at 65 dB SPL. The order of sentence presentation was the same for all subjects and is listed in table 1.

Results

Figure 2 shows the mean perception scores and corresponding standard deviations as a function of interval duration. Speech recognition decreased with increasing interval duration and with increasing gap ratio. However, differences can be observed for the different gap ratios.

50% Gap Ratio

With the exception of one score at interval duration (ID) = 400 ms a large uniform decrease in perception rate is shown. For IDs <400 ms 4 subjects exhibited recognition of 100%. For IDs >400 ms no subject reached 100% perception rate. Mean values range from 92% at ID = 250

ms to 65% at ID = 700 ms. A mean value of 50% seems to be a reasonable prediction for very large interval durations: an interval duration of a whole sentence length will lead to the muting of half a sentence (approximately 50% of the words). Hence, about 50% of the words will be recognized. In our study this limit of 50% perception rate was not obtained, due to accentuation on interval durations that were shorter than the whole sentence length.

65% Gap Ratio

Compared to the 50% condition, all perception rates were lower for the 65% condition when comparing the same interval duration. Two subjects achieved a perception rate of 100% for interval durations of 100 ms (mean: 92%). For interval durations above 100 ms none of the subjects reached 100% perception rates. Surprisingly, at ID = 350 ms the perception rate was lower than for 400 ms. For high gap ratios of 65%, we expect a perception rate of 35% for the same reasons as described above. This score was reached for the mean value at ID = 500 ms.

80% Gap Ratio

As expected, perception rates for conditions with gap ratio = 80% were the lowest compared to 50 and 65% con-

Table 1. Summary of investigated conditions

Con- dition No.	ID ms	gr %	Speech duration ms	GD ms	HSM list
1	700	50	350	350	17-1
2	550	50	275	275	18-2
3	500	50	250	250	19-1
4	450	50	225	225	19-2
5	400	50	200	200	20-1
6	350	50	175	175	20-2
7	300	50	150	150	21-1
8	250	50	125	125	21-2
9	500	65	175	325	22-1
10	450	65	157	293	22-2
11	400	65	140	260	14-1
12	350	65	122	228	14-2
13	300	65	105	195	15-1
14	250	65	87	163	15-2
15	200	65	70	130	16-1
16	150	65	52	98	16-2
17	100	65	35	65	18-1
18	250	80	50	200	10-1
19	200	80	40	160	10-2
20	150	80	30	120	11-1
21	100	80	20	80	11-2
22	50	80	10	40	12-1

Interval duration (ID) denotes the length of the processed speech intervals; gap ratio (gr, %) refers to the amount of speech deleted in every sentence; GD refers to the deleted time in milliseconds. The last column shows the investigated list of the Hochmair-Schulz-Moser (HSM) test.

ditions. A recognition rate of 100% was reached by 5 subjects for interval durations of 50 ms. For interval durations of 50–250 ms, the mean perception rate dropped from 95 to 35%. This decrease was remarkably steep and strictly monotonous. The expected limit of 20% for large interval durations was not determined since interval durations were restricted to a maximum of 250 ms.

A repeated analysis of variance (ANOVA) was used to estimate the significance of the differences in recognition performance across the different interval durations and gap ratios. Significant main effects were identified for gap ratio [$F(2, 11) = 244.01$, $p < 0.0001$] and for interval duration [$F(2, 11) = 66.22$, $p < 0.0001$].

In order to describe perception as a function of information reduction, a least square fit was conducted with a sigmoid fit function. We slightly modified the fit function of Kollmeier and Wesselkamp [17] to take the differ-

Fig. 2. Mean perception rate across all listeners is plotted as a function of interval duration for gap ratios of 50, 65 and 80%. Error bars indicate $\pm$ one standard deviation. The solid lines illustrate results of the fitted psychometric function.

ent limits into account: for large interval durations perception converges on the gap ratio as described above. For very small interval durations, the intelligibility is forced to be nearly 100%. Hence, intelligibility P depends on interval duration and gap ratio (gr) as described in equation 1:

$$P(ID, gr[\%]) = \frac{gr[\%]}{1 + e^{4m_0(ID - ID_0)}} + (100 - gr[\%]) \quad (1)$$

As mentioned above, gap ratios of 50, 65 and 80% were investigated and interval durations varied between 50 and 700 ms (table 1). Obviously, equation 1 fulfils the above-listed requirements for sigmoid function. At $ID = ID_0$ perception is $100\% - (gr[\%])/2$. Mathematically, ID_0 corresponds to the point of inflection of the sigmoid function. Obviously, this value depends upon the specific gap ratio. For our setting, it was 75, 67.5 and 60% for gap ratios of 50, 65, and 80%, respectively. The parameter m_0 represents the slope of the perception function at the point of inflection.

For gap ratios of 50%, the fit procedure yields parameters $m_0 = (1.4 \pm 0.2) \cdot 10^{-3} \cdot 1/\text{ms}$ and $ID_0 = (458 \pm 16)$ ms. For $gr = 65\%$ resulting parameters were $m_0 = (3.7 \pm 0.6) \cdot 10^{-3} \cdot 1/\text{ms}$ and $ID_0 = (265 \pm 11)$ ms. For a gap ratio of 80%, $m_0 = (5.7 \pm 0.5) \cdot 10^{-3} \cdot 1/\text{ms}$ and $ID_0 = (197 \pm 4)$ ms were obtained.

Figure 3 shows the same data for the perception rates shown in figure 2 as a function of gap duration. Gap duration (GD) can be calculated from interval duration by equation 2:

$$GD = ID (gr[\%]/100) \quad (2)$$

For example, for $gr = 50\%$, GD is exactly half of the corresponding interval duration. While the perception rates show a similar course as in figure 2, they seem to be less influenced by the amount of informational loss, i.e. the different gap ratios.

In order to apply the same fit procedure to the dependency of the perception rate on the measured gap durations, equation 2 was used:

$$P(GD, gr[\%]) = \frac{gr[\%]}{1 + e^{-4m_0(GD - GD_0)}} + (100 - gr[\%]) \quad (3)$$

Again, GD_0 is the point of inflection and m_0 is the slope at this point. Results of the fit were $m_0 = (3.6 \pm 0.9) \cdot 10^{-3} \cdot 1/\text{ms}$ and $GD_0 = (214 \pm 12) \text{ ms}$ for gap ratios of 50%, $m_0 = (4.7 \pm 0.7) \cdot 10^{-3} \cdot 1/\text{ms}$ and $GD_0 = (181 \pm 7) \text{ ms}$ and for gap ratios of 65% as well as $m_0 = (7.2 \pm 0.6) \cdot 10^{-3} \cdot 1/\text{ms}$ and $GD_0 = (158 \pm 3) \text{ ms}$ for test sentences measured at 80% gap ratio.

Discussion

This study shows that speech recognition is reduced when short gaps are inserted in sentences. Two main determining factors were investigated for speech recognition: the duration of the gaps and the complete length of the manipulated intervals. Speech recognition depends to a similar extent on both and can be described by a sigmoid curve starting at about 100% for small gap durations and interval duration with the steepest slope at about 150–200 ms. However, distinct differences in the courses of the recognition functions were observed.

Influence of Gap Ratio

Deletions of 50, 65 and 80% were investigated. When other parameters were kept constant, a constant decrease from 50 to 80% in speech perception was observed. However, when gap durations were below 100 ms, mean speech recognition rates were above 85% for all gap ratios investigated. For gap ratios of 50%, the influence of gap insertion was very moderate and individuals rarely experienced a significant problem understanding speech. For higher gap ratios listeners usually had to make additional central efforts in order to recover the original speech signal.

Fig. 3. Mean perception rate across all listeners plotted as a function of gap duration for gap ratios of 50, 65 and 80%. Influence of gap duration on mean perception rate. For the investigated gap ratios of 50, 65 and 80%, mean perception rates across all listeners are shown as a function of gap duration.

Influence of Gap Duration

In particular, when gap duration was in the range of syllable length, speech recognition depended substantially on gap duration. As shown in figure 3, the steepest decrease in recognition function was observed at gap durations of about 150 ms for all gap ratios. Since the mean syllable duration of the test was at about 270 ms, the proposed 'half-syllable' condition was nearly met at the point of maximal slope.

Syllables that are only heard by half can be reconstructed completely [18]. When gap durations are above the syllable duration of 270 ms more than one syllable lies within a gap. Therefore, the main contributing factor for the reduced speech recognition here is inaudibility. For gap durations that cover whole syllables or more, the corresponding speech segment may be completely inaudible, leading to recognition scores of about the fraction of the blanked intervals (i.e. gap ratios). Central completion is not based on the reconstruction of partially destroyed syllables but rather on known rules of grammar [18].

Looking at a fixed gap ratio, the perception rate decreases evenly with increasing gap duration. However, distinct sharp decreases are observable for gap ratios of 65 and 50% at gap durations of 228 and 200 ms, respectively. This might be related to the average syllable length

which is in the same range. This may be investigated in future studies using different syllable rates.

The perception rate as a function of both interval duration and gap duration was modelled by a sigmoid function. These functions can be described by the point of inflection (ID_0 and GD_0 , respectively) and the slope m_0 at this point. Fitting for both variables yielded a monotonous increase in the slope and decrease for the point of inflection with increasing gap ratio. However, this dependency was less for gap duration. This might indicate that the absolute duration of gaps in speech is more important than the duration of manipulated intervals or the gap ratio.

One might criticize that gap positions in the test were randomly distributed and that it is important what information in detail is exactly discarded. Considering that syllable duration is not constant throughout the test, it was necessary to implement a fixed pattern for the modulation. Because the interruptions employed in these studies were not linked systematically to specific speech components, interpretation of the results in terms of possible sampling limits for perceptual restoration must be considered as statistical [14].

The test presented here may be used in clinical practice for the investigation of central auditory processing disorders. For long gap durations the test results are primarily determined by central completion abilities (i.e.

top-down processes). For short gap durations, the perception rate may be assumed to be more a peripheral evaluation (bottom-up process). Therefore, the particular dependency of perception rate on gap duration may deliver information on auditory speech processing at different time scales.

Conclusion

The study presented here shows that massively destroyed speech can be reconstructed and correctly understood. The study shows that the perception of interrupted speech signals does not only depend on the percentage amount of deletions but also on the absolute duration of individual deleted intervals. A remarkable correlation between syllable duration and critical gap duration was identified. The test may be used for quantifying the reconstruction abilities of patients with central auditory processing disorders.

Acknowledgments

The authors gratefully acknowledge Dr. Gary Weismer for his helpful comments. The study was supported by the Johannes und Frieda Marohn-Stiftung, Erlangen.

References

- 1 Nelson PB, Jin SH, Carney AE, Nelson DA: Understanding speech in modulated interference: cochlear implant users and normal hearing listeners. *J Acoust Soc Am* 2003;113:916–969.
- 2 Jerger S, Jerger S, Pirozollo F: Correlational analysis of speech audiometric scores, hearing loss, age and cognitive abilities in the elderly. *Ear Hear* 1991;12:103–109.
- 3 Stuart A: Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear Hear* 2005;26:78–88.
- 4 Möbes J, Matzke M, Lamers M, Rüssler J, Münte TF: Electrophysical correlates of the perception of time-inverted speech. *Klin Neurophysiol* 2003;34:75–79.
- 5 Nie K, Barco A, Zeng FG: Spectral and temporal cues in cochlear implant speech perception. *Ear Hear* 2006;27:208–217.
- 6 Grant KW, Seitz PF: The recognition of isolated words and words in sentences: individual variability in the use of sentence context. *J Acoust Soc Am* 2000;107:1000–1011.
- 7 Humes LE: Do ‘auditory processing’ tests measure auditory procession in the elderly. *Ear Hear* 2005;26:109–119.
- 8 Holube I, Kollmeier B: Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. *J Acoust Soc Am* 2006;100:1703–1716.
- 9 Cooke M: A glimpsing model of speech perception in noise. *J Acoust Soc Am* 2006;119:1562–1573.
- 10 Stuart A, Phillips DP: Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing and presbycusis listeners. *Ear Hear* 1996;17:478–489.
- 11 Powers GL, Wilcox JC: Intelligibility of temporally interrupted speech with and without intervening noise. *J Acoust Soc Am* 1977;61:195–199.
- 12 Shinn-Cunningham BG, Wang D: Influence of auditory object formation on phonemic restoration. *J Acoust Soc Am* 2007;123:295–301.
- 13 Miller GA, Licklider JCR: The intelligibility of interrupted speech. *J Acoust Soc Am* 1950;22:167–173.
- 14 Bashford JA, Meyers MD, Brubaker BS, Warren RM: Illusory continuity of interrupted speech: speech rate determines durational limits. *J Acoust Soc Am* 1988;84:1635–1638.
- 15 Hochmair-Desoyer I, Schulz E, Moser L, Schmidt M: The HSM sentence test as a tool for evaluating speech understanding in noise of cochlear implant users. *Am J Otol* 1997;15(6 suppl):S83.
- 16 Zwicker E, Fastl H: *Psychoacoustics*. Berlin, Springer, 1999.
- 17 Kollmeier B, Wesselkamp M: Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J Acoust Soc Am* 1997;102:2412–2421.
- 18 Greenberg S, Arai T: What are the essential cues for understanding spoken language? *IEICE Trans Inf Syst* 2004;E87D:1059–1070.