Time for a shave? Does facial hair interfere with visual speech intelligibility?

Chapter · January 2010

DOI: 10.13140/2.1.2903.7123

6 authors, including:

Susanne Fuchs
Centre for General Linguistics

Melanie Weirich
Friedrich Schiller University Jena

Christian Kroos
University of Surrey

Daniel Pape
McMaster University

X

Time for a shave? Does facial hair interfere with visual speech intelligibility?

SUSANNE FUCHS
MELANIE WEIRICH
CHRISTIAN KROOS
NATALIE FECHER
DANIEL PAPE
SABINE KOPPETSCH

Abstract: Watching the speaker's face under noisy auditory conditions generally increases the intelligibility of speech. The aim of this study is to investigate the impact of facial hair, covering parts of the articulators, on visual speech intelligibility under noise. We hypothesized that facial hair decreases intelligibility, leads to a longer reaction time and a lower confidence to perceive the relevant target word. Thus, is it time for a shave now? Three speakers were recorded with a video camera in three different conditions: wearing no beard, a moustache and a long chin beard. The moustache and the long chin beard consisted of natural hair which was attached to surface of the face. Speakers uttered the relevant sentences in all three conditions. These video recordings provided the basis for an audio-visual perception experiment: 44 subjects, separated in three groups, perceived either the speakers in the no beard condition or the moustache condition or the long chin beard condition. Audio only and audiovisual data were presented in a randomized order. Subjects were instructed to type in a text input field what they perceived. Their reaction time was measured and they were asked to rate how confident they were in perceiving the target word. Our findings show that the impact of facial hair (moustache) on speech intelligibility is generally rather small and not significant, but subjects seeing speakers with a moustache improve more from the audio only to the audiovisual condition in comparison to those who see speakers with no beard. Moreover, they need a longer reaction time and they are less confident to perceive the relevant target word. We interpret this result with respect to a greater attentiveness when subjects are confronted with visual impoverishment of the articulators.

1. INTRODUCTION

If one walks through the first level of the main building at the Humboldt University in Berlin and looks at all the portraits of researchers who studied there, became professors, and in some cases won a Nobel prize, one may come to the conclusion that the most important visible sign for a famous person is to be a man and to have a beard. About 80 per cent of the various portraits in this gallery show men with a beard (except for some of the philosophers and the women), independent of the generation. Nowadays, this trend might be in a transition towards a more beardless fashion, since more females become famous, the equipment to shave facial hair has improved. However, in our community of speech researchers, there is clearly no lack of famous men with facial hair (see Figure 1).

This paper is not concerned with the social status attributed to facial hair on a man (see Appendix for some recent contributions in magazines on this topic), but with the influence of facial hair on audio-visual speech intelligibility in noise.

It is generally known that watching the speaker's face increases the intelligibility of speech in noisy environments (Sumby & Pollack, 1954; Grant & Seitz, 2000). By observing the cyclical opening and closing of the visible jaw, one may be able to identify the rhythmic structure of the spoken utterance or even the focus of a particular sequence (Dohen, Lœvenbruck & Hill, 2005). Moreover, certain minimal pairs, e.g. the /m/-/n/ contrast, may be better identified with the additional visual information of their articulators than with only the purely acoustic information (Schwartz et al., 2002).

Figure 1: Well-known researchers from the area of speech production and perception with various beard shapes. All photos have been printed with the permission of the individuals.

Facial hair can cover parts of the face such as the upper lip, the teeth or the larynx. If the upper lip and teeth are partially hidden, the area of the mouth opening is modified, and facial hair is responsible for a kind of *natural visual impoverishment* of the visual intelligibility of speech. Under normal auditory conditions such an impoverishment may be marginal for the intelligibility of speech, since the auditory information is fully available. However, under noisy auditory conditions, the visual cues may be crucial for increasing speech intelligibility. Based on these considerations, we hypothesize that:

(1) Facial hair hiding visible articulatory movements leads to lower speech intelligibility under noisy auditory conditions, a longer reaction time, and a lower confidence in recognizing the relevant target words.

(2) The shape and location of the beard is crucial for the reduced speech intelligibility in noise. A moustache hiding the upper lip movement has a larger impact on visual speech intelligibility than a long chin beard, hiding the larynx only.

So in terms of speech intelligibility is it now time for a shave or not?

2. METHODOLOGY

Investigating the interference of facial hair with visual speech intelligibility poses the problem of accurately controlling the amount and shape of facial hair across several speakers while at the same time keeping the recording situation constant. Since it is difficult to find participants willing to grow and then cut their beards as needed for this study and since this would lead to large time spans between the recordings, we decided to use artificial beards made from natural hair. Two different beard types were chosen, a moustache and a long chin beard.

2.1. STIMULI

2.1.1. VIDEO RECORDINGS

Three male speakers (hereafter SP1, SP2, and SP3) in their mid 20s were recorded (see Figure 2). None of them had substantial natural facial hair above 3mm during the time of the recording.

Figure 2: Subjects 1, 2 and 3 without beard (left), with a moustache (middle) and with a long chin beard (right).

The speakers were selected according to their hair colour and texture, which had to fit the colour and texture of the attached facial hair. Two types of facial hair were obtained in a specialist mask shop. They consist of natural hair which is woven into a strip of gauze that provides an attachment area. With the help of professional glue (Mastix) used in the film industry, the gauze strip was attached to the facial surface.
Every speaker read a set of 40 sentences (each of the 20 target words in two different carrier sentences) in three conditions: with no beard (beard0), with the moustache (beard1), and with the long chin beard (beard2). The recordings were carried out in a soundproof room at the

Centre of General Linguistics (ZAS) in Berlin using a Sony DCR-TRV14E video camera. The speaker was seated approximately 80 cm away from the camera in front of a uniformly white background and was uniformly lit. The video frame comprised the entire head and neck of the speaker.

2.1.2. SPEECH MATERIAL

Twenty nouns were selected as target words on the basis of their high frequency in the lexicon (Haas, Knetschke & Sperlbaum, 1984) and their semantic content (see Table 1). Their meaning had to fit four carrier sentences without allowing the prediction of the target word from the semantic context of the carrier sentence.

Table 1: Speech material with target words (English translation), phonological structure and IPA (Standard German), frequency in the lexicon, and number of voiceless obstruents.

Target Word	Phonological structure and IPA		Frequency	No. of vl. obstruents
Abend (evening)	VCVCC	/ˈa:bənt/	129	0
Beispiel (example)	CVVCCVC	/ˈb̥æʃpi:l/	774	2: /ʃp/
Beruf (job)	CVCVC	/b̥ɛˈru:f/	342	1: /f/
Bücher (books)	CVCVC	/ˈb̥y:çɐ/	128	1: /ç/
Dinge (things)	CVCV	/ˈdɪŋə/	216	0
Freude (happiness)	CCVVCV	/ˈfrɔɪdə/	158	1: /f/
Kinder (children)	CVCCVC	/ˈkɪndɐ/	628	1: /k/
Klasse (class)	CCVCV	/ˈklasə/	180	2: /k, s/
Leben (life)	CVCVC	/ˈle:bən/	215	0
Leute (people)	CVVCV	/ˈlɔɪtə/	544	1: /t/
Mädchen (girl)	CVCCVC	/ˈmɛ:dçən/	198	1: /ç/
Musik (music)	CVCVC	/muˈzi:k/	156	1: /k/
Schule (school)	CVCV	/ˈʃu:lə/	719	1: /ʃ/
Seite (side)	CVVCV	/ˈzætə/	187	1: /t/
Sommer (summer)	CVCVC	/ˈzɔmɐ/	249	0
Strasse (street)	CCCVCV	/ˈʃtra:sə/	142	3: /ʃ, t, s/
Vater (father)	CVCVC	/ˈfa:tɐ/	289	2: /f, t/
Wasser (water)	CVCVC	/ˈvasɐ/	163	1: /s/
Wetter (weather)	CVCVC	/ˈvɛtɐ/	206	1: /t/
Woche (week)	CVCV	/ˈvɔxə/	181	1: / x/

There were four carrier sentences:

Meine Freundin hat __ gesagt. (My friend said __ .)
Auf dem Blatt hat __ gestanden. (__ was written on the paper.)
Meine Mutter hat __ geschrieben. (My mother wrote ___ .)
Mein Nachbar hat __ geschrien. (My neighbour cried __ .)

For the video recording every target word occurred in two of the four carrier sentences, but for further analysis only one of the two sentences was selected to make the perception experiment feasible in terms of its duration. Target words are all bi-syllabic. In general, we tried to make the corpus as phonetically balanced as possible (see Table 1).

2.1.3. STIMULI PROCESSING

From the continuous video recording short clips of equal length were extracted containing one sentence each. The starting point was set at 0.5s before the speaker uttered the first syllable and the endpoint at 0.5s after the completion of the sentence. Using Adobe Premiere 1.0 the original video frames of 720 x 576 pixels were cut on the left and right side to be square-sized (576 x 576 pixels) and if necessary, slightly adjusted in size to position all speakers in the exact same size at the exact same location in the frame. The original sound track was extracted and its average loudness computed. Commercial multi-speaker babble noise was added with its loudness set to result in a final signal-to-noise ratio of 3dB. For the audio-only condition the video frames showing the speaker were replaced by black frames. The acoustic signals were low-pass filtered at 12dB/octave with a cut-off frequency of 8000 Hz in order to remove high-frequency noise from the recording and, more importantly, to match the frequency range of the stimuli to the frequency range of the multi-speaker babble noise.

2.2. DESIGN

To avoid any participant seeing the same speaker with different beards and becoming aware of the aim of the study, the beard condition was made a between-subject factor while 'speaker' was kept within-subject. Thus, a participant would see all three speakers with the same beard type. Furthermore, despite the fact that each speaker was recorded in all beard conditions in a single session, changes in the acoustics among the different beard conditions cannot be ruled out. The audio-only control condition (A) was designed to mirror the audio-visual (AV) condition: The audio data in the A condition are the same as those in the AV condition, but they are presented with a black video. Each target word plus carrier sentence was presented in six different versions (beard0-AV, beard1-AV, beard2-AV, beard0-A, beard1-A, beard2-A), not counting repetitions.

2.3. PROCEDURE

The experiment control software Alvin (see Hillenbrand and Gayvert, 2005) was used to present the stimuli on a monitor and register the participants' responses. The participants were seated approximately 50cm away from the monitor and listened to the stimuli via Sennheiser HD 201 headphones. They were instructed to type the target word into a text input field on the screen using the computer keyboard as soon as they thought they had recognized it after (or while) watching the stimulus video clip. They were told that their response times were measured by pressing the enter button after they typed in the perceived word. Subjects were subsequently prompted to rate their confidence in having correctly identified the target word on a 5-point Likert scale by pressing a software button with the computer mouse. The test trials (20 target words * 6 conditions * 2 repetitions) were preceded by 5 practise trials. The experiment took approximately 30 minutes.

2.4. PARTICIPANTS

Forty-six participants took part in the experiment. Two of them had to be excluded after the experiment as it became apparent that they were (for different reasons) familiar with the target words. The participants were randomly assigned to one of three stimuli groups (speakers without beard, speakers with a moustache, speakers with a long chin beard) though across groups the same gender ratio (4 males, 11 females, except beard1 with only 10 females) and a similar age range (between 20 and 35 years) were maintained. The participants received 10 euro for their participation.

2.5. STATISTICAL DESIGN

The statistical package R (R Development Core Team, 2008) was used for further quantitative analyses. Since our dataset was not equally well balanced (only data from 14 subjects were included in the beard1 condition whereas data from 15 speakers were included in the beard0 and beard2 conditions), we used linear mixed models. Linear mixed models consist of fixed effects and random effects. Beard condition (beard0, beard1, beard2) served as the fixed effect and listener and target word as random effects. Two packages were loaded for the analyses: the lme4 (linear mixed models) and the language R (to use the Monte Carlo Markov Chain method and provide pMCMC values). The details are given in the results section.

3. RESULTS

Intelligibility: There is clear evidence that speech intelligibility increases when watching the speaker's face (AV) in comparison to the audio-only (A) condition. This increase is on average 17 % for no beard, 20 % for the moustache and 12 % for the long chin beard (see Figure 3). Speakers with a moustache have in all cases the lowest speech intelligibility, whereas speakers with a long chin beard have a similar or even better intelligibility in comparison to speakers with no beard. Running a linear mixed model for the A data with beard condition as the fixed factor and listener and target word as random factors provides evidence that the moustache differs from the other two beard conditions: from beard0 (no beard) with pMCMC=0.043 and from beard2 (long chin beard) with pMCMC=0.0002. Similar significant differences could not be found for the AV data.

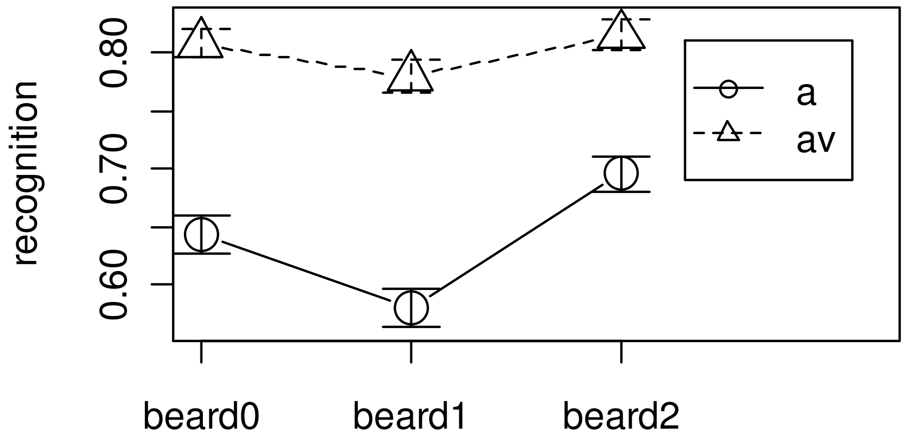

Figure 3: Means and standard errors for target word recognition; a = audio data (circle markers), av = audio-visual data (triangle markers), 3 beard conditions: beard0=no beard, beard1=moustache, beard2=long chin beard.

Reaction times: Reaction times were first transformed into a logarithmic scale to make them more normally distributed. Second, we checked whether the linear mixed model provided a good fit to the data. A quantile-quantile plot showed that the model was stressed when it tried to fit the longest reaction times. Hence, we removed the extreme outliers with a standardized residual at a distance greater than 2 standard deviations from zero (trimmed model in chapter 7 of Baayen, 2008). These extreme outliers (up to 22000 ms) were probably due to typos or to the fact that listeners forgot to press the enter button after typing the relevant word.

Our findings provide evidence that reaction times are similar for the A and AV data (Figure 4); hence we pooled all data together. However, results differ significantly with respect to the beard condition. Subjects showed significantly longer reaction times in the beard1 condition (moustache) in comparison to beard0 (pMCMC=0.0042) and beard2 (pMCMC=0.0001).

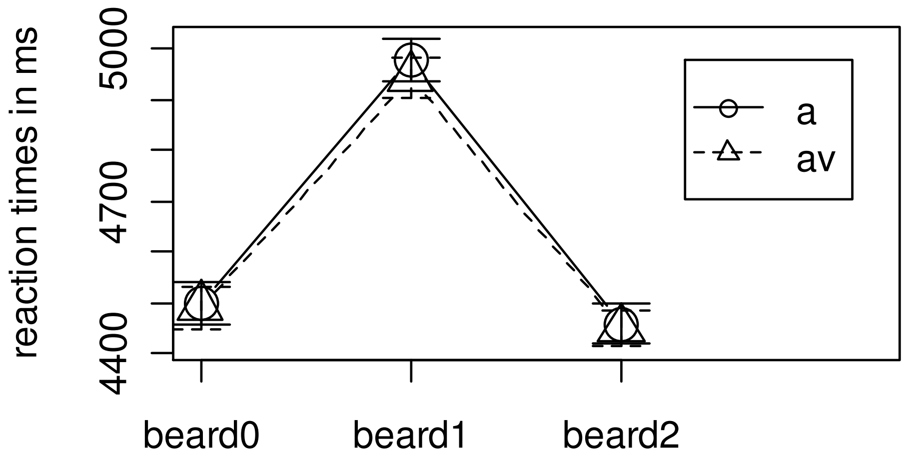

Figure 4: Means and standard errors for reaction times (real data in ms); a = audio data (circle markers), av = audio-visual data (triangle markers), 3 beard conditions: beard0=no beard, beard1=moustache, beard2=long chin beard.

Listeners' confidence: Similarly to intelligibility, subjects were most confident on a 5-point Likert scale when they rated the AV data for the speakers with a long chin beard (Figure 5). They were least confident when they rated the audio-only data for speakers with a moustache. The confidence level was significantly larger in the video condition than in the audio-only condition (pMCMC=0.0001) for all beard types. In addition, confidence ratings were significantly different between beard1 (moustache) and beard2 (long chin beard) in the AV condition (pMCMC=0.0058) and in the A condition (pMCMC=0.01).

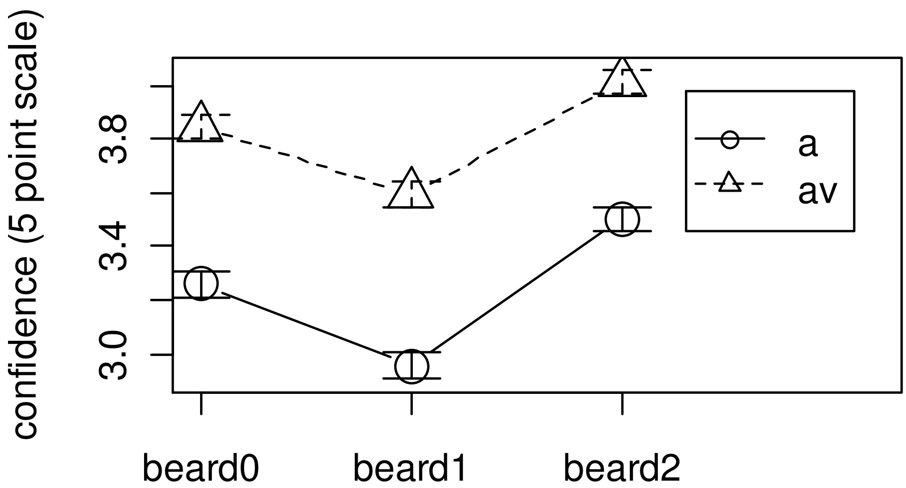

Figure 5: Means and standard errors for confidence ratings; a = audio data (circle markers), av = audio-visual data (triangle markers), 3 beard conditions: beard0=no beard, beard1=moustache, beard2=long chin beard.

So far the comparison between no beard versus moustache shows the trend in the expected direction: the moustache has a reduced intelligibility, longer reaction time, and listeners are less confident to perceive the relevant target word than in the no beard condition. However, the findings for the long chin beard go against our expectations. We found an effect of facial hair in the audio-only condition, where no visual information is available. Two explanations for this finding might be possible: First, since the A and AV conditions were presented randomly and the number of target words was limited to keep the experiment feasible, a strong learning effect could have taken place. Hence, listeners may even become better when they have no visual information available, since they have learnt the relevant target word.

Second, since we glued the beards on the facial skin of the speakers, it may have caused some irritation, so that our speakers produced the relevant sentences in a different way with the long chin beard in place. During the recording session and during the data preparation, we were not aware of such an effect, thus we assume a learning effect.

3.1. LEARNING

To check the potential effects of learning, all the A and AV data were pooled together, but recognition was split by beard condition and occurrence of the target words from the first to the sixth trial (see Figure 6).

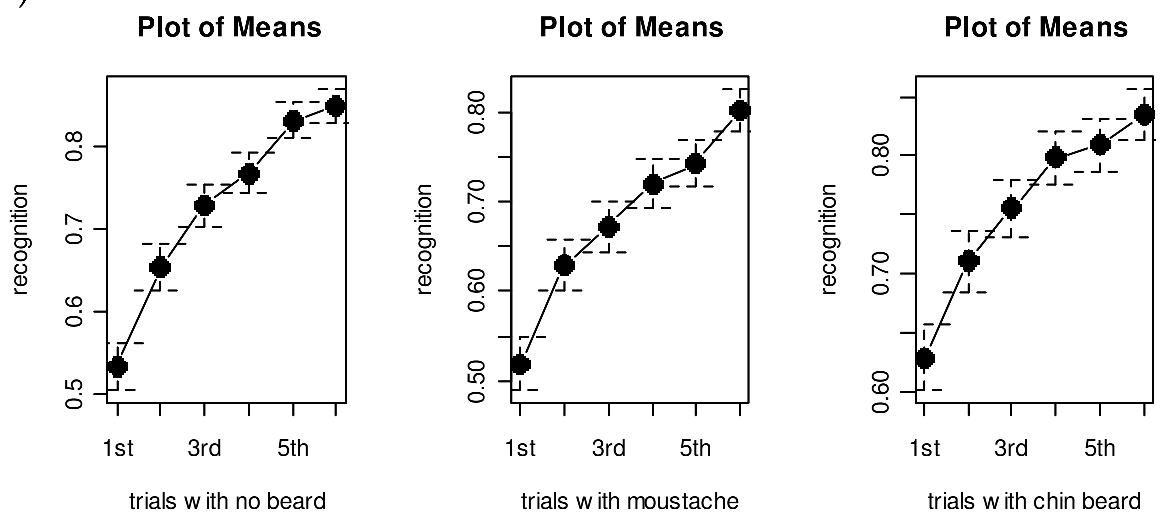

Figure 6: Means and standard errors of recognition rate (A and AV data pooled together) split by beard condition (from left to right) and trial.

Figure 6 displays the results and clearly shows an effect of learning. The recognition rate increases from one trial to the next, independently of the beard condition. The learning effect is greatest from the first trial of a word to the second one. Next we ran a linear regression analysis for the first trial only to test whether differences between the beard conditions are already evident in the first trial.

No statistical differences were found for the AV condition. In the A condition beard0 and beard2 reached borderline significance (pMCMC=0.0498), but beard1 differs from beard2 (with pMCMC=0.0028). This result leads us to conclude that our speakers used different strategies when wearing a long chin beard. We can only speculate about these strategies. One of the reasons might be that for the long chin beard, opening the jaw stretches the surface of the skin, but the artificial beard did not stretch as much as the skin. Such a difference might have caused a different speaking behaviour of our subjects. These results would be in agreement with recent findings on the effects of skin stretching on speech production and perception by Ito and colleagues (Ito, Tiede & Ostry, 2009). Although this finding could be very interesting for the role of somato-sensory information in speech production and perception, it is disruptive for our perception experiment, since it is difficult to derive any reliable conclusions when comparing the long chin beard with the other conditions. Hence, we will only discuss the comparison between no beard and the moustache.

3.2. WORD EFFECTS

Splitting intelligibility by the different words used in this corpus reveals a large impact of the individual word (see Figure 7). Words like *Leben* (life), and *Woche* (week) have a relatively low recognition rate in the audio-only condition and their intelligibility increases substantially (up to about 40%) in the AV condition. An intermediate improvement was found for words like *Abend* (evening) or *Freude* (happiness). All other words, e.g. *Strasse* (street), *Mädchen* (girl), show a lower difference between the A and the AV conditions and in most cases words are already recognized successfully in the A condition (above 80 percent). These differences among the words cannot be explained with respect to the variation in word frequency presented in Table 1. We noticed that in the word group with a high intelligibility in the audio-only condition, voiceless obstruents occur relatively often. Voiceless obstruents, in particular voiceless fricatives, are perceptually very salient (Balise & Diehl, 1994) and might therefore be recognized even under noisy auditory conditions.

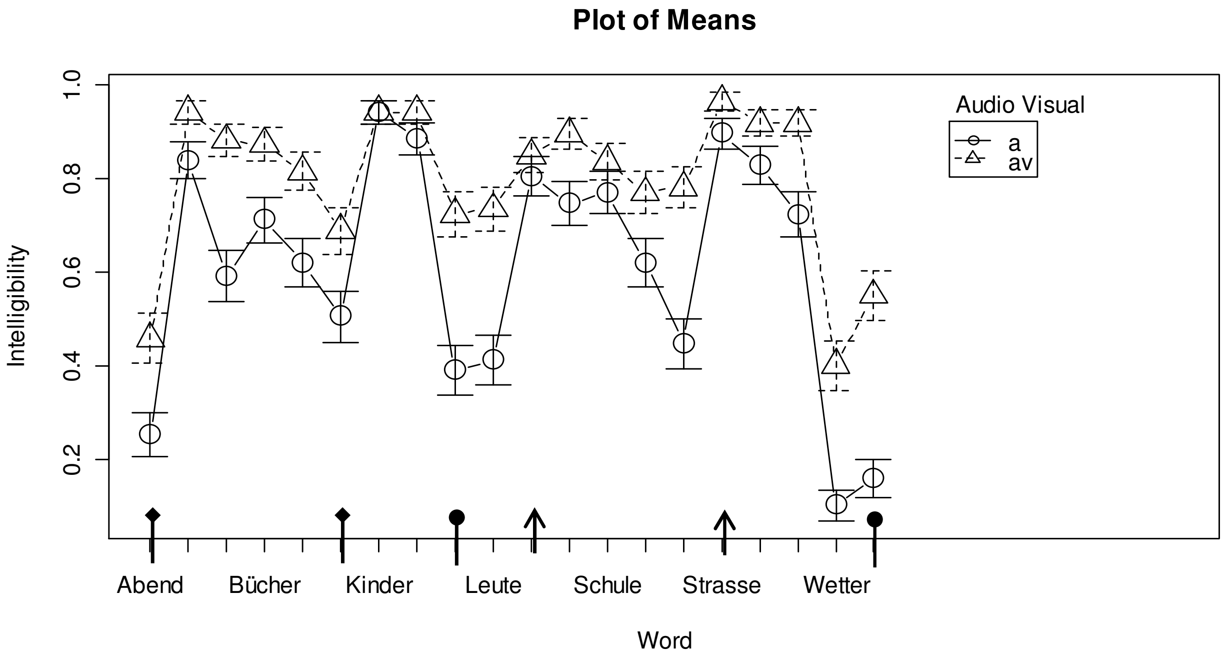

Figure 7: Means and standard errors of intelligibility split by word (x-axis), a=audio, av=audiovisual condition, beard0 and beard1 data are pooled together. Markers on the x-axis correspond to the words mentioned in the text with different degree of improvement: Markers with an arrow are the ones with the least improvement, markers with a dot to the ones with the largest improvement and markers with a square to an intermediate stage

At this point, we present the intelligibility plots for the group of 0 obstruents and the group of 2 obstruents (for 3 obstruents the sample size is rather small) in the no beard and moustache conditions.

In Figure 8 it can be seen that the number of voiceless obstruents in bisyllabic words has an impact on the intelligibility in the two selected cases. If the target word includes no voiceless obstruents, the improvement from A to AV is larger in the no beard condition than in the moustache condition and the no beard AV condition is significantly better than the moustache.

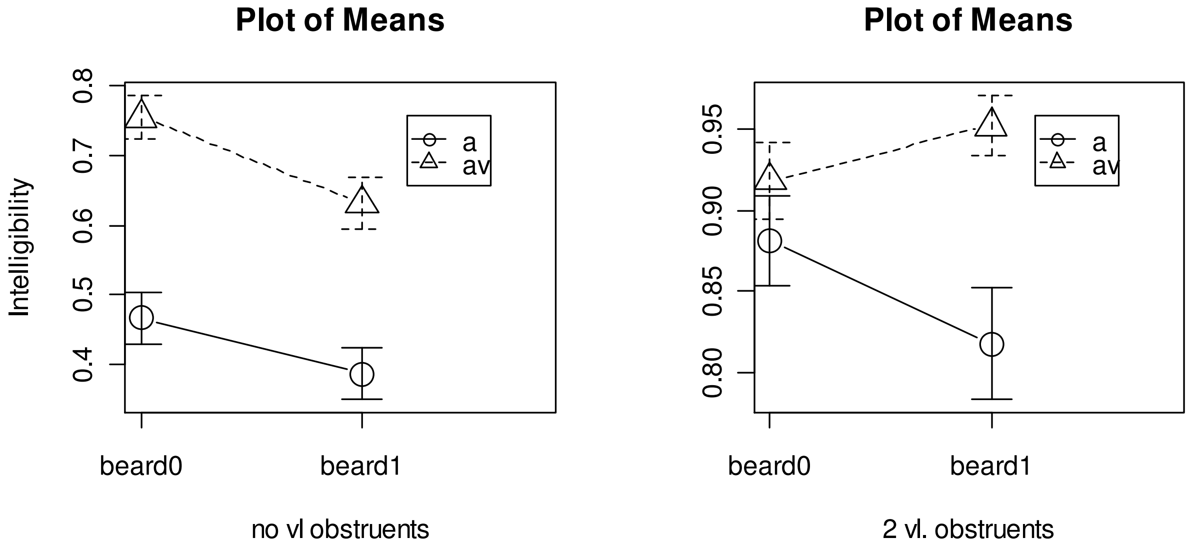

Figure 8: Means and standard errors of intelligibility split by number of obstruents in the target word: left subplot: 0 voiceless obstruents, right subplot: 2 voiceless obstruents, a=audio, av=audiovisual condition, beard0: left column, beard1: right column.

Similar differences are not evident in words with two voiceless obstruents, where intelligibility for no beard is already 90 per cent in the audio-only condition and it is not significantly different from the AV condition. For the moustache (beard1) the results for AV-condition improve in relation to the audio only condition, but within a given condition (A or AV) the findings for no beard and beard1 do not differ.

4 SUMMARY AND CONCLUSION

We have good news for all those who have facial hair longer than 3 mm: There is no need to shave! Although a trend towards reduced intelligibility was found in the beard1 condition (moustache), this trend was not significantly different from the beard0 (no beard) condition. Moreover, the improvement from the audio-only to the audio-visual intelligibility is larger for beard1 (moustache) than beard0 (no beard). We interpret this result with respect to greater attentiveness. Listeners who are presented with an impoverished visual signal pay more attention to this visual information, thereby enhancing intelligibility. The greater attentiveness may also be reflected in the significantly longer reaction time found for beard1. Thus, if you wear a moustache in a noisy auditory environment, please do not speak fast and take a break from time to time, otherwise listeners may not be able to process your speech. Moreover, be aware that people may be attracted by your beard and focus on it (that may be relevant particularly for politicians). Similar to the findings for intelligibility, listeners show a non significant trend towards greater confidence that they have perceived the right word when they see speakers without any facial hair in comparison to when they see speakers wearing a moustache. However, findings are influenced by the number of voiceless obstruents in a given word: the more voiceless obstruents in a word, the easier the recognition of the word even under auditory noise. We interpret this result with respect to the perceptual salience of these sounds.

Based on our data, we were not able to verify whether differences in the shape of the facial hair would affect intelligibility. Such an investigation may be carried out in the future.

APPENDIX

Sächsische Zeitung 27th / 28th of june 2009, front page

"Haare im Gesicht sind Männersache – ob es den Frauen gefällt oder nicht….. Da soll doch einer sagen, Bärte seien nicht gut für die Karriere….Zeige mir was in deinem Gesicht wächst und ich sage dir, wer du bist."

[Facial hair is man's business – whether it suits women or not….How can one dare to say, facial hair would not be good for the career…. Show me your facial hair and I will tell you, who you are.]

Page M5: "Geschichten im Gesicht"

"Bärte sind mehr als nur Haare am Kinn. Sie machen aus Männern Charakterköpfe mit Markenzeichen."

[Facial hair is more than only hair at the chin. It makes striking heads out of men.]

Prince, September 2009, page 130

"Seitdem Brad Pitt, George Clooney und andere Hollywoordstars nicht mehr oben ohne herumlaufen, hat der Schnäuzer die Straßen zurückerobert. Ist der Hype berechtigt?"

[Ever since Brad Pitt, George Clooney and other Hollywood stars do not walk about topless, the moustache is back in the streets. Is the hype justifiable?]

"Im Volksmund wird der Schnäuzer auch "Rotzbremse", "Suppensieb", "Popelfänger", "Schenkelbürste" oder "Pornobalken" genannt. "

[In vernacular speech the moustache is also called "snot bracket", "soup filter", "bogey catcher", "crural brush" or "porn baulk".]

"Träumt eine Frau von einem Mann mit Schnäuzer, hat das laut Esoterikanbieter Traumdeuter.ch folgende Bedeutung: „Ein Mann, den Sie immer für einen Helden hielten, entpuppt sich als armseliger Tropf.""
[If a woman dreams about a man with a moustache, it has the following meaning according to the esoteric offerer Traumdeuter.ch: A man, who you accounted as a hero, turns out to be a poor blighter.]

"Schnurrbärte sind eine Bakterienschleuder. "
[Moustaches are bacteria catapults.]

"Salvador Dali war der festen Überzeugung, mit seinen Schnurrbartspitzen göttliche Botschaften empfangen zu können."
[Salvador Dali had the firm conviction that he would pick up godly messages by means of his moustache ends.]

"Heute brauchen Typen dicke Karren., früher reichte der dickste Schnurrbart: Zu Kaiserzeiten wies die Größe des Soldatenschnäuzers auf den Rang innerhalb der Armee hin. "
[Nowadays, guys need big cars, in former times the largest moustache was sufficient: In Roman Iron Age the size of the soldier's moustache was a sign of its status in the army.]

"Man kann sich mit seinem Schnurrbart auch in Geduld und Selbstdisziplin üben: Der Inder Badamsinh Juwansinh Gurjar ließ seinen Schnäuzer in 25 Jahren auf knapp vier Meter Länge wachsen – Weltrekord."
[One can also exercise patience and self-discipline with a moustache: The Indian Badamsinh Juwansinh Gurjar grew his facial hair for 25 years to just under 4 metres – world record.]

ACKNOWLEDGEMENTS

We would like to thank Jörg Dreyer for the technical support, all participants of this study, Sandra Pape-Hagushikowa for the graphical design of Figure 1, all researchers who provided their photos, and Jean-Luc Schwartz for useful comments on an earlier version of this paper. This work was supported by a grant from the BMBF to the PB1 at ZAS. It is dedicated to all men with facial hair and Dieter Fuchs.

REFERENCES

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R.* Cambridge University Press.

Balise, R. R. & Diehl, R.L. (1994). Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51, 99-110.

Dohen, M., Loevenbruck, H. & Hill, H. (2005). A multi-measurement approach to the identification of the audiovisual facial correlates of contrastive focus in French. *AVSP-2005 British Columbia*, 115-116.

Grant, K.W. and Seitz, P.-F. (2000) The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America* 108(3), 1197-1208.

Haas, W., Knetschke, E., Sperlbaum, M. (eds.) (1984). *Textkorpora 1: Einführungs- und Registerband.* PHONAI, Lautbibliothek der deutschen Sprache 28. Max Niemeyer Verlag, Tübingen.

Hillenbrand, J.M. and Gayvert, R.T. (2005). Open source software for experiment design and control. *Journal of Speech, Language and Hearing Research* 48, 45-60.

Ito T., Tiede M., and Ostry D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academic Sciences* U S A 106(4), 1245–1248.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Retrieved October, 20th 2008 from http://www.r-project.org/

Schwartz, J.L., Abry, C., Boë, L.J. & Cathiard, M. (2002). Phonology in a Theory of Perception-for-Action-Control. In: J. Durand and B. Laks (eds.), *Phonetics, Phonology and Cognition.* Oxford Studies in Theoretical Linguistics. Oxford University Press, New York, 254–280.

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