

Can you hear me now? Reducing the Lombard effect in a driving car using an In-Car Communication system

Rabea Landgraf¹, Johannes Köhler-Kaeß, Christian Lüke²,
Oliver Niebuhr³, Gerhard Schmidt²

¹General Linguistics, Kiel University, Germany

²Digital Signal Processing and System Theorie (DSS), Kiel University, Germany

³Dept. of Design and Communication, IRCA, University of Southern Denmark, Denmark

landgraf@isfas.uni-kiel.de, johans2@web.de, cl@tf.uni-kiel.de, olni@sdu.dk, gus@tf.uni-kiel.de

Abstract

This study aimed to evaluate an In-Car Communication system (ICC), which was developed to improve the communication between passengers inside a driving car. The evaluation was conducted by assessing parameters involved in the Lombard effect, i.e. modifications of speech production in the presence of loud noises. Speech recordings were made inside a stationary car at Kiel University using an acoustic and visual ambiance simulation to imitate real driving situations. In this way, background noises of different driving speeds can be removed from the signals after the recordings, thus allowing undisturbed analyses of acoustic parameters. Recordings were done at noise conditions of silence, 50 km/h and 130 km/h, with and without the use of ICC. 16 subjects participated in the production experiment. Analyses showed that - both with and without ICC - fundamental frequency and intensity increased at higher noise levels, thus confirming the Lombard effect. But, this phenomenon was reduced by the use of ICC, and both pitch and intensity decreased. Furthermore, the reduction of the Lombard effect due to ICC was greater in the back seat compared with the front seat.

Index Terms: Lombard effect, In-Car Communication, ambiance simulation.

1. Introduction

Speech communication inside a driving car is often very difficult and stressful, particularly at higher driving speeds. In this situation, the Lombard effect plays an important role. It represents modifications of speech production and communicative behaviour that are produced at loud noises [1] [7] [18]. Speakers want to remain intelligible to their listeners and maintain control of their own voices [8] [9]. Changes of speech production in loud environments not only concern an increase of intensity, as the French otolaryngologist Étienne Lombard (1869-1920) once discovered [1], but involve many other parameters, such as pitch [2] [10], formant frequencies [2] [7], gestures, and facial expressions [5] [6] [9].

Investigating the Lombard effect in a driving car can be very difficult because measurements of acoustic parameters such as F0 or speech intensity can interfere with loud background noises and thus lead to misinterpretations [13]. Therefore, we developed an acoustic ambiance simulation to investigate speech communication in a car under controlled conditions [13] [14]. During this simulation, speakers sit in a stationary car and hear the noises of different driving speeds from loudspeakers that are placed inside the windows and on the seats.

To analyse speech signals without background noise, the simulated driving noises can be removed from the signals after the recordings [13] [14]. Thus, it is possible to conduct undisturbed acoustic-prosodic as well as spectral analyses of the speech signals without having the dialogue partners wearing headphones.

Previous investigations based on the Lombard effect have shown that our simulation constitutes a realistic driving ambiance [15] [17] [18]. Recently, the acoustic simulation has been further improved and extended by an additional visual simulation. That is, a large screen is placed in front of the car, showing a video of driving situations that have been recorded together with the audio signals [17] [20]. Images of the acoustic and visual simulation are shown in figure 1.



Figure 1: The acoustic and visual ambiance simulation at Kiel University.

To enhance the communication between passengers, so-called In-Car Communication systems (ICC) transmit speech signals inside the driving car from the speaker to the listener(s). Such an ICC System was also developed at Kiel University, whereby the speech signal of the talking person is recorded via microphones inside his/her seat belt and played back by loudspeakers that are placed near the other passengers' heads. [11] [16]

2. Aims

The present study is part of a larger investigation, which aims to evaluate the ICC system described above from a linguistic point of view. That is, it will be examined how the use of ICC influences speech communication and whether it actually improves communication in the presence of loud driving noises. In the current study, this was investigated by measuring parameters involved in the Lombard effect and by answering the following research questions:

- How does speech communication in a car change under Lombard conditions (without ICC)?

- How does the Lombard effect change by using an ICC system?

To address these questions, speech recordings were performed at different driving noises and with or without using the ICC system. These recordings are referred to as the SPID speech corpus (SPontaneous In-car Dialogues) [19] and were elicited using the Map Task paradigm. That is, one dialogue partner takes the role of the instruction giver and his/her interlocutor is the instruction follower. Differences in mean pitch and mean intensity in the speech signals were analysed. The present investigation additionally aimed at searching for differences in speech production between instruction giver and follower. The potential differences relate to the speech task and seating position in the car.

3. Recording conditions of SPID

Recordings were made using the acoustic and visual ambiance simulation. For the elicitation of spontaneous speech, each dialogue pair performed Map Tasks in the presence of three different noise conditions, i.e. *silence* and the driving noises of 50 km/h and 130 km/h. All of these noise conditions were combined both with and without an ICC system, so that $3 \times 2 = 6$ experimental conditions were included. Speech signals without ICC were recorded in the first session. Recordings with ICC followed about three to four months later. This break was to reduce any learning/familiarization effects of the subjects with regard to the speech task.

16 subjects (8 male, 8 female) participated in the production experiment and were divided into 8 same-gender dialogue pairs, thus reducing artefacts of phonetic convergence and controlling gender-related prosodic differences. The subjects in each dialogue pair were acquaintances or friends to ensure that a relaxed and authentic speech communication would be achieved. One of the dialogue partners sat in the front passenger seat and the other sat behind him or her in the back seat so that they could not make eye contact. All participants were native speakers of standard German and between 22 and 31 years old. The mean speaker age was 26.4 in the first recording session without ICC, and 26.6 in the second recording session with ICC.

In addition, after each of the two recording sessions, further recordings were conducted in a *real* driving situation using the same car. These recordings also included the noise conditions of 50 km/h and 130 km/h, and were conducted either with or without ICC, i.e. analogously to the previous recordings in the laboratory. Altogether, 3×2 (laboratory) + 2×2 (driving) = 10 experimental conditions were included in the study, and each of the two recording sessions took between two and four hours. The break between the two sessions also prevented symptoms of fatigue due to long recording durations. The following investigation only focuses on speech signals of the laboratory condition, as the simulation's ecological validity has already been attested [15] [17] [18].

During the speech recordings, the person in the front seat talked towards the windshield and sat with his/her back to the dialogue partner in the back seat [11] [16]. That is, the subjects' acoustic energy radiation was directed to the front of the car, making it more difficult for the speaker in the back to understand his/her interlocutor. The ICC system was adjusted in such a way that speech signals from the front seat were transmitted louder to the loudspeakers in the back seat than speech signals from the back seat were transmitted to the front seat.

To allow for channel separation, each dialogue partner had

his/her own microphones. Their speech signals were recorded both by means of a head-mounted microphone and a stationary microphone placed to the left of each subject. For the purpose of the present analyses, only the signals recorded by the head-mounted microphones were used. Video recordings of the subjects were made during the production experiment to allow further analyses of visual information in the communication situation. The camera was placed at the dashboard of the car. Altogether, we recorded approximately 15.78 hours of speech recordings in the acoustic and visual ambiance simulation. In addition, the dialogues recorded in the laboratory had a duration of about 4.6 to 23.6 minutes. This duration varied according to the experimental condition and the respective dialogue partners.

4. The Map Task scenario

The Map Task scenario was used to elicit spontaneous speech for the SPID corpus [3]. Both dialogue partners received a map, with the instruction giver having a route drawn on it. The other speaker was the instruction follower. The giver had to explain the route to the follower, who had to draw it on his/her own map. Different maps were designed for each of the 10 recording conditions. To allow further analyses of the influences of different maps on the speech signals, 6 maps were only used in the laboratory conditions, and 4 of them only for recordings in the real driving situations. Seats and roles of giver and follower were chosen by the subjects, but never changed within each dialogue pair for the rest of the experiment.

4.1. Maps

The two maps given to a pair of subjects shared many common features, but there were some keywords that only appeared on the giver's map or on the follower's map. In addition, names of streets and persons as well as house numbers were slightly different. Some parts of the maps even had slightly different structures. Very often, the giver had a name on his/her map like "Schulz", whereas the follower had two names, "Schulz" and "Scholz", so that a controversial discussion was fostered. Maps and noise conditions were randomised for both recordings in the laboratory and the driving car.

4.2. Instructions

Prior to the production experiment, subjects were told that there would be two recording sessions. They also knew that in every session, recordings would be made at different noise conditions in the acoustic and visual ambiance simulation (which was briefly explained) and during real driving situations in the same car afterwards. In preparation for the Map Task scenario, a short text was given to the dialogue partners before each of the two sessions. The text was slightly different for instruction givers and followers and described a short scenario for the speech task. That is, the giver was invited to a wedding of his/her cousin and promised to pick up several other people on the way to the event. For this purpose, his/her cousin drew a self-made town map. But now, he/she cannot go to the wedding and asks the follower to go and to pick up the other guests. In order to do this, the giver has to dictate the route on his/her map (which they knew was wrong in several points) to the follower. The participants were informed that their maps were different. When the participants finished reading the introductory text, it was again orally summarised by the experimenter and upcoming questions were answered. After the first recording session, the subjects were told that an ICC system would be employed

during the next session, and the experimenter shortly explained how this system works.

5. Transliterations and annotations

Detailed orthographic transliterations were made for 5 of 8 dialogue pairs. On the basis of these transliterations, the free online software MAUS (Munich AUtomatic Segmentation system) [4] [12] was used to obtain segmentations and orthographic labelling of the speech signals. The output was returned in the form of so-called TextGrid files for Praat [21].

6. Results

For statistical examination of the research questions described above (see section 2), we conducted a linear least squares regression using the programming language and environment R [22]. The mean pitch and mean intensity of every speech signal were the investigated parameters. The influences of the two factors *ICC* and *noise* on the measurements were tested, with *ICC* having two different levels (*without ICC*, *with ICC*), and *noise* having three (*silence*, *50 km/h*, *130 km/h*). Furthermore, we investigated interactions between these two factors and conducted regression analyses to examine the effects of the factors *seat* (*front*, *rear*) and *role* (*giver*, *follower*) on the measurements.

We applied a within-subject normalisation procedure in order to prepare the data for the linear regression model (LRM). For every subject s , a mean value \bar{x}_s was calculated over all experimental conditions i . The mean was then subtracted from the respective individual values $x_{s,i}$, resulting in the mean-adjusted values

$$x'_{s,i} = x_{s,i} - \bar{x}_s. \quad (1)$$

Based on the mean-adjusted values $x'_{s,i}$, the deviations from the speaker's own mean values were used to compare the holistic behaviour of the speakers in all experimental conditions. Therefore, we are analysing $\Delta F0$ and Δ intensity.

6.1. Mean pitch

Analyses of the mean pitch were done using Praat, with individual pitch ranges for men (50 Hz - 300 Hz) and women (100 Hz - 500 Hz). Since the dialogue pairs were separated by gender, and as we applied a within-subject normalisation procedure, we did not need to convert pitch to semitones. Descriptive analyses of the measurements in every speech signal revealed that the Lombard effect appeared both with and without the use of ICC. That is, the mean pitch increased for almost every speaker from silence to 50 km/h and from 50 km/h to 130 km/h. These effects are illustrated in figure 2, where changes between the measurements seem to be quite linear.

Regarding the outputs of the LRM for the $\Delta F0$ measurements, both *ICC* and *noise* had highly significant influences on the mean pitch. For *ICC* the p-value was <0.0001 . For *noise* the maximum p-value was 0.0001. There was also an interaction between the two factors with $p=0.0009$. Furthermore, the factor *seat* had a nearly significant influence with $p=0.0836$. In addition, there was a marginally significant interaction between the factors *ICC* and *seat* ($p=0.0152$). This means that individuals appear to adapt differently to the ICC-condition, according to their seating location. The factor *role* had no significant influence on the mean pitch.

The estimated values resulting from the LRM showed absolute changes in overall mean pitch for different factor combi-

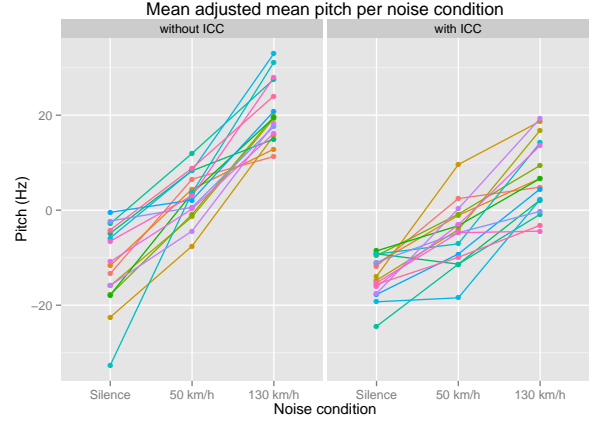


Figure 2: Normalised mean pitch ($\Delta F0$) per noise condition, with and without the use of ICC. Different lines represent different subjects.

nations. To ensure better comparability of these values, a linear transformation was calculated, where the value of *without ICC*, *silence*, *rear* was subtracted from every factor combination. In this way, *without ICC*, *silence*, *rear* became zero and were taken as a baseline. The transformed estimated values are shown in table 1. They confirm that both with and without ICC, there was a Lombard effect at higher noise levels with regard to the mean pitch. In addition, the estimated values reveal that with ICC, the mean pitch was lower than without ICC in almost all noise conditions (with the exception of *silence*, *front*). It can further be seen that the Lombard effect is less strong when an ICC system is used.

Table 1: Estimated values for mean pitch for each factor-combination of “ICC”, “noise”, and “seat”. “Without ICC, *silence*, *rear*” is taken as a baseline.

	rear	silence	50 km/h	130 km/h
without ICC		0 Hz	14.548 Hz	32.192 Hz
with ICC		-5.779 Hz	3.496 Hz	15.202 Hz
	front	silence	50 km/h	130 km/h
without ICC		-3.282 Hz	11.266 Hz	28.910 Hz
with ICC		-2.497 Hz	6.778 Hz	18.484 Hz

Table 1 shows that subjects adapted differently to the situation with an ICC system depending on the seating position of each speaker. That is, in the silence condition, measurements in the back seat decreased with ICC, whereas they slightly increased in the front seat when the ICC system was activated. In addition, the differences in the mean pitch values between *without ICC* and *with ICC* increased at higher driving noises. In the front seat, however, this effect was not as strong as in the back seat.

6.2. Mean intensity

Measurements of mean intensity were also done in Praat. Descriptive analyses yielded results similar to those of mean pitch. There was an increase in loudness with increasing background noises with and without ICC, as is shown in figure 3.

Regression analyses for the Δ intensity measurements

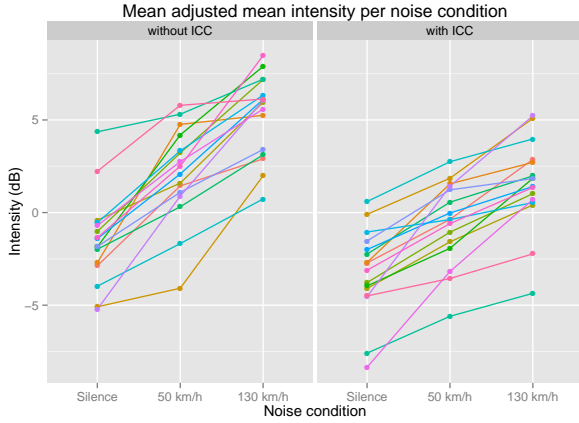


Figure 3: Normalised mean intensity (Δ intensity) per noise condition, with and without the use of ICC. Different lines represent different subjects.

show that *ICC* (with $p < 0.0001$) and *noise* (with a maximum of $p = 0.0055$) both had significant influences. Again, the LRM showed an interaction between the two factors. With regard to the direction of influence, this interaction is comparable to that of the mean pitch analyses. However, the interaction between *ICC* and *noise* had a p -value of 0.0513. Therefore, only a marginally significant influence of the interaction can be assumed. In addition, *seat* had a highly significant influence on mean intensity ($p = 0.0002$). Furthermore, there was a significant interaction between the factors *ICC* and *seat* ($p = 0.0001$). This confirms the results for mean pitch that individuals adapt differently to the ICC system according to their seating position. Again, the analyses showed that there was no impact of the factor *role* on mean intensity.

The estimated values in table 2 (*without ICC*, *silence*, *rear* is again taken as a baseline) confirm that there was a Lombard effect both with and without ICC. Again the values are lower with activated ICC in almost every noise condition, except for *silence*, *front*. Furthermore, the Lombard effect was less strong with ICC.

Table 2: Estimated values for mean intensity for each factor-combination of “*ICC*”, “*noise*”, and “*seat*”. “*Without ICC*, *silence*, *rear*” is taken as a baseline.

rear	silence	50 km/h	130 km/h
without ICC	0 dB	3.617 dB	6.793 dB
with ICC	-4.042 dB	-1.371 dB	0.710 dB
front	silence	50 km/h	130 km/h
without ICC	-2.367 dB	1.250 dB	4.426 dB
with ICC	-1.737 dB	0.934 dB	3.015 dB

Similar to mean pitch, table 2 shows that the dialogue partners adapted differently to the situation with ICC, depending on their seating position. Again, when ICC was activated, measurements in the silence condition decreased in the back seat and increased in the front seat. The differences in the mean intensity values between *without ICC* and *with ICC* increased at higher driving noises. The only exception was that, in the front seat, the difference in mean intensity slightly decreased from *silence* to 50 km/h.

7. Discussion

Analyses of mean pitch and mean intensity in the speech signals yielded parallel results for both parameters. The Lombard effect was found for speech communication inside a driving car. Subjects spoke louder and with a higher fundamental frequency when the driving noises in the background increased from silence to 50 km/h, and further from 50 km/h to 130 km/h. The Lombard effect occurred with and without the use of an ICC system. But, with ICC, loudness and pitch decreased in all noise conditions except for the silence condition in the front seat. In addition, the magnitude of the Lombard effect was reduced with ICC. Regarding mean pitch, it is striking that the Lombard effect became stronger at higher noise levels, as differences between the pitch levels were greater for 50 km/h and 130 km/h than for silence and 50 km/h.

Further analyses of the seating position showed that there was a difference in the adaptation of speech production to the ICC-condition between front seat and back seat. This was true for both pitch and intensity. The differences between the measurements *without ICC* and *with ICC* were higher for the back seat than the front seat in every noise condition. Thus, the ICC was more effective in the back seat. This result could be caused by the fact that the ICC system was adjusted to transmit speech signals from the front seat to the back more strongly than in the other direction. Furthermore, the values decreased in the back seat with ICC being switched on in the silence condition. One possible explanation could be the direction of sound energy radiation (see section 3), which was unfavourable for the person in the back seat. It could have caused an increased vocal effort that was then compensated by the activation of the ICC system. On the contrary, the measurements in the silence condition in the front seat were almost equal or just slightly increased from *without ICC* to *with ICC*. In addition, it is striking that, for mean intensity in the condition of *rear with ICC 130 km/h* in table 2, the estimated value (0.710) approached the baseline of *rear without ICC silence*. That is, with ICC, the Lombard effect could be reduced for the back seat in such a way that subjects were almost as loud as they would be in the silent condition. Additional analyses of the different speech roles *giver* and *follower* revealed that they had no influence on the measurements.

8. Conclusions

Our study shows that the use of an ICC system influences speech communication in a driving car. We conclude from the findings of this study that it takes less effort for dialogue partners to communicate at higher driving noises when using ICC because the Lombard effect is reduced with regard to pitch and loudness. Subsequent to this study, we will analyse additional parameters in the speech signals of the SPID corpus to further support the hypotheses described above and to supplement the findings of this investigation. We will also check if conversations between dialogue partners are actually more comfortable and less stressful with ICC. Relaxed communication inside the driving car may also result in increased driving safety. In addition, we will investigate why the use of ICC changes the magnitude of the Lombard effect. One possible reason is that subjects perceive their own voices from the loudspeakers and thus gain more control of them. Another explanation could be that the speaker’s knowledge about his/her increased intelligibility for the dialogue partner reduces the necessity of a strong Lombard effect.

9. References

- [1] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *Journal of Speech and Hearing Research*, vol. 14, pp. 677–709, 1971.
- [2] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *Journal of the Acoustical Society of America*, vol. 84, pp. 917–928, 1988.
- [3] H. S. Thompson, A. Anderson, E. G. Bard, G. Doherty-Sneddon, A. Newlands, and C. F. Sotillo, "The HCRC Map Task Corpus: Natural dialogue for speech recognition," *Proc. Human Language Technology Workshop*, pp. 25–30, 1993.
- [4] F. Schiel, "Automatic Phonetic Transcription of Non-Prompted Speech," *In Proc. of the ICPhS*, pp. 607–610, 1999.
- [5] C. Davis, J. Kim, and A. Sironic, "In-quiet and Lombard audiovisual speech," *7th Proc. International Seminar on Speech Production*, 2006.
- [6] E. Vatikiotis-Bateson, A. V. Barbosa, C. Y. Chow, M. Oberg, J. Tan, and H. C. Yehia, "Audiovisual Lombard speech: Reconciling production and perception," *Proc. International Conference of Auditory-Visual Speech Processing*, pp. 45–50, 2007.
- [7] P. Lau, "The Lombard Effect as a communicative phenomenon," *UC Berkeley Phonology Lab Annual Report*, no. 39, pp. 1–9, 2008.
- [8] R. Patel and K. W. Schell, "The influence of linguistic content on the Lombard Effect," *Journal of Speech, Language, and Hearing Research*, vol. 51, pp. 209–220, 2008.
- [9] M. Garnier, N. Henrich, and D. Dubois, "Influence of sound immersion and communicative interaction on the Lombard Effect," *Journal of Speech, Language, and Hearing Research*, vol. 53, pp. 588–608, 2010.
- [10] L. Folk and F. Schiel, "The Lombard Effect in spontaneous dialog speech," *Proc. 12th International Interspeech Conference*, pp. 2701–2704, 2011.
- [11] C. Lüke, H. Özer, G. Schmidt, A. Theiß, and J. Withopf, "Signal Processing for In-Car Communication Systems," *DSP Workshop for In-Vehicle Systems*, pp. 1–8, 2011.
- [12] T. Kislser, F. Schiel, and H. Sloetjes, "Signal processing via web services: the use case WebMAUS," *In Proceedings Digital Humanities 2012*, pp. 30–34, 2012.
- [13] T. John, O. Niebuhr, G. Schmidt, and A. Theiß, "Phonetic analysis vs. dirty signals: Fixing the paradox," *Proc. der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV)*, pp. 1–8, 2013.
- [14] C. Lüke, A. Theiß, G. Schmidt, O. Niebuhr, and T. John, "Creation of a Lombard speech database using an acoustic ambiance simulation with loudspeakers," *Proc. of the 6th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems*, pp. 1–8, 2013.
- [15] T. John, R. Landgraf, C. Lüke, S. Rohde, G. Schmidt, A. Theiß, and J. Withopf, "Über die Verbesserung der Sprachkommunikation in geräuschbehafteten Umgebungen," *In: O. Niebuhr (ed), Formen des Nicht-Verstehens*, pp. 185–203, Frankfurt: Peter Lang, 2014.
- [16] A. Theiß, G. Schmidt, J. Withopf, C. Lüke, "Instrumental Evaluation of In-Car Communication Systems," *Proc. ITG Fachtagung Sprachkommunikation*, pp. 1–4, 2014.
- [17] R. Landgraf, O. Niebuhr, G. Schmidt, T. John, C. Lüke, and A. Theiß, "Von der Straße ins Labor: Die Modifikation der Sprachproduktion bei lauten Fahrgeräuschen," *41. Jahrestagung für Akustik (DAGA)*, pp. 414–417, 2015.
- [18] R. Landgraf, "Simulating complex speech-production environments," *In: O. Niebuhr and R. Skarantz (eds), Tackling the Complexity of Speech*, pp. 97–110, Prague: Eponcha, 2015.
- [19] O. Niebuhr, B. Peters, R. Landgraf, and G. Schmidt, "The Kiel Corpora of "Speech & Emotion" - A Summary," *41. Jahrestagung für Akustik (DAGA)*, pp. 1011–1014, 2015.
- [20] A. Warhadpande, C. Lüke, A. Theiß, and G. Schmidt, "Improvement by Adding Video Feature in an Acoustic Ambiance Simulation for Automobiles," *Proc. of the 7th Biennial Workshop on DSP for In-Vehicle Systems and Safety*, 2015. (in print)
- [21] <http://www.fon.hum.uva.nl/praat/>
- [22] <https://www.r-project.org/>