# Multitask Learning for Grapheme-to-Phoneme Conversion of Anglicisms in German Speech Recognition

Julia Maria Pritzen[12], Michael Gref[1], Dietlind Zühlke[2], Christoph Andreas Schmidt[1]

[1] Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Germany
[2] TH Köln - University of Applied Sciences, Cologne, Germany

# Agenda

1. Problem Definition
2. Related Work
3. Proposed Approach
4. Experimental Setup
5. Evaluation and Results
6. Conclusion

Fraunhofer IAIS

Chapter 01

# Problem Definition

Fraunhofer IAIS

# Problem Definition

## Anglicisms in German Speech

> An Anglicism is a word or idiom that is recognizably English in its form (spelling, pronunciation, morphology, or at least one of the three), but is accepted as an item in the vocabulary of the receptor language.
>
> **(Görlach, 1994, p.224)**

- 4.5 % of words in German speech are Anglicisms (Hunt, 2019)
- Use of Anglicisms is steadily increasing (Burmasova, 2010)

# Problem Definition

## Anglicisms in German Speech

- Anglicisms (mostly) do not follow German pronunciation rules

- German G2P model uses German "pronunciation rules" to generate pronunciations
  - Generating pronunciations for words of English heritage with a German system can lead to wrong pronunciations
  - Potentially wrong entries in the pronunciation dictionary for Anglicisms

- Why not just use the English source pronunciations?
  - Can contain phonemes that are not present in the DE phoneme set
  - Anglicisms are inflected using the target languages grammar
    - Example: downgeloadet

→ **We need a different solution!**

Chapter 02

# Related Work

# Related Work

## Multitask Learning, Seq2Seq LSTM and Multilingual G2P Approaches

- **Caruana (1993): Multitask Learning: A Knowledge-Based Source of Inductive Bias**
  - Approach of modeling the human concept of inductive transfer to a machine learning model
  - MTL model will learn multiple tasks in parallel
- **Yao and Zweig (2015): Sequence-to-Sequence Neural Net Models for Grapheme-to-Phoneme Conversion**
  - Based on Sutskever et al. (2014) → Seq2Seq LSTM model for machine translation
  - DNN approach utilizing LSTM cells in an encoder-decoder architecture
- **Milde et al. (2017): Multitask Sequence-to-Sequence Models for Grapheme-to-Phoneme Conversion**
  - Multilingual Seq2Seq G2P models utilizing multitask learning to train simultaneously on a German and English G2P task
  - MTL models did not outperform the baseline Sequitur G2P model, not even for English loan words
- **Sokolov et al. (2019): Neural Machine Translation for Multilingual Grapheme-to-Phoneme Conversion**
  - Multilingual Seq2Seq G2P model using transfer learning to improve performance on foreign words from different languages
  - Increased PER and WER for German with multilingual model

Chapter 03

# Proposed Approach

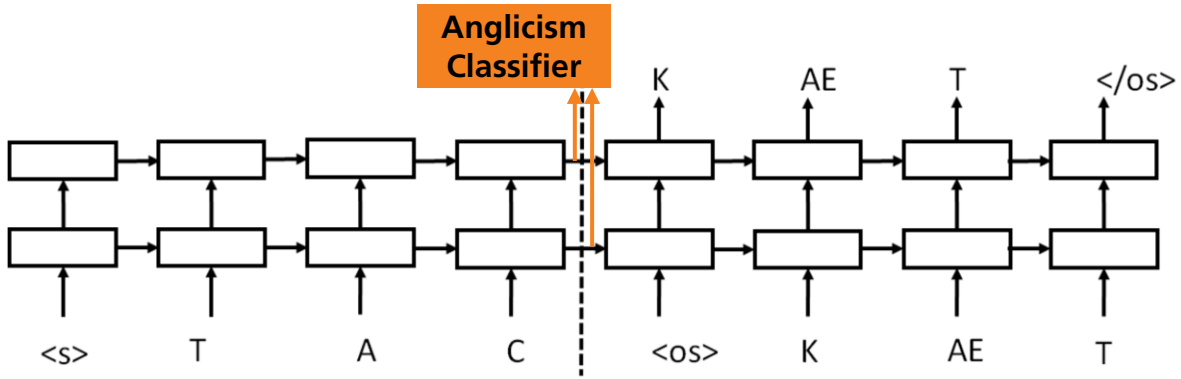
 

Fraunhofer IAIS

# Proposed Approach

## Multitask Seq2Seq G2P Model

- DE Seq2Seq G2P model based on Yao and Zweig (2015) with an additional classification task




- Model will learn how to distinguish Anglicisms based on encoder output
- Anglicism pronunciations will be generated with different "pronunciation rules" than native German words

Chapter 04

# Experimental Setup

Fraunhofer IAIS

# Experimental Setup

## Datasets

- **Anglicism list** derived from Wiktionary and VDS Anglizismenindex[1] (18,967 words)
  - Wiktionary provides lists of "Anglizismen[2]" and "Scheinanglizismen[3]" (pseudo-anglicisms)
  - Additionally tried to find inflections of Anglicisms
- **German pronunciation dictionary** PHONOLEX Core (65,427 entries)
- **Pronunciations from Wiktionary** based on Anglicism list (9,802 pronunciations)
- **Anglicism ASR test set** including segments with Anglicism usage (1,3 h)

| | Person | Wortform |
|---|---|---|
| **Präsens** | ich | streame |
| | du | streamst |
| | er, sie, es | streamt |
| **Präteritum** | ich | streamte |
| **Konjunktiv II** | ich | streamte |
| **Imperativ** | Singular | stream!<br>streame! |
| | Plural | streamt! |
| **Perfekt** | **Partizip II** | **Hilfsverb** |
| | gestreamt | haben |
| *Alle weiteren Formen:* Flexion:streamen | | |

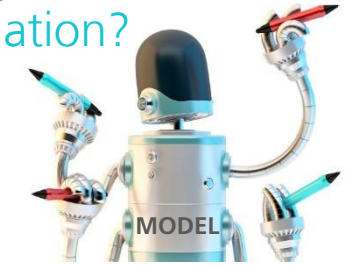
Inflection box on a Wiktionary page (https://de.wiktionary.org/wiki/streamen)

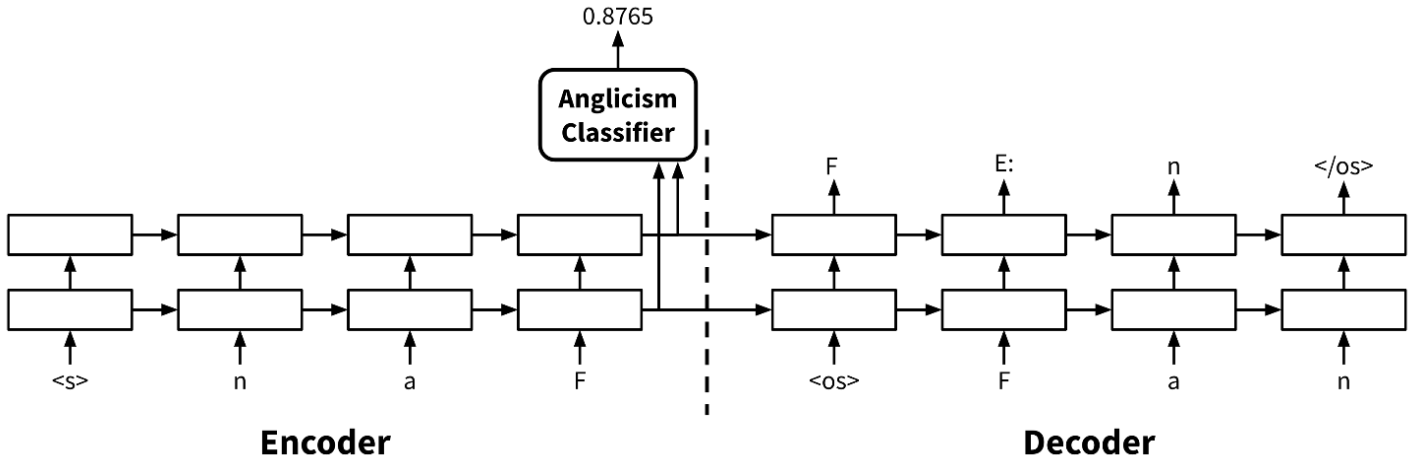
1. https://vds-ev.de/arbeitsgruppen/deutsch-in-der-oeffentlichkeit/ag-anglizismenindex
2. https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Anglizismen
3. https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Anglizismen/Scheinanglizismen

# Experimental Setup

## Implementation

Anglicism or not?
Pronunciation?




- We rebuilt the encoder-decoder LSTM model from Yao and Zweig (2015)
- Added a binary classification task after the encoder step to "detect" Anglicisms
- Losses were combined to optimize on both tasks

→ Resulting model generates a phoneme sequence and classifies whether the input sequence is considered an Anglicism

Chapter 05

# Evaluation and Results

Fraunhofer IAIS

# Evaluation and Results

## MTL Evaluation

| | | | | G2P Task | | Anglicism Classification Task | | | |
|---|---|---|---|---|---|---|---|---|---|
| **G2P Model** | **Data Source & Specifics** | **Epochs** | **Iter. / Epoch** | **PER** | **WER** | **Accu.** | **Prec.** | **Recall** | **F1** |
| $MTL_{Base}$ | PHONOLEX Core | 7 | 2498 | *5.68* | *24.43* | *98.03* | *0.00* | *0.00* | *0.00* |
| $MTL_{Wiki}$ | $MTL_{Base}$ + Wiktionary | 6 | 2845 | 8.63 | 30.89 | 91.24 | 80.69 | 54.26 | 64.89 |
| $MTL_{WL}$ | $MTL_{Wiki}$ + weighed losses ($\alpha = 0.7$) | 7 | 2845 | **7.87** | **28.03** | **92.42** | 86.71 | 58.14 | 69.61 |
| $MTL_{DS}$ | $MTL_{Wiki}$ + downsampled data | 16 | 806 | 11.21 | 39.63 | 88.66 | **90.30** | **86.63** | **88.43** |

- **$MTL_{Base}$**: Only negative classifications because of the class imbalance in PHONOLEX Core (~2 % Anglicism ratio)
- **$MTL_{Wiki}$**: Adding in Wiktionary Anglicism pronunciations to raise Anglicism ratio to ~16 %
- **$MTL_{WL}$**: Altered loss summation by including an additional $\alpha$ parameter ($\alpha = 0.7$) to weight the tasks accordingly:

$$Total\ Loss = \alpha * Decoder\ Loss + (1 - \alpha) * Classifier\ Loss$$

- **$MTL_{DS}$**: Downsampled training data (50 % Anglicism ratio) with equal loss summation (no $\alpha$ parameter)

# Evaluation and Results

## ASR Evaluation

- Chose models $MTL_{WL}$ and $MTL_{DS}$ to test the results in an ASR model
- Generated Anglicism pronunciations based on the Anglicism list (18,967 entries)
- Added resulting Anglicism pronunciations to a baseline ASR models pronunciation dictionary
- Created two control ASR models by generating Anglicism pronunciations with a DE Sequitur and a DE Seq2Seq G2P model

# Evaluation and Results

## ASR Evaluation

- ASR test sets:
  - Anglicisms 2020 → German test set with heavy Anglicism usage
  - German Broadcast 2020
  - Challenging Broadcast 2018

  } Typical German broadcast test sets (control groups)

- Measured WER to determine overall performance of the added pronunciations

- Additionally measured Anglicism error rate (AER) for test set "Anglicisms 2020"
  - Every Anglicism in the test set was flagged as an entity
  - AER represents ratio of wrongly recognized Anglicisms

Levenshtein Distance

WER

Julia ist cool
Julia isst Kuh

→ **2 word errors**

Anglicism not recognized

AER

Julia ist **cool**
Julia isst **Kuh**

→ **1 Anglicism error**

# Evaluation and Results

## ASR Results

| | Anglicisms 2020 | | | | |
|---|---|---|---|---|---|
| **Model** | **WER (%)** | **AER (%)** | **Recognized Anglicisms** | **German Broadcast 2020 (%)** | **Challenging Broadcast 2018 (%)** |
| Baseline (Gref et al., 2019) | 15.80 | 39.50 | 824 | **6.56** | **10.84** |
| Sequitur | 15.76 | 39.35 | 826 | 6.56 | 10.82 |
| Seq2Seq | 15.75 | 39.28 | 827 | 6.56 | 10.91 |
| $MTL_{WL}$ | **15.65** | 38.33 | 840 | 6.57 | 10.86 |
| $MTL_{DS}$ | 15.67 | 38.40 | 839 | 6.60 | 10.90 |
| Wav2Vec2 | 15.69 | 42.07 | 789 | 9.34 | 9.48 |

- MTL models outperformed the non-MTL models → $MTL_{WL}$ showed the best WER
- Additional Anglicism pronunciations did not significantly impact the performance of German control test sets
- Wav2Vec2 showed highest AER (42.07 %) which even exceeded the baseline model

# Evaluation and Results

## Real World Example

- Correct pronunciation for "Views" (→ /v j u: s/) eliminated a word error
- "Views" (also singular "View") was not included in the training / validation data

| Segment 8 in "Rezo - Die Zerstörung der Presse" from test set "Anglicisms 2020" | |
|---|---|
| Reference | […] haben teilweise Millionen von Views |
| Baseline | […] haben teilweise Millionen von Virus |
| Sequitur | […] haben teilweise Millionen von Virus |
| Seq2Seq | […] haben teilweise Millionen von Virus |
| $MTL_{WL}$ | […] haben teilweise Millionen von Views |
| $MTL_{DS}$ | […] haben teilweise Millionen von Views |

| Views | |
|---|---|
| **Model** | **Pronunciation** |
| Baseline | v j E f s |
| Sequitur | v j E f s |
| Seq2Seq | f i: v s |
| $MTL_{WL}$ | v j u: s |
| $MTL_{DS}$ | v j u: s |

Chapter 06

# Conclusion

# Conclusion

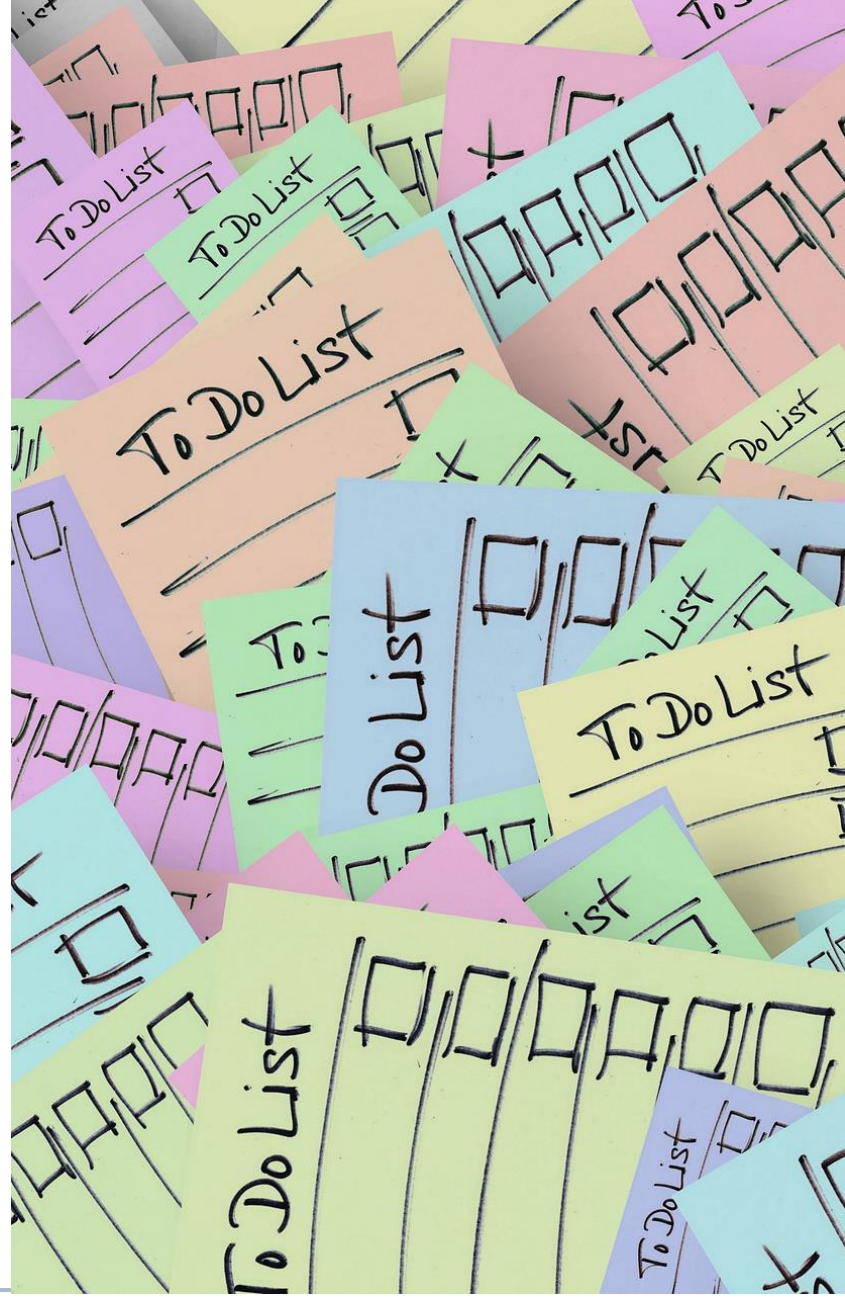
## Summary

- Improved Anglicism recognition results by generating and adding Anglicism pronunciations to the ASR model's pronunciation dictionary
- WER $-1\%$ relative and AER $-3\%$ relative
- No performance loss in control test sets, which are typical use cases in the DE broadcast domain
- Only uses phonemes of the German phoneme set
- Can also be used for other target languages (e.g. Spanish)

# Conclusion

## Future Work

- Experiment with the tuning criteria, e.g. focusing more on the classification results to better deal with the class imbalance
- Assess different architectures to improve the shared knowledge of tasks in the model
- Look more into end-to-end models



Fraunhofer IAIS

# Thank you for your attention!

## Contact information

**Julia Pritzen**
Research Engineer / NetMedia
julia.pritzen@iais.fraunhofer.de

Fraunhofer-Institut für intelligente
Analyse- und Informationssysteme IAIS
Schloss Birlinghoven 1
53757 Sankt Augustin

**www.iais.fraunhofer.de**

# Sources

## Images

- Agenda: https://pixabay.com/photos/a-book-agenda-table-notes-notebook-3043275
- "Denglisch" comic: https://static.dw.com/image/50102546_403.png
- Dictionary icon: https://www.iconpacks.net/free-icon/pink-book-4999.html
- Screenshot Wiktionary: https://de.wiktionary.org/wiki/streamen
- Multitasking robot: https://www.slon.pics/wp-content/uploads/2018/07/Four-arm-robot-with-pencils.-Multitasking-concept-20036.jpg
- ToDo lists: https://pixabay.com/illustrations/lists-to-do-paper-checkbox-tasks-6131213/

# Sources

## Literature


- Baevski, A., Zhou, H., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations.
- Bisani, M. and Ney, H. (2008). Joint-sequence models for grapheme-to-phoneme conversion. Speech Communication, 50(5):434–451.
- Burmasova, S. (2010). Empirische Untersuchung der Anglizismen im Deutschen am Material der Zeitung Die WELT (Jahrgänge 1994 und 2004).
- Caruana, R. (1993). Multitask Learning: A Knowledge-Based Source of Inductive Bias. In Proceedings of the Tenth International Conference on Machine Learning, pages 41–48. Morgan Kaufmann.
- Caruana, R. (1997). Multitask Learning. Ph.D. thesis, Carnegie Mellon University.
- Coats, S. (2019). Lexicon geupdated: New German anglicisms in a social media corpus. European Journal of Applied Linguistics, 7:255 – 280.
- Conneau, A., Baevski, A., Collobert, R., Mohamed, A., and Auli, M. (2020). Unsupervised Cross-lingual Representation Learning for Speech Recognition.
- Gref, M., Schmidt, C., Behnke, S., and K¨ohler, J. (2019). Two-Staged Acoustic Modeling Adaption for Robust Speech Recognition by the Example of German Oral History Interviews. In IEEE International Conference on Multimedia and Expo, ICME 2019, Shanghai, China, July 8-12, 2019, pages 796–801. IEEE.
- Grosman, J. (2021). Wav2Vec2-Large-XLSR-53-German. https://huggingface.co/jonatasgrosman/wav2vec2-largexlsr-53-german.
- Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8):1735–1780, November.
- Hunt, J.W. (2019). Anglicisms in German: tsunami or trickle? In Amei Koll-Stobbe, editor, Informalization and Hybridization of Speech Practices: Polylingual Meaning-Making across Domains, Genres, and Media, pages 25–58. Peter Lang, Bern, Schweiz.
- Max Planck Institute for Psycholinguistics, The Language Archive. (2020). ELAN (Version 5.9) and Simple-ELAN (Version 1.3) [Computer software]. https://archive.mpi.nl/tla/elan

# Sources

## Literature


- Milde, B., Schmidt, C., and K¨ohler, J. (2017). Multitask Sequence-to-Sequence Models for Graphemeto-Phoneme Conversion. In INTERSPEECH 2017, pages 2536–2540.
- Sokolov, A., Rohlin, T., and Rastrow, A. (2019). Neural Machine Translation for Multilingual Graphemeto-Phoneme Conversion. In Interspeech 2019, pages 2065–2069.
- Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, pages 3104—3112, Cambridge, MA, USA. MIT Press.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38–45, Online, October. Association for Computational Linguistics.
- Yao, K. and Zweig, G. (2015). Sequence-to-Sequence Neural Net Models for Grapheme-to-Phoneme Conversion. In INTERSPEECH 2015, pages 3330–3334. 9.

# Sources

## Language Resources


- Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., and Weber, G. (2020). Common voice: A massively-multilingual speech corpus. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 4218–4222, Marseille, France, May. European Language Resources Association (ELRA).
- Elfers, A. (2020). Der Anglizismen-Index 2020 - Deutsch statt Denglisch. https://vds-ev.de/denglisch-undanglizismen/anglizismenindex/aganglizismenindex/
- Schiel, F. (1998). Speech and speech releated resources at BAS. In Proceedings of the First International Conference on Language Resources and Evaluation, pages 343–350, Granada, Spain., May. European Language Resources Association (ELRA).
- Stadtschnitzer, M., Schwenninger, J., Stein, D., and Koehler, J. (2014). Exploiting the large-scale German broadcast corpus to boost the Fraunhofer IAIS Speech Recognition System. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, pages 3887–3890, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Wiktionary. (2019). Verzeichnis: Deutsch/Anglizismen. https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Anglizismen
- Wiktionary. (2020). Verzeichnis:Deutsch/Anglizismen/Scheinanglizismen. https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Anglizismen/Scheinanglizismen