

# Enriched communication across the lifespan

## *Comunicación enriquecida a lo largo de la vida*

Martin Cooke<sup>1</sup>, Simon King<sup>2</sup>, Valerie Hazan<sup>3</sup>,  
Yannis Stylianou<sup>4</sup>, Esther Janse<sup>5</sup>, Deniz Baskent<sup>6</sup>,  
Volker Hohmann<sup>7</sup>, Axel Winneke<sup>8</sup>, Inma Hernaez<sup>9</sup>

<sup>1</sup>Ikerbasque, Basque Science Foundation, Spain

<sup>2</sup>University of Edinburgh, UK

<sup>3</sup>University College London, UK

<sup>4</sup>University of Crete, Greece

<sup>5</sup>Radboud University Nijmegen, The Netherlands

<sup>6</sup>University Medical Center Groningen, The Netherlands

<sup>7</sup>Hörzentrum GmbH, Germany

<sup>8</sup>Fraunhofer Institute for Digital Media Technology, Germany

<sup>9</sup>Universidad del País Vasco, Spain

m.cooke@ikerbasque.org

**Abstract:** Speech is a hugely efficient means of communication: a reduced capacity in listening or speaking creates a significant barrier to social inclusion at all points through the lifespan, in education, work and at home. Hearing devices and speech synthesis can help address this reduced capacity but their use imposes greater listener effort. The goal of the EU-funded ENRICH project is to modify or augment speech with additional information to make it easier to process. Enrichment reduces listening burden by minimising cognitive load, while maintaining or improving intelligibility. ENRICH investigates the relationship between cognitive effort and natural and synthetic speech. Non-intrusive metrics for listening effort will be developed and used to design modification techniques which result in low-burden speech. The value of various enrichment approaches will be evaluated with individuals and cohorts with typically sub-optimal communication ability, such as children, hearing- or speech-impaired adults, non-native listeners and individuals engaged in simultaneous tasks.

**Keywords:** Intelligibility, listening effort

**Resumen:** El habla es un medio de comunicación sumamente eficiente: la capacidad reducida para oír o hablar crea barreras importantes para la inclusión social a lo largo de la vida en la educación, en el trabajo o en el hogar. Los audífonos y la síntesis del habla pueden ayudar a abordar esta capacidad reducida, pero su uso impone un mayor esfuerzo por parte del oyente. El objetivo del proyecto europeo ENRICH es modificar o aumentar el habla con información adicional resultando así más fácil de procesar. El enriquecimiento reduce el esfuerzo de escucha minimizando la carga cognitiva, mientras se mantiene o mejora la inteligibilidad. ENRICH investigará la relación entre el esfuerzo cognitivo y las diferentes formas de habla natural y sintética. Se desarrollarán métricas no intrusivas para el esfuerzo de escucha que se utilizarán para diseñar modificaciones que resulten en un habla de baja carga. El valor de los diversos enfoques de enriquecimiento se evaluará con individuos y cohortes con habilidades de comunicación típicamente subóptimas, como niños, adultos con problemas de audición o de habla, oyentes no nativos e individuos que realizan tareas simultáneas.

**Palabras clave:** Inteligibilidad, esfuerzo de escucha

## 1 Introduction

Speech is highly variable in terms of its clarity and intelligibility. Conversational speech is more difficult to understand than formal speech (e.g., broadcast speech) because speakers tend to reduce many sounds in casual conversations. Furthermore, speech may be difficult to understand when it is produced by non-native speakers or young children, or when it is produced in a very noisy or reverberant environment. Much of the focus in robust speech research has been on developing speech processing techniques aimed at improving speech intelligibility in challenging conditions and many advances have been made in this regard. For example, it is possible to improve the intelligibility of speech heard in noisy conditions via modifications carried out on speech before it is transmitted to a listener's ears (Sauert and Vary, 2006; Skowronski and Harris, 2006; Taal, Jensen, and Leijon, 2013; Zorila, Kandia, and Stylianou, 2012; Cooke et al., 2013; Godoy and Stylianou, 2013; Tang, Fazenda, and Cox, 2018). Other research has shown that manipulating the speech produced by adults with dysarthria can improve its intelligibility (Kain et al., 2007).

Because of these advances, improving intelligibility *per se* is no longer a sufficient goal. Understanding less-than-optimal speech also imposes a high cognitive load (Rönnberg et al., 2010; Pals, Sarampalis, and Baskent, 2013; Hornsby, 2013), and the resulting increase in listener effort may have a significant cost, in terms of comprehension lagging behind the speech, and in terms of its effect on an individual's well-being and ability to multitask successfully (e.g., listening to speech while driving). Less-than-optimal speech can be expected to be particularly troublesome for those listening in a non-native language, or with immature auditory-cognitive systems, or early hearing loss, or older adults facing age-related cognitive decline, or users who rely on hearing prostheses. Nearly all listeners fall into one or more of these categories at various points during their lifetime.

Given the success of speech modification in improving speech intelligibility in noise, the extent to which speech can be further transformed to increase its effectiveness in reducing cognitive load is an open question. Enrichment denotes both modifications to the

speech signal itself as well as augmentation, for instance by embedding additional contextual information such as text or visual cues into the speech signal, to be decoded at the receiver.

The goal of ENRICH, a Marie Skłodowska-Curie European Training Network, is to equip researchers with the capacity to produce innovative advances in the new field of speech enrichment. This pursuit is naturally inter-disciplinary, involving cognitive studies, modelling, engineering and real-world field evaluation with a range of listener groups, and will require future researchers who possess a broad range of skills as well as an appreciation of business practices and the needs of end-users. ENRICH members represent complementary academic disciplines with strong links to hospitals and schools, as well as businesses focused on the development of assistive technologies in speech and hearing.

ENRICH addresses the unmet need for multi-skilled practitioners and engineers in this rapidly growing sector currently facing a serious workforce shortage. Through a comprehensive training programme driven by the needs of industry and clinical practice, ENRICH is equipping 14 early-stage researchers (predoctoral fellows) with not just the necessary cross-disciplinary knowledge and research techniques, but also with experience of entrepreneurship and technology transfer so they can translate research findings into meaningful products and services that will facilitate spoken language communication in the coming decades.

## 2 Objectives

1. To provide world-class research training in a multidisciplinary/multi-sectorial setting
2. To generate new scientific findings in speech perception and production
3. To develop novel algorithms for speech enrichment
4. To demonstrate the benefit of enrichment on cognitive effort for listener groups and individuals
5. To pilot the application of speech enrichment in assistive technology

### 3 Partners

The ENRICH consortium consists of 16 beneficiaries and partners from academia, industry and clinical practice, collectively providing infrastructure for investigating spoken communication and for applying innovations to end-user populations.

#### Beneficiaries:

- University of the Basque Country (UPV-EHU), Spain (Coordinator)
- University of Edinburgh, UK
- University College London (UCL), UK
- Radboud University Nijmegen, Netherlands
- University of Crete, Greece
- University Medical Center Groningen, Netherlands
- Horzentrum, Germany
- Fraunhofer Institute for Digital Media Technology, Germany

#### Partners:

- Voxygen, France
- Toshiba, UK
- Basque Center for Cognition, Brain and Language, Spain
- FORTH, Greece
- Hacettepe University, Turkey
- Sonova, Switzerland
- Chania General Hospital, Greece
- University of Oldenburg, Germany

The consortium provides expertise in computational modelling of human speech perception in noise; metrics for intelligibility prediction; speech synthesis; casual and formal speech differences; relating variability in language performance across the adult life span to individual characteristics such as cognitive and hearing abilities; speech signal processing and modification; speech perception with hearing impairment and hearing devices, in relation to cognitive effort; commercial hearing-aid evaluation.

### 4 Research themes

The proposed work programme revolves around three main themes (Fig. 1).

Theme 1 explores the dimensions of cognitive effort in listening, and its relation with different forms of speech, natural and artificial. Building on this understanding, non-intrusive metrics for effort based on information in the speech signal are being developed

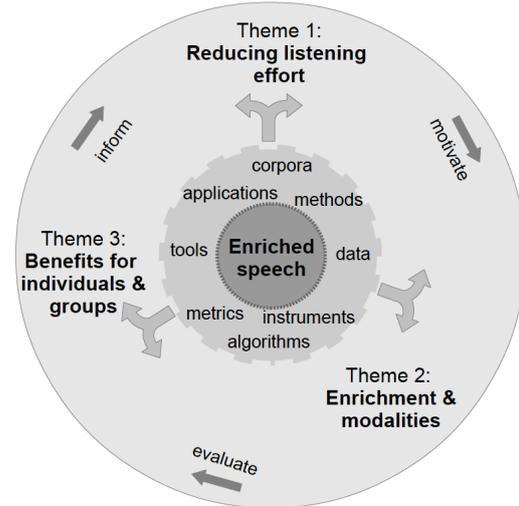


Figure 1: ENRICH research themes

to inform the design of effort-reducing enrichment techniques.

Theme 2 examines practical aspects of speech enrichment and explores issues such as how the speech signal can be augmented to contain further information useful to certain classes of listeners, or devices used by the hearing impaired. It also examines what forms of enrichment are suitable to different communication modes (e.g., passive listening, remote two- or multiparty conversations, near- and far-field broadcast).

Variability in the response of different individuals or specific populations to distinct forms of speech provides the setting for theme 3. This theme investigates the impact of specific enrichments on individuals and groups in order to better understand perceptual and cognitive processes. Key questions covered in this theme include: What can we learn from highly-trained listeners such as musicians or interpreters? Can enrichment ameliorate the consequences of gradual hearing deterioration over the long-term? To what extent does speech enrichment transfer across languages?

Theme 1 provides data and ideas to support the development of approaches to speech enrichment in theme 2. The value of proposed techniques for different classes of listeners is being evaluated in theme 3, which subsequently informs the design of further behavioural studies in theme 1. ENRICH benefits from network-wide speech data, experimental methods, instrumentation, tools and knowledge. This common infrastructure is being augmented with novel metrics for measuring

cognitive effort, algorithms for speech modification, evaluation frameworks and real-life technology applications.

## 5 Selected highlights

Studies to date have demonstrated that poor quality synthetic speech is more effortful to process (Govender and King, 2018) and that existing intelligibility enhancement techniques can also reduce cognitive load (Simantiraki, Cooke, and King, 2018). Intelligibility enhancing modifications can also be learnt by example (Shifas, Tsiaras, and Stylianou, 2018) or these can also occur on the perceptual side, for example, due to music training (Kaplan, Wagner, and Baskent, 2018).

More information about ENRICH is available at <http://www.enrich-etn.eu> and via the ENRICH video [https://youtu.be/\\_2W52Y3IE\\_Y](https://youtu.be/_2W52Y3IE_Y).

## Acknowledgements

ENRICH has received funding from the EU H2020 research and innovation programme under MSCA GA 675324.

## Bibliografía

Cooke, M., C. Mayo, C. Valentini-Botinhao, Y. Stylianou, B. Sauert, and Y. Tang. 2013. Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication*, 55:572–585.

Godoy, E. and Y. Stylianou. 2013. Increasing speech intelligibility via spectral shaping with frequency warping and dynamic range compression plus transient enhancement. In *Proc. Interspeech*, pages 3572–3576.

Govender, A. and S. King. 2018. Using Pupilometry to Measure the Cognitive Load of Synthetic Speech. In *Proc. Interspeech 2018*, pages 2838–2842.

Hornsby, B. 2013. The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hearing*, 34:523–534.

Kain, A., J.-P. Hosom, X. Niu, J. van Santen, M. Fried-Oken, and J. Staehely. 2007. Improving the intelligibility of dysarthric speech. *Speech Communication*, 49:743–759.

Kaplan, E., A. Wagner, and D. Baskent. 2018. Are musicians at an advantage when processing speech on speech? In *Proceedings of ICMPC15/ESCOM10. Graz, Austria: Centre for Systematic Musicology, University of Graz*.

Pals, C., A. Sarampalis, and D. Baskent. 2013. Listening effort with cochlear implant simulations. *J. Speech Hearing Language Res.*, 56:1075–1084.

Rönnerberg, J., M. Rudner, T. Lunner, and A. A. Zekveld. 2010. When cognition kicks in: Working memory and speech understanding in noise. *Noise and Health*, 12:263–269.

Sauert, B. and P. Vary. 2006. Near end listening enhancement: Speech intelligibility improvement in noisy environments. In *Proc. ICASSP*, pages 493–496, Toulouse, France, May.

Shifas, M., V. Tsiaras, and Y. Stylianou. 2018. Speech Intelligibility Enhancement Based on a Non-causal Wavenet-like Model. In *Proc. Interspeech 2018*, pages 1868–1872.

Simantiraki, O., M. Cooke, and S. King. 2018. Impact of Different Speech Types on Listening Effort. In *Proc. Interspeech 2018*, pages 2267–2271.

Skowronski, M. D. and J. G. Harris. 2006. Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments. *Speech Communication*, 48(5):549–558.

Taal, C. H., J. Jensen, and A. Leijon. 2013. On optimal linear filtering of speech for near-end listening enhancement. *IEEE Signal Proc. Let.*, 20(3):225–228.

Tang, Y., B. Fazenda, and T. Cox. 2018. Automatic speech-to-background ratio selection to maintain speech intelligibility in broadcasts using an objective intelligibility metric. *Applied Sciences*, 8:59, 01.

Zorila, T., V. Kandia, and Y. Stylianou. 2012. Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression. In *Proc. Interspeech*, pages 635–638.