

Paper Instructions and Template for Interspeech 2025

Anonymous submission to Interspeech 2025

Abstract

Manuscripts submitted to Interspeech 2025 must use this document as both an instruction set and as a template. Do not use a past paper as a template. Always start from a fresh copy, and read it all before replacing the content with your own. The main changes with respect to previous years' instructions are highlighted in blue.

Before submitting, check that your manuscript conforms to this template. If it does not, it may be rejected. Do not be tempted to adjust the format! Instead, edit your content to fit the allowed space. The maximum number of manuscript pages is 5. The 5th page is reserved exclusively for acknowledgements and references, which may begin on an earlier page if there is space.

The abstract is limited to 1000 characters. The one in your manuscript and the one entered in the submission form must be identical. Avoid non-ASCII characters, symbols, maths, italics, etc as they may not display correctly in the abstract book. Do not use citations in the abstract: the abstract booklet will not include a bibliography. Index terms appear immediately below the abstract.

Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

Templates are provided on the conference website for Microsoft Word®, and L^AT_EX. We strongly recommend L^AT_EX which can be used conveniently in a web browser on <https://overleaf.com> where this template is available in the Template Gallery.

1.1. General guidelines

Authors are encouraged to describe how their work relates to prior work by themselves and by others, and to make clear statements about the novelty of their work. This may require careful wording to avoid unmasking the identity of the authors in the version submitted for review (guidance in Section 1.2). All submissions must be compliant with the **ISCA Code of Ethics for Authors**, the **Pre-prints Policy**, and the **Conference Policy**. These can be found on the conference website under the "For authors" section.

1.1.1. Originality

All papers submitted to Interspeech 2025 must be original contributions that are not currently submitted to any other conference, workshop, or journal, nor will be submitted to any other conference, workshop, or journal during the review process of

Interspeech 2025. [Cross-checks with submissions from other conferences will be carried out to enforce this rule.](#)

1.1.2. Generative AI tools

All (co-)authors must be responsible and accountable for the work and content of the paper, and they must consent to its submission. Generative AI tools cannot be a co-author of the paper. They can be used for editing and polishing manuscripts, but should not be used for producing a significant part of the manuscript.

1.1.3. Conference theme

The theme of Interspeech 2025 is *Fair and Inclusive Speech Science and Technology*. Interspeech 2025 continues to be fully committed to advancing speech science and technology while meeting new challenges. Please refer to the conference website for further detail.

1.1.4. Participation in reviewing

Interspeech uses peer review to evaluate submitted manuscripts. At least one co-author of each paper should be volunteering to review for the Interspeech conference. [If you are eligible, but not yet a reviewer for ISCA conferences, you should sign up at](#)

<https://isca-speech.org/Reviewing>

[using the same email address as in your CMT account. The eligibility criteria can be located through the above link. Papers for which no author qualifies as ISCA reviewer do not have to follow this guideline. During submission authors should indicate the email of any co-authors who should be added to the reviewer database](#)

1.1.5. Accessibility recommendations

Since the theme of Interspeech 2025 is *Fair and Inclusive Speech Science and Technology*, we want this year's conference to take a step in the direction of more accessible and inclusive published papers as well. Authors are encouraged to have their papers fulfill specific requirements relating to **figures** (see Section 2.2) and **tables** (see Section 2.3). Additionally, while the authors should write in a style with which they feel comfortable, we encourage them to consider the use of **inclusive language**: [please do not unnecessarily use gendered nouns or pronouns when ungendered alternatives exist, unless you are referring to specific individuals.](#)

84	<i>1.1.6. Scientific reporting check list</i>		
85	The following checklist will be part of the submission form and		
86	includes a list of guidelines we expect Interspeech 2025 papers		
87	to follow. A compact version of the checklist is available under		
88	“For authors” section . Not every point in the check list will be		
89	applicable to every paper. Ideally, all guidelines that do apply to		
90	a certain paper should be satisfied. Nevertheless, not satisfying		
91	a guideline that applies to your paper is not necessarily grounds		
92	for rejection.		
93	1. Claims and limitations - for all papers		
94	• The paper clearly states what claims are being investigated,		
95	and why .		
96	• The novelties are clearly explained .		
97	• The main claims made in the abstract and introduction ac-		
98	curely reflect the paper’s contributions and scope.		
99	• The limitations of your work are described.		
100	• The relevant assumptions made in your work are stated in		
101	the paper.		
102			
103	2. If models are used, the paper includes information about the		
104	following :		
105	• Justification for each selected model, algorithm, or archi-		
106	tecture .		
107	• Sufficient details of the model (architecture), including ref-		
108	erence(s) to prior literature .		
109	3. If data sets are used, the paper includes information about the		
110	following:		
111	• Relevant details such as languages, audio duration distri-		
112	bution, number of examples, and label distributions, or a		
113	reference where that information can be found.		
114	• Details of train/validation (development)/test splits. Ide-		
115	ally, the lists should be released with the supplementary		
116	material if not already publicly available.		
117	• Explanation of all pre-processing steps, including any cri-		
118	teria used to exclude samples, if applicable.		
119	• Reference(s) to all data set(s) drawn from the existing lit-		
120	erature.		
121	• For newly collected data, a complete description of the		
122	data collection process, such as subjects, instructions to		
123	annotators and methods for quality control and a statement		
124	on whether ethics approval was necessary.		
125	4. If using non-public data sets or pre-trained models :		
126	• The paper includes a discussion on the reason/s for using		
127	non-public data sets.		
128	• Full details of the dataset are included in the paper to en-		
129	able comparison to similar data sets and tasks.		
130	• For papers that use publicly available, pre-trained (e.g.		
131	foundational or self-supervised) models, link to the model		
132	is provided . For papers that use non-public pre-trained		
133	models, description of the data (including quantity) used		
134	for training that model is provided .		
135	5. If reporting experimental results, the paper includes:		
136	• An explanation of evaluation metrics used.		
137	• An explanation of how models are initialized, if applicable.		
138	• Some measure of statistical significance of the reported		
139	gains or confidence intervals (a python toolkit and brief		
140	tutorial for computing confidence intervals with the boot-		
		strapping approach can be found in https://github.com/luferrer/ConfidenceIntervals .	141
		• A description of the computing infrastructure used and the	142
		average runtime for each model or algorithm (e.g. training,	143
		inference etc).	144
		• The number of parameters in each model.	145
		6. If hyperparameter search (including choice of architecture or	146
		features and any other development decision) was done, the	147
		paper includes:	148
		• Final results on a held-out evaluation set not used for hy-	149
		perparameter tuning.	150
		• Hyperparameter configurations for best-performing mod-	151
		els.	152
		• The method for choosing hyperparameter values to ex-	153
		ply, the range of the hyperparameter values , and the cri-	154
		terion used to select among them.	155
		7. If source code is used:	156
		• The code is or will be made publicly available and/or suf-	157
		ficient details are included in the paper for the work to be	158
		reproduced.	159
		• For publicly available software, the corresponding version	160
		numbers and links and/or references to the software.	161
		For any paper that uses machine learning, we encour-	162
		age authors to additionally consult <i>Good practices for eval-</i>	163
		uation machine learning systems https://arxiv.org/abs/2412.03700	164
		to understand the reasons for the above	165
		reporting guidelines .	166
			167
		1.2. Double-blind review	168
		Interspeech 2025 uses double-blind review to support the in-	169
		tegrity of the review process.	170
		<i>1.2.1. Version submitted for review</i>	171
		The manuscript submitted for review must not include any in-	172
		formation that might reveal the authors’ identities or affilia-	173
		tions. This also applies to the metadata in the submitted PDF	174
		file (guidance in Section 4.1.2), uploaded multimedia and on-	175
		line material (guidance in Section 4.2).	176
		Take particular care to cite your own work in a way that	177
		does not reveal that you are also the author of that work. For	178
		example, do not use constructions like “In previous work [23],	179
		we showed that ...”. Instead use something like “Jones et al.	180
		[23] showed that ...”.	181
		Acknowledgements should not be included in the ver-	182
		sion submitted for review.	183
		Papers that reveal the identity of the authors will be rejected	184
		without review. Note that the full list of authors must still be	185
		provided in the online submission system, since this is neces-	186
		sary for detecting conflicts of interest.	187
		<i>1.2.2. Camera-ready version</i>	188
		Authors should include names and affiliations in the final ver-	189
		sion of the manuscript, for publication. \LaTeX users can do this	190
		simply by uncommenting <code>\interspeechcameraready</code> .	191
		The maximum number of authors in the author list is 20. If the	192
		number of contributing authors is more than this, they should	193
		be listed in a footnote or the Acknowledgements section. In-	194
		clude the country as part of each affiliation. Do not use com-	195
		pany logos anywhere in the manuscript, including in affilia-	196

197 tions and Acknowledgements. After acceptance, authors may of
198 course reveal their identity in other ways, including: adjusting
199 the wording around self-citations; adding further acknowledge-
200 ments; updating multimedia and online material. Acknowledg-
201 egements, if any, should be added in the camera-ready version.

202 1.2.3. Pre-prints

203 Please, see [https://interspeech2025.org/
204 submission-policy/](https://interspeech2025.org/submission-policy/) for a detailed explanation of
205 the policy on pre-prints for this Interspeech. The period of
206 anonymity during which a version of the submitted paper
207 may not be posted or updated online starts a month before the
208 Interspeech submission deadline and runs until accept/reject
209 decisions are announced.

210 2. Format

211 2.1. Layout

212 Authors should observe the following specification for page lay-
213 out by using the provided template. Do not modify the template
214 layout! Do not reduce the line spacing!

215 2.1.1. Page layout

- 216 • Paper size must be DIN A4.
- 217 • Two columns are used except for the title section and for large
218 figures or tables that may need a full page width.
- 219 • Left and right margin are 20 mm each.
- 220 • Column width is 80 mm.
- 221 • Spacing between columns is 10 mm.
- 222 • Top margin is 25 mm (except for the first page which is
223 30 mm to the title top).
- 224 • Bottom margin is 35 mm.
- 225 • Text height (without headers and footers) is maximum
226 235 mm.
- 227 • Page headers and footers must be left empty.
- 228 • No page numbers.
- 229 • Check indentations and spacing by comparing to the example
230 PDF file.

231 2.1.2. Section headings

232 Section headings are centred in boldface with the first word cap-
233 italised and the rest of the heading in lower case. Sub-headings
234 appear like major headings, except they start at the left margin
235 in the column. Sub-sub-headings appear like sub-headings, ex-
236 cept they are in italics and not boldface. See the examples in
237 this file. No more than 3 levels of headings should be used.

238 2.1.3. Fonts

239 Times or Times Roman font is used for the main text. Font size
240 in the main text must be 9 points, and in the References section
241 8 points. Other font types may be used if needed for special pur-
242 poses. L^AT_EX users should use Adobe Type 1 fonts such as Times
243 or Times Roman, which is done automatically by the provided
244 L^AT_EX class. Do not use Type 3 (bitmap) fonts. Phonetic tran-
245 scriptions should be placed between forward slashes and pho-
246 netic transcriptions between square brackets, for example /lɔː
247 ænd ɔːdɜː/ vs. [lɔːrɛnɔːdɜː], and authors are encouraged to use
248 the terms ‘phoneme’ and ‘phone’ correctly [1].

249 2.1.4. Hyperlinks

250 For technical reasons, the proceedings editor will strip all ac-
251 tive links from the papers during processing. URLs can be
252 included in your paper if written in full, e.g., [https://
253 www.interspeech2025.org/](https://www.interspeech2025.org/). The text must be all black.
254 Please make sure that they are legible when printed on paper.

255 2.2. Figures

256 Figures must be centred in the column or page. Figures
257 which span 2 columns must be placed at the top or bottom
258 of a page. Captions should follow each figure and have the
259 format used in Figure 1. Diagrams should preferably be
260 vector graphics. Figures must be legible when printed in
261 monochrome on DIN A4 paper; a minimum font size of 8
262 points for all text within figures is recommended. Diagrams
263 must not use stipple fill patterns because they will not reproduce
264 properly in Adobe PDF. See *APA Guidelines regarding the
265 use of color in Figures* ([https://apastyle.apa.org/
266 style-grammar-guidelines/tables-figures/
267 colors](https://apastyle.apa.org/style-grammar-guidelines/tables-figures/colors)).

268 From an accessibility standpoint, please ensure that your
269 paper remains coherent for a reader who cannot see your fig-
270 ures. Please make sure that:

- 271 • all figures are referenced in the text.
- 272 • figures are fully explained either in text, in their caption, or
273 in a combination of the two.
- 274 • instead of visualizing differences between areas in your fig-
275 ures only by color, those differences can also be illustrated
276 by using different textures or patterns.
- 277 • all text, including axis labels, is at least as large as the content
278 text surrounding the figures.

279 2.3. Tables

280 An example of a table is shown in Table 1. The caption text
281 must be above the table. Tables must be legible when printed in
282 monochrome on DIN A4 paper; a minimum font size of 8 points
283 is recommended. From an accessibility standpoint, please make
284 sure that:

- 285 • tables are not images.
- 286 • all table headings are clearly marked as such, and are distinct
287 from table cells.
- 288 • all tables are referenced in the text.

Table 1: *This is an example of a table*

Ratio	Decibels
1/10	-20
1/1	0
2/1	≈ 6
3.16/1	10
10/1	20

289 2.4. Equations

290 Equations should be placed on separate lines and numbered. We
291 define

$$\begin{aligned} x(t) &= s(t') \\ &= s(f_{\omega}(t)) \end{aligned} \quad (1)$$

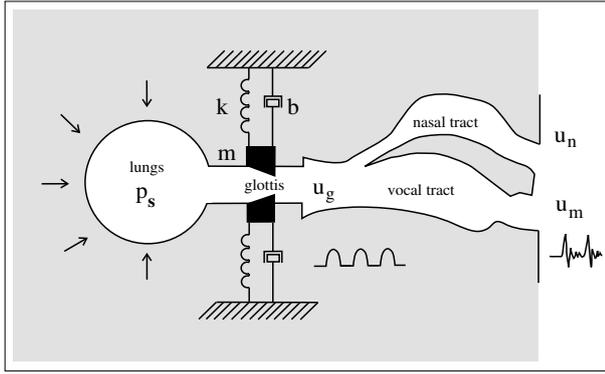


Figure 1: Schematic diagram of speech production.

where $f_\omega(t)$ is a special warping function. Equation 2 is a little more complicated.

$$f_\omega(t) = \frac{1}{2\pi j} \oint_C \frac{\nu^{-1k} d\nu}{(1 - \beta\nu^{-1})(\nu^{-1} - \beta)} \quad (2)$$

2.5. Style

Manuscripts must be written in English. Either US or UK spelling is acceptable (but do not mix them).

2.5.1. References

It is ISCA policy that papers submitted to Interspeech should refer to peer-reviewed publications. References to non-peer-reviewed publications (including public repositories such as arXiv, Preprints, and HAL, software, and personal communications) should only be made if there is no peer-reviewed publication available, should be kept to a minimum, and should appear as footnotes in the text (i.e., not listed in the References).

References should be in standard IEEE format, numbered in order of appearance, for example [2] is cited before [3]. For longer works such as books, provide a single entry for the complete work in the References, then cite specific pages [4, pp. 417–422] or a chapter [4, Chapter 2]. Multiple references may be cited in a list [5, 6].

2.5.2. International System of Units (SI)

Use SI units, correctly formatted with a non-breaking space between the quantity and the unit. In \LaTeX this is best achieved using the `siunitx` package (which is already included by the provided \LaTeX class). This will produce 25 ms, 44.1 kHz and so on.

3. Specific information for Microsoft Word

For ease of formatting, please use the styles listed in Table 2. The styles are defined in the Word version of this template and are shown in the order in which they would be used when writing a paper. When the heading styles in Table 2 are used, section numbers are no longer required to be typed in because they will be automatically numbered by Word. Similarly, reference items will be automatically numbered by Word when the “Reference” style is used.

If your Word document contains equations, you must not save your Word document from “.docx” to “.doc” because this

will convert all equations to images of unacceptably low resolution.

4. Submissions

Information on how and when to submit your paper is provided on the conference website.

4.1. Manuscript

Authors are required to submit a single PDF file of each manuscript. The PDF file should comply with the following requirements: (a) no password protection; (b) all fonts must be embedded; (c) text searchable (do ctrl-F and try to find a common word such as “the”), and (d) the document properties do not reveal the identity or username. The conference organisers may contact authors of non-complying files to obtain a replacement. Papers for which an acceptable replacement is not provided in a timely manner will be withdrawn.

4.1.1. Embed all fonts

It is very important that the PDF file embeds all fonts! PDF files created using \LaTeX , including on <https://overleaf.com>, will generally embed all fonts from the body text. However, it is possible that included figures (especially those in PDF or PS format) may use additional fonts that are not embedded, depending how they were created.

On Windows, the bullzip printer can convert any PDF to have embedded and subsetted fonts. On Linux & MacOS, converting to and from Postscript will embed all fonts:

```
pdf2ps file.pdf
ps2pdf -dPDFSETTINGS=/prepress file.ps file.pdf
```

4.1.2. Sanitise PDF metadata

Check that author identity is not revealed in the PDF metadata. The provided \LaTeX class ensures this. Metadata can be inspected using a PDF viewer.

Table 2: Main predefined styles in Word

Style Name	Entities in a Paper
Title	Title
Author	Author name
Affiliation	Author affiliation
Email	Email address
AbstractHeading	Abstract section heading
Body Text	First paragraph in abstract
Body Text Next	Following paragraphs in abstract
Index	Index terms
1. Heading 1	1 st level section heading
1.1 Heading 2	2 nd level section heading
1.1.1 Heading 3	3 rd level section heading
Body Text	First paragraph in section
Body Text Next	Following paragraphs in section
Figure Caption	Figure caption
Table Caption	Table caption
Equation	Equations
• List Bullet	Bulleted lists
[1] Reference	References

360 4.2. Optional supplementary material or links to online 361 material

362 4.2.1. Supplementary material

363 Interspeech offers the option of submitting supplementary ma-
364 terial. These files are meant for code, data or audio-visual il-
365 lustrations that cannot be conveyed in text, tables and graphs.
366 Just as with figures used in your manuscript, make sure that you
367 have sufficient author rights to all other materials that you sub-
368 mit. This material will be available to reviewers and committee
369 members for evaluating the paper, but it will not be published
370 with the paper or accessible in the archives. To make material
371 available for future readers, please use online resources (see
372 section 4.2.2). Supplementary material is intended to support
373 the reviewers in assessing the paper; however, the paper should
374 be understandable on its own. Reviewers may choose to exam-
375 ine the supplementary material but are not required to do so.

376 In line with the double-blind review policy, make sure that
377 the supplementary material is anonymised. Unless strictly nec-
378 essary, avoid to provide code at this step, which is usually harder
379 to anonymise.

380 Your supplementary material must be submitted in a single
381 ZIP file limited to 100 MB for each separate paper. Within the
382 ZIP file you can use folders to organise the files. In the ZIP file
383 you should include a README.txt or index.html file to
384 describe the content. In the manuscript, refer to a supplement-
385 ary file by filename. Use short file names with no spaces.

386 4.2.2. Online resources such as web sites, blog posts, code, and 387 data

388 It is common for manuscripts to provide links to web sites (e.g.,
389 as an alternative to including a ZIP file as supplementary ma-
390 terial), code repositories, data sets, or other online resources.
391 Provision of such materials is generally encouraged; however,
392 they should not be used to circumvent the limit on manuscript
393 length. In particular, reviewers will not be required to read any
394 content other than the main manuscript.

395 Links included in the version submitted for review should
396 not reveal the identity or affiliation of the authors. If this is not
397 possible, then the links should not be provided in the version
398 submitted for review, but they can be added in the final version
399 of the paper, if accepted. A placeholder can be included in the
400 paper with a text like: “[to ensure author anonymity, the link
401 to the resource will be added after the review process]”. Links
402 and repositories provided for review should be anonymous. In
403 particular, names in repositories are forbidden.

404 Authors are encouraged to group any supplementary
405 material in a single repository. If the paper is accepted,
406 authors will have the option to include a single link to
407 this repository in the paper’s metadata on ISCA Archives.
408 For that purpose, the link must be a DOI link. Free
409 DOI links can for instance be obtained through Zen-
410 do, with additional guidance available in GitHub Docs
411 ([https://docs.github.com/en/repositories/
412 archiving-a-github-repository/
413 referencing-and-citing-content](https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content))

414 5. Discussion

415 Authors must proofread their PDF file prior to submission, to
416 ensure it is correct. Do not rely on proofreading the L^AT_EX
417 source or Word document. **Please proofread the PDF file be-
418 fore it is submitted.**

6. Acknowledgements

Acknowledgement should only be included in the camera-ready
version, not in the version submitted for review. The 5th page
is reserved exclusively for acknowledgements and references.
No other content must appear on the 5th page. Appendices, if
any, must be within the first 4 pages. The acknowledgments and
references may start on an earlier page, if there is space.

The authors would like to thank ISCA and the organising
committees of past Interspeech conferences for their help and
for kindly providing the previous version of this template.

7. References

- [1] R. K. Moore and L. Skidmore, “On the use/misuse of the term ‘Phoneme’,” in *Proc. INTERSPEECH 2019 – 20th Annual Conference of the International Speech Communication Association*, Graz, Austria, Sep. 2019, pp. 2340–2344.
- [2] S. B. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980.
- [3] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.
- [4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. New York: Springer, 2009.
- [5] J. Smith, F. Lastname2, and F. Lastname3, “A really good paper about Dynamic Time Warping,” in *Proc. INTERSPEECH 2022 – 23rd Annual Conference of the International Speech Communication Association*, Incheon, Korea, Sep. 2022, pp. 100–104.
- [6] R. Jones, F. Lastname2, and F. Lastname3, “An excellent paper introducing the ABC toolkit,” in *Proc. INTERSPEECH 2022 – 23rd Annual Conference of the International Speech Communication Association*, Incheon, Korea, Sep. 2022, pp. 105–109.