

# Contrastive Learning for Speech Enhancement

郭品辰  
F14066127  
NCKU SNAME

黃仁鴻  
P76094169  
NCKU CSIE

## Abstract

### 1. Introduction

許多與日常生活息息相關的任務都是仰賴語音作為資訊傳遞的媒介，像是電話通訊、語音辨識與助聽器等等。然而在現實環境中充滿著各種不可預期的噪音干擾，這嚴重影響了語音訊號技術的效能。因此，將這些雜訊去除的語音增強技術就成了對語音任務很重要前置處理單元。[3]

而語音增強的問題就是不論在何種噪音環境，面對相同的語音，模型都能夠抽取出相同的特徵，進而利用與特徵來重構語音。這部分想法與近年流行自監督方法中的對比學習(Contrastive Learning)不謀而合，比對學習希望相似的樣本，其特徵編碼能越像越好，而不同樣本的特徵編碼則是相距越遠越好。

我們認為，藉由 CL 的方法來學習語音特徵的潛在編碼，再利用此特徵編碼還原語音，應該會具備比一般深度學習的語音增強方法更高的通用性與泛化能力。

#### 1.1. Language

All manuscripts must be in English [4].

#### 1.2. Dual submission

Please refer to the author guidelines on the CVPR 2020 web page for a discussion of the policy on dual submissions. [3]

#### 1.3. Paper length

Papers, excluding the references section, must be no longer than eight pages in length. The references section will not be included in the page count, and there is no limit on the length of the references section. For example, a paper of eight pages with two pages of references would have a total length of 10 pages. **There will be no extra page charges for CVPR 2020.**

Overlength papers will simply not be reviewed. This includes papers where the margins and formatting are deemed

to have been significantly altered from those laid down by this style guide. Note that this L<sup>A</sup>T<sub>E</sub>X guide already sets figure captions and references in a smaller font. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in eight pages if it is reviewed in eleven.

#### 1.4. The ruler

The L<sup>A</sup>T<sub>E</sub>X style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-L<sup>A</sup>T<sub>E</sub>X document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (L<sup>A</sup>T<sub>E</sub>X users may uncomment the `\cvprfinalcopy` command in the document preamble.) Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 095.5), although in most cases one would expect that the approximate location will be adequate.

#### 1.5. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: <http://www.pamitc.org/documents/mermin.pdf>.

## 1.6. Blind review

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one's own work—in fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words “my” or “our” when citing previous work. That is all. (But see below for techreports.)

Saying “this builds on the work of Lucy Smith [1]” does not say that you are Lucy Smith; it says that you are building on her work. If you are Smith and Jones, do not say “as we show in [7]”, say “as Smith and Jones show in [7]” and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper just asking to be rejected:

An analysis of the frobnicable foo filter.

In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Removed for blind review

An example of an acceptable paper:

An analysis of the frobnicable foo filter.

In this paper we present a performance analysis of the paper of Smith *et al.* [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.

[1] Smith, L and Jones, C. “The frobnicable foo filter, a fundamental contribution to human knowledge”. *Nature* 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [4] as additional material and cite it as

[1] Authors. “The frobnicable foo filter”, F&G 2014 Submission ID 324, Supplied as additional material `fg324.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go

to a techreport for further details. Thus, you may say in the body of the paper “further details may be found in [5]”. Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled “Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties”, by Zeus *et al.*

You can handle this paper like any other. Don't write “We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]”. That would be silly, and would immediately identify the authors. Instead write the following:

We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus *et al.* 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.

...

The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al.*, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

### FAQ

**Q:** Are acknowledgements OK?

**A:** No. Leave them for the final copy.

**Q:** How do I cite my results reported in open challenges?

**A:** To conform with the double blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

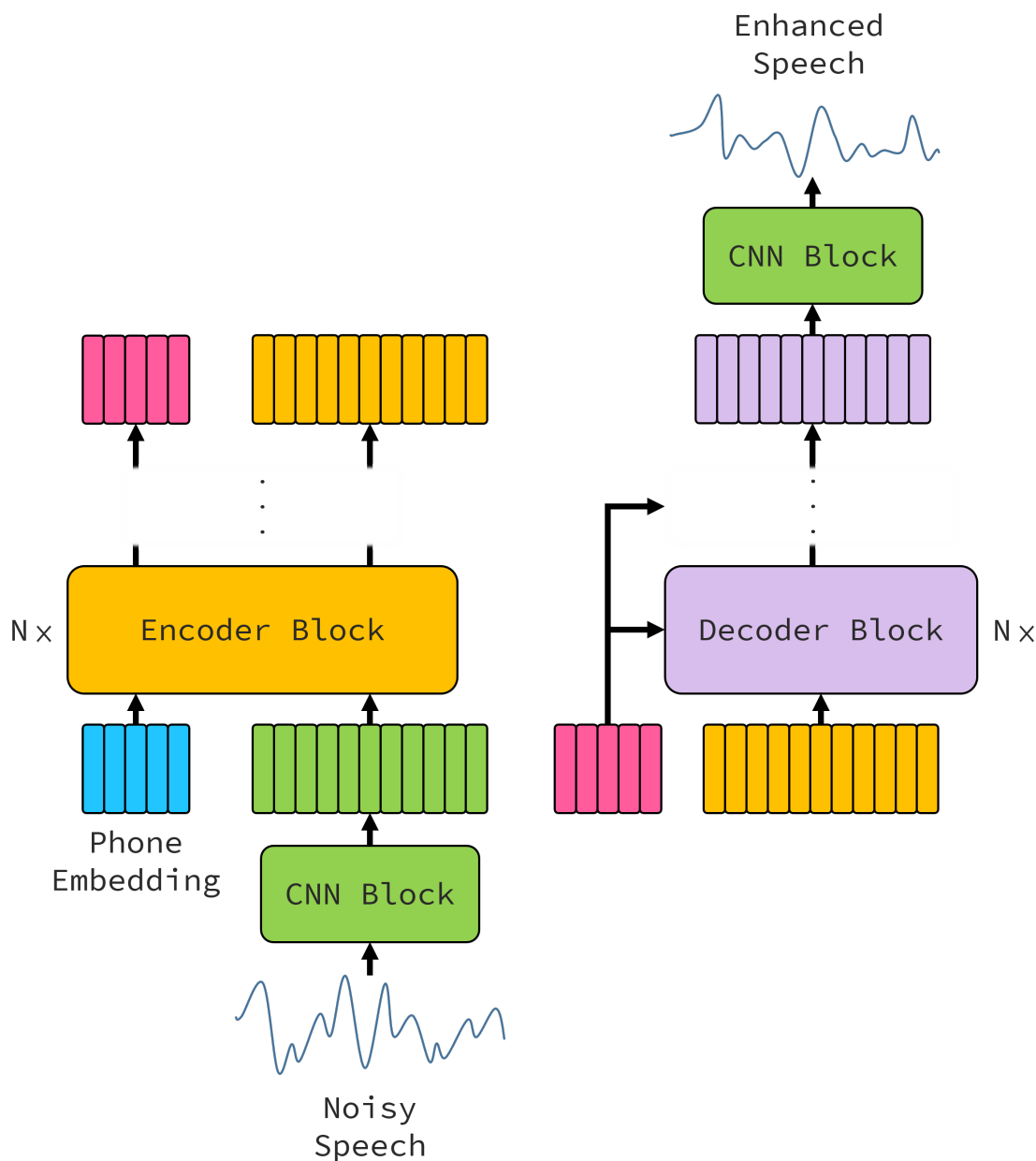


Figure 1. Example of a short caption, which should be centered.

## 1.7. Miscellaneous

Compare the following:

`$conf_a$`                       $conf_a$   
`$\mathit{conf}_a$`               $conf_a$   
 See The  $\text{\TeX}$ book, p165.

The space after *e.g.*, meaning “for example”, should not be a sentence-ending space. So *e.g.* is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space

by using “et alia”, shortened to “*et al.*” (not “*et. al.*” as “*et*” is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: “Frobination has been trendy lately. It was introduced by Alpher [1], and subsequently developed by Alpher and Fotheringham-Smythe [2], and Alpher *et al.* [3].”

This is incorrect: “... subsequently developed by Alpher *et al.* [2] ...” because reference [2] has just two authors. If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sen-

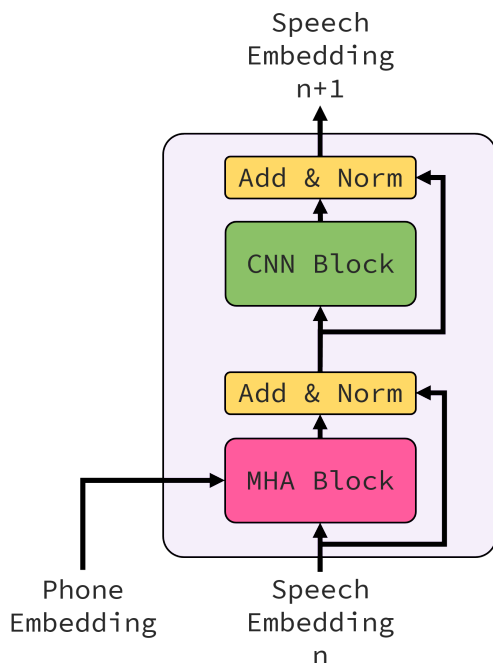


Figure 2. Example of caption. It is set in Roman so that mathematics (always set in Roman:  $B \sin A = A \sin B$ ) may be included without an ugly clash.

tence as in Alpher *et al.*

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [2, 1, 4] to [1, 2, 4].

## 2. System framework

因應實驗環境的硬體限制，本專題將會基於 BYOL 與 SimSiam 這兩種在小型 batch size 也有良好成效的對比學習方法進行實驗。

All text must be in a two-column format. The total allowable width of the text area is  $6\frac{7}{8}$  inches (17.5 cm) wide by  $8\frac{7}{8}$  inches (22.54 cm) high. Columns are to be  $3\frac{1}{4}$  inches (8.25 cm) wide, with a  $\frac{5}{16}$  inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for  $8.5 \times 11$ -inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

### 2.1. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area  $6\frac{7}{8}$  inches (17.5 cm) wide by  $8\frac{7}{8}$  inches (22.54 cm) high. Page numbers should be in footer with page numbers, centered and .75 inches from the bottom of the page and make it start at the

correct page number rather than the 4321 in the example. To do this fine the line (around line 23)

```
%\ifcvprfinal\pagestyle{empty}\fi
\setcounter{page}{4321}
```

where the number 4321 is your assigned starting page.

Make sure the first page is numbered by commenting out the first page being empty on line 46

```
%\thispagestyle{empty}
```

### 2.2. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

**MAIN TITLE.** Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

**AUTHOR NAME(s) and AFFILIATION(s)** are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The **ABSTRACT** and **MAIN TEXT** are to be in a two-column format.

**MAIN TEXT.** Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 2 and 3. Short captions should be centred.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

**FIRST-ORDER HEADINGS.** (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

**SECOND-ORDER HEADINGS.** (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

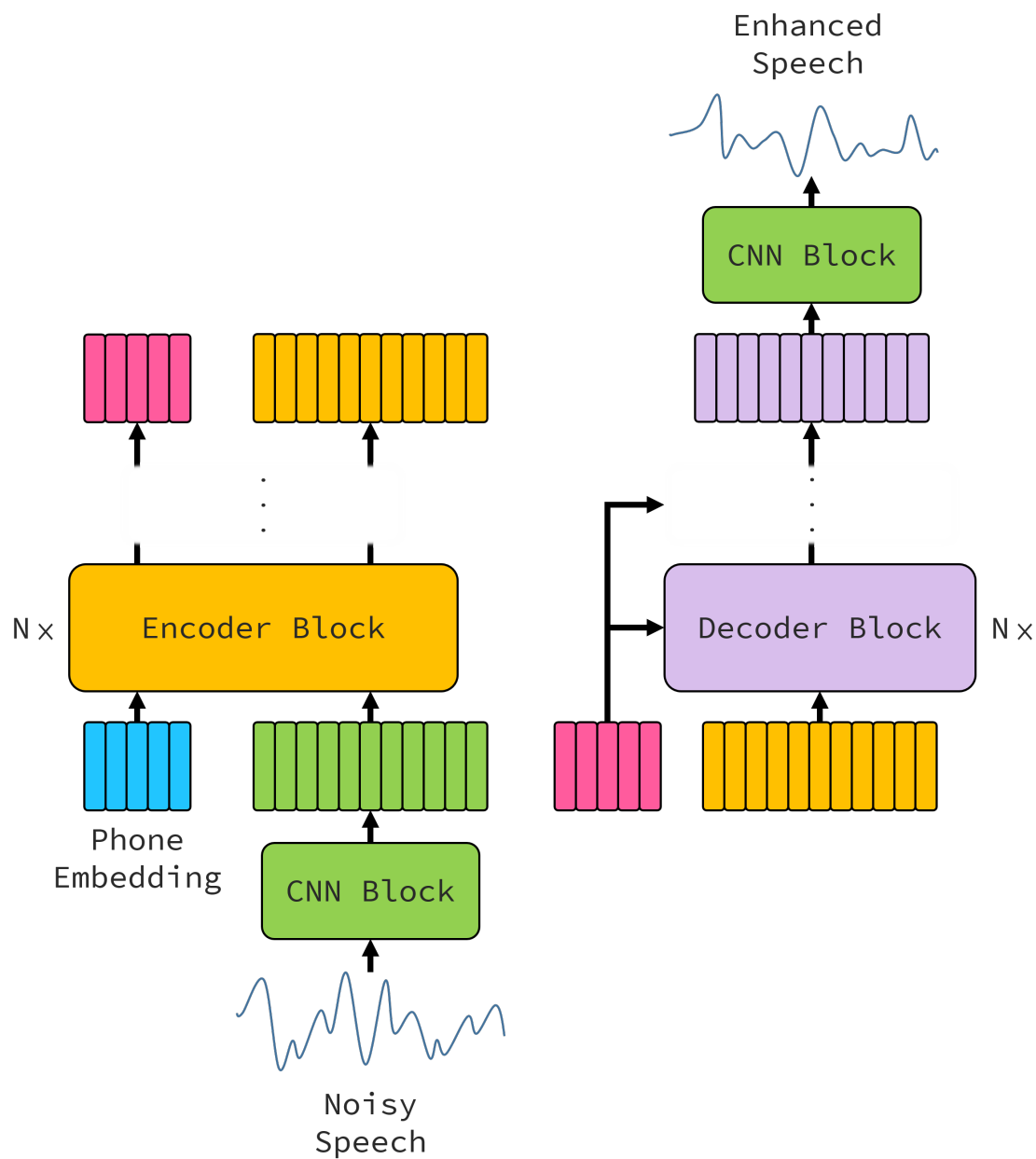


Figure 3. Example of a short caption, which should be centered.

## 2.3. Footnotes

Please use footnotes<sup>1</sup> sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

<sup>1</sup>This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

## 2.4. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [4]. Where appropriate, include the name(s) of editors of referenced books.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 1. Results. Ours is better.

## 2.5. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in L<sup>A</sup>T<sub>E</sub>X, it's almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]
{myfile.eps}
```

## 2.6. Color

Please refer to the author guidelines on the CVPR 2020 web page for a discussion of the use of color in your document.

## 3. Expected results

本專題預計將會進行以下幾項實驗：

1. 直接對 AutoEncoder 訓練 Speech Enhancement 任務。
2. 基於 BYOL 對 Encoder 與 Bottleneck 進行預訓練後凍結參數，然後接上 Decoder 訓練 Speech Enhancement 任務。
3. 基於 SimSiam 對 Encoder 與 Bottleneck 進行預訓練後凍結參數，然後接上 Decoder 訓練 Speech Enhancement 任務。
4. 將 BYOL 訓練得到的參數作為初始權重後，對 AutoEncoder 訓練 Speech Enhancement 任務。
5. 將 SimSiam 訓練得到的參數作為初始權重後，對 AutoEncoder 訓練 Speech Enhancement 任務。
6. 基於不同數量的樣本進行上述實驗，以測試各種方法的泛化能力。

並比較上述模型的性能優劣，並將最好的模型包裝為桌面應用。

## References

- [1] FirstName Alpher. Frobnication. *Journal of Foo*, 12(1):234–778, 2002.
- [2] FirstName Alpher and FirstName Fotheringham-Smythe. Frobnication revisited. *Journal of Foo*, 13(1):234–778, 2003.
- [3] FirstName Alpher, FirstName Fotheringham-Smythe, and FirstName Gamow. Can a machine frobnicate? *Journal of Foo*, 14(1):234–778, 2004.
- [4] Authors. The frobnicable foo filter, 2014. Face and Gesture submission ID 324. Supplied as additional material `fg324.pdf`.
- [5] Authors. Frobnication tutorial, 2014. Supplied as additional material `tr.pdf`.