

# CONSTRUCTING PRONUNCIATION LEXICON WITHOUT PHONEME INVENTORY

A. Guntario Sukma Cahyani\*

Institut Teknologi Bandung  
Departement of Informatics  
Jl. Ganeca no. 10 Bandung, Indonesia

B. Sakriani Sakti, C. Satoshi Nakamura†

Nara Institut of Science and Technology  
Departement of Information System  
891615, Takayamacho, Ikoma, Nara, Japan

## ABSTRACT

Pronunciation lexicon plays an important role in automatic speech recognition (ASR) system. Often times, pronunciation lexicon uses canonical form i.e. the correct pronunciations as agreed by experts, but in some cases, the language does not have known phoneme inventory, as in the case of low resource language, or the available phoneme inventory i.e. the canonical form of the pronunciation does not represent the surface pronunciation that are uttered by the target speaker, as in the case of non-native and code-switching ASR. For such cases, pronunciation lexicon cannot be built manually, since the phonemes are practically unknown. In this paper, we investigate several methods of lexicon building for the case of low-resource language: data-driven mapping which generated pronunciations of low-resource language words using rich-resource language phonemes inventory, unsupervised learning that discovered new acoustic units, and we proposed a method that is a combination of the two. We carried out the experiment in two general scenarios: first scenario that served as an analysis, using Indonesian language as a low-resource language and English as rich-resource language, and the second scenario as an implementation using Indonesian Ethnic language as the low-resource language and Indonesian as the rich resource language. We found that the best method for building lexicon without phoneme inventory is the method we proposed, and we implement it to the real low-resource language ASR.

**Index Terms**— pronunciation lexicon, automatic speech recognition, lexicon building, low-resource language

## 1. INTRODUCTION

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. Please follow them and if you have any questions, direct them to Conference Management Services, Inc.: Phone +1-979-846-6800 or email to [papers@asru2017.org](mailto:papers@asru2017.org).

---

\*

†

## 2. FORMATTING YOUR PAPER

All printed material, including text, illustrations, and charts, must be kept within a print area of 7 inches (178 mm) wide by 9 inches (229 mm) high. Do not write or print anything outside the print area. The top margin must be 1 inch (25 mm), except for the title page, and the left margin must be 0.75 inch (19 mm). All *text* must be in a two-column format. Columns are to be 3.39 inches (86 mm) wide, with a 0.24 inch (6 mm) space between them. Text must be fully justified.

## 3. PAGE TITLE SECTION

The paper title (on the first page) should begin 1.38 inches (35 mm) from the top edge of the page, centered, completely capitalized, and in Times 14-point, boldface type. The authors' name(s) and affiliation(s) appear below the title in capital and lower case letters. Papers with multiple authors and affiliations may require two or more lines for this information. Please note that papers should not be submitted blind; include the authors' names on the PDF.

## 4. TYPE-STYLE AND FONTS

To achieve the best rendering both in printed proceedings and electronic proceedings, we strongly encourage you to use Times-Roman font. In addition, this will give the proceedings a more uniform look. Use a font that is no smaller than nine point type throughout the paper, including figure captions.

In nine point type font, capital letters are 2 mm high. **If you use the smallest point size, there should be no more than 3.2 lines/cm (8 lines/inch) vertically.** This is a minimum spacing; 2.75 lines/cm (7 lines/inch) will make the paper much more readable. Larger type sizes require correspondingly larger vertical spacing. Please do not double-space your paper. TrueType or Postscript Type 1 fonts are preferred.

The first paragraph in each section should not be indented, but all the following paragraphs within the section should be indented as these paragraphs demonstrate.

## 5. MAJOR HEADINGS

Major headings, for example, "1. Introduction", should appear in all capital letters, bold face if possible, centered in the column, with one blank line before, and one blank line after. Use a period (".") after the heading number, not a colon.

### 5.1. Subheadings

Subheadings should appear in lower case (initial word capitalized) in boldface. They should start at the left margin on a separate line.

#### 5.1.1. Sub-subheadings

Sub-subheadings, as in this paragraph, are discouraged. However, if you must use them, they should appear in lower case (initial word capitalized) and start at the left margin on a separate line, with paragraph text beginning on the following line. They should be in italics.

## 6. PRINTING YOUR PAPER

Print your properly formatted text on high-quality, 8.5 x 11-inch white printer paper. A4 paper is also acceptable, but please leave the extra 0.5 inch (12 mm) empty at the BOTTOM of the page and follow the top and left margins as specified. If the last page of your paper is only partially filled, arrange the columns so that they are evenly balanced if possible, rather than having one long column.

In LaTeX, to start a new column (but not a new page) and help balance the last-page column lengths, you can use the command "`\pagebreak`" as demonstrated on this page (see the LaTeX source below).

## 7. PAGE NUMBERING

Please do **not** paginate your paper. Page numbers, session numbers, and conference identification will be inserted when the paper is included in the proceedings.

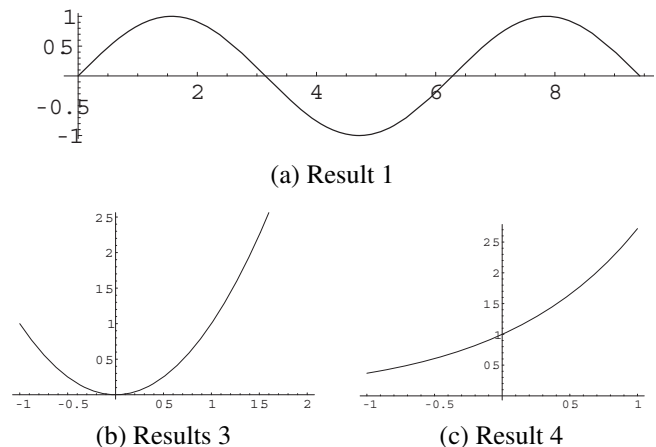
## 8. ILLUSTRATIONS, GRAPHS, AND PHOTOGRAPHS

Illustrations must appear within the designated margins. They may span the two columns. If possible, position illustrations at the top of columns, rather than in the middle or at the bottom. Caption and number every illustration. All halftone illustrations must be clear black and white prints. Colors may be used, but they should be selected so as to be readable when printed on a black-only printer.

Since there are many ways, often incompatible, of including images (e.g., with experimental results) in a LaTeX document, below is an example of how to do this [?].

## 9. FOOTNOTES

Use footnotes sparingly (or not at all!) and place them at the bottom of the column on the page on which they are referenced. Use Times 9-point type, single-spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).



**Fig. 1.** Example of placing a figure with experimental results.

## 10. COPYRIGHT FORMS

You must include your fully completed, signed IEEE copyright release form when you submit your paper. We **must** have this form before your paper can be published in the proceedings.

## 11. REFERENCES

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [?]. An additional final page (the fifth page, in most cases) is allowed, but must contain only references to the prior literature.