

Lark VC-English Transcription Instruction-Transcriber

Need Native English Speaker ONLY

Please check this file frequently, if there are any updates, we'll mark them in light yellow.

- **What do you need to do?**

Read instructions → do the test on google forms → pass and get the message from Feishu (Our colleagues will contact you as well) → do the trial on the test queue → start the tasks on formal queue

Video Instruction:

- ★ Lark VC-English Transcription-Training: <https://youtu.be/lLLs83qMNE4>
- ★ Lark VC Summary of Questions: <https://youtu.be/mzwOiM65o8A>
- ★ Lark VC Mistakes Example: <https://youtu.be/sEM2AfNYmik>
- ★ **LarkVC Feedback from Client (Mistakes):**
<https://www.feishu.cn/docs/doccnDW2YNSNgT9WIhxrOag6zc>

Lark VC_Summary of Questions

- **Please check the above ↑ file frequently, and avoid any mistakes that are mentioned in the file!!!**

Frequently Made Mistakes:

1. Please capitalized **i if I is used as a first-person pronoun**

eg:

I have a dream. **100**

I'm online.

i have a dream. **×**

i'm online **×**

2. If at the end of the audio, the speaker obviously did not finish the sentence, please do not add a period (full stop) in the end!!!

I'm not sure if my mother went to the. **×**

(If you can't tell whether the speaker has finished speaking at the end, please judge from the text, if the text constitutes a complete sentence, add then please all period.)

3. Please pay much attention to the **punctuation marks in the middle of audio!!!!!! Don't forget to add them!!! (Please add punctuation marks according to the context, instead of the speaker's pause!!)**

4. Proper nouns must be capitalized!!!

eg: Mike, US, Chinese....

5. Missing words, please transcribe stuttering/model particles according to rules.

(There is no uhm, umm...ect that does not exist in the rules)

6. ok okey okay uniformly written as ok[critical]

7. Better do not write ... be careful of using ; (better to use ,)

8. when you cut off audio parts, please pay MUCH attention to it !!! Because after submitting your work, the cut part can't be restored by reviewers if you cut too much. (but if you can still edit it before you submit and move to the next audio)

9. Please mark punctuation marks according to the actual sentence structure: If the structure of the sentence should have a question sentence, mark the question mark without considering the context. If it is the structure should have a period, then mark a period

10. Do not add punctuation marks between stuttering!!! (Incorrect case below)

☒ non-discard ☐ discard

look I mean, I, I would love to say that I would have the courage and honor to go do that, but I didn't. you know I didn't go do that. so who knows

11. Please use English punctuation marks (half-width, don't use full-width)

☒ non-discard ☐ discard

it's it's it's caught you know, it's kind a book.

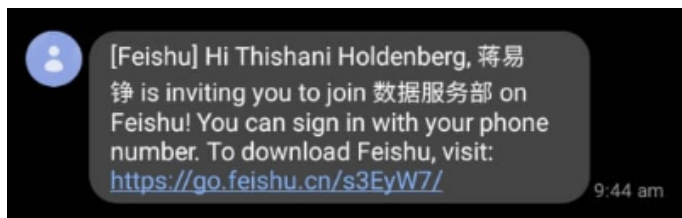
12. If there is another speaker intercepting the main speaker, please do not add punctuation marks before/after the interception. (Incorrect case below)

☒ non-discard ☐ discard

on the time um and yeah, if she kind of has this text in the city vibe. yeah. she's very she's coming from Texas. she also comes military base. uh huh. so she's got this strength she has a lot of hardware kind of jewelry, but she also has this like Navajo Indian feel. we order a lot of jewelry made by like Navajo and the like actual

● How to use the platform

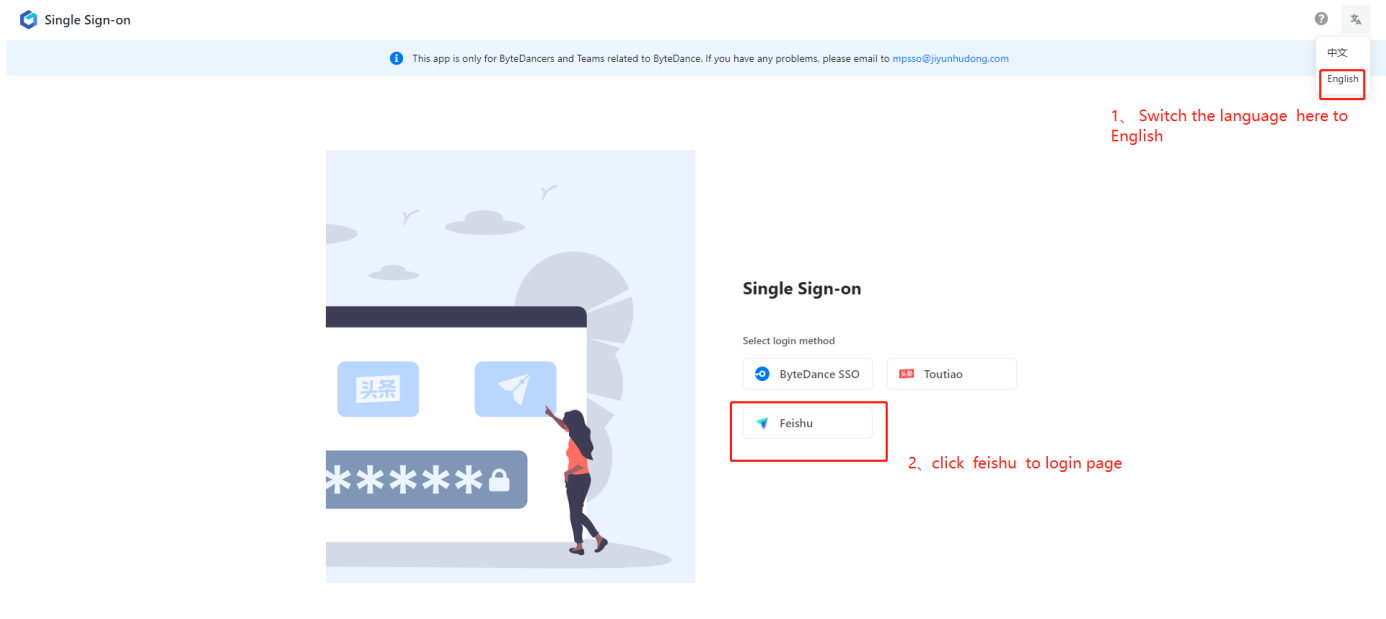
(If you pass the test, you will get the following message and your phone number will have the permission to login the platform (transcription website), you don't need to download the APP. Our colleague will contact you and add you in the group as well)



1. login the website <https://tcs.jiyunhudong.com/> by your phone number and verification code

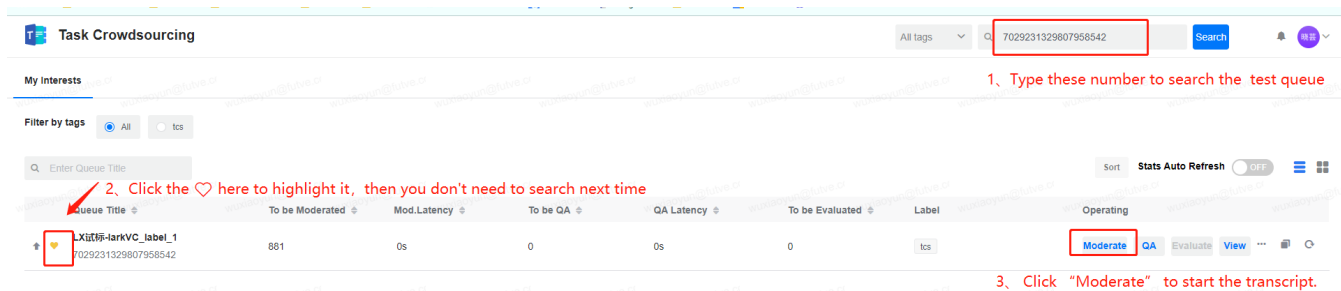
(If you have not passed the test, you will not be able to obtain account permissions)

- 1) Select switch language in the upper right corner
- 2) Choose **Feishu** to login
- 3) Click on **the upper right corner to switch to SMS verification login**
- 4) Type the mobile phone number and tick agree, then you will receive a SMS, type it then you can login



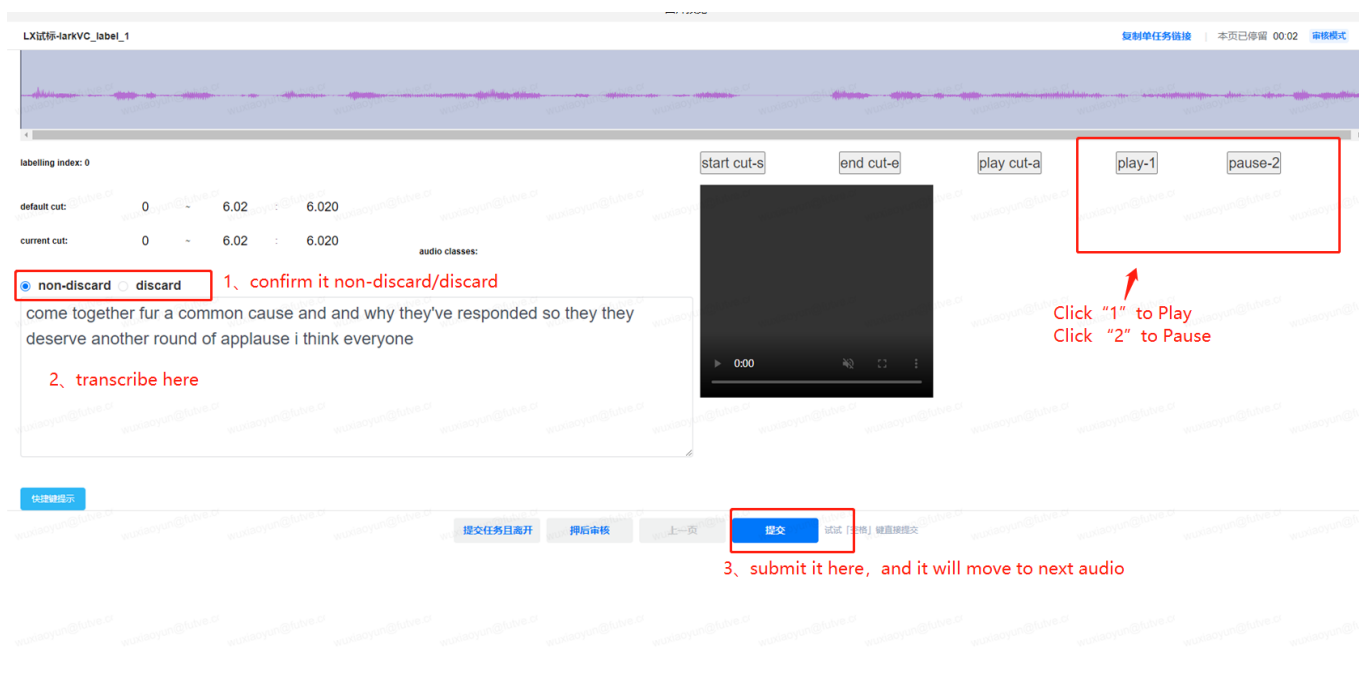
2. Test queue: 7029231329807958542

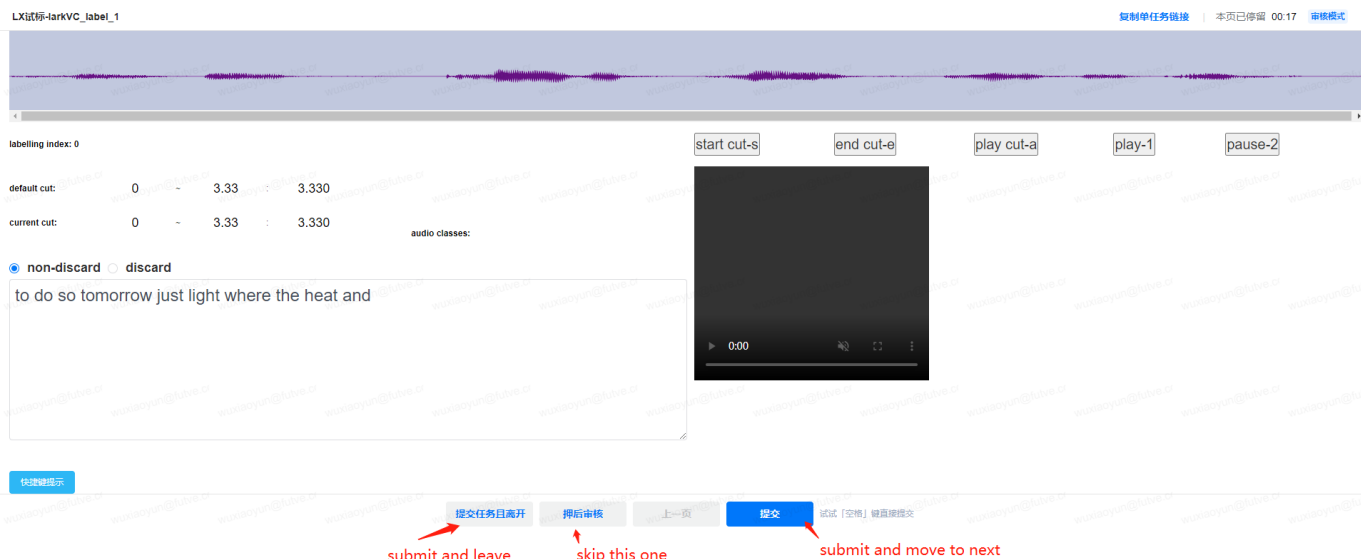
- 1) Type the queue number in the search box (The search box is next to your name)
- 2) Light up the ♥ in front of the queue number to avoid secondary searches
- 3) Select moderate to start transcribing



3. How to use the buttons

- 1) Hit “1” to play the whole audio (1 on your keyboard)
- 2) Hit “2” to stop the audio (2 on your keyboard)
- 3) Hit “A” to play the audio you interrupt (A on your keyboard)
- 4) Hit the “S” to start and “E” for the end if you want to intercept the audio (S E on your keyboard)
- 5) “deferment” here means to “skip”
- 6) Click “submit” if you complete the transcribe, if you found you need to modify after submit, click the “Previous”, but if you skip it, then you can’t click the “previous”



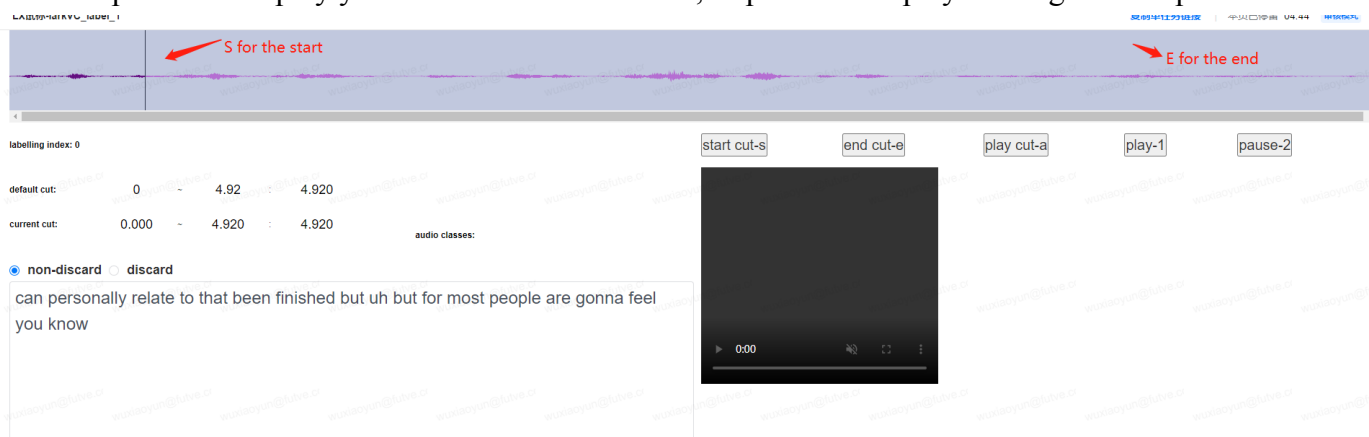


4. How to intercept (cut) audio

Click the mouse where you want to keep as the beginning of the audio, then press "S", and then move the mouse to the end of the audio you want to keep, and press "E", so you can intercept the audio you want to save

(Please pay much attention to cutting off audios, if it's not necessary, do not cut! If you are not sure, please do not cut it! If you cut too much and submit, the incorrect cut can be restored!!!)

You can press "A" to play your reserved audio content, or press "1" to play the original complete audio.



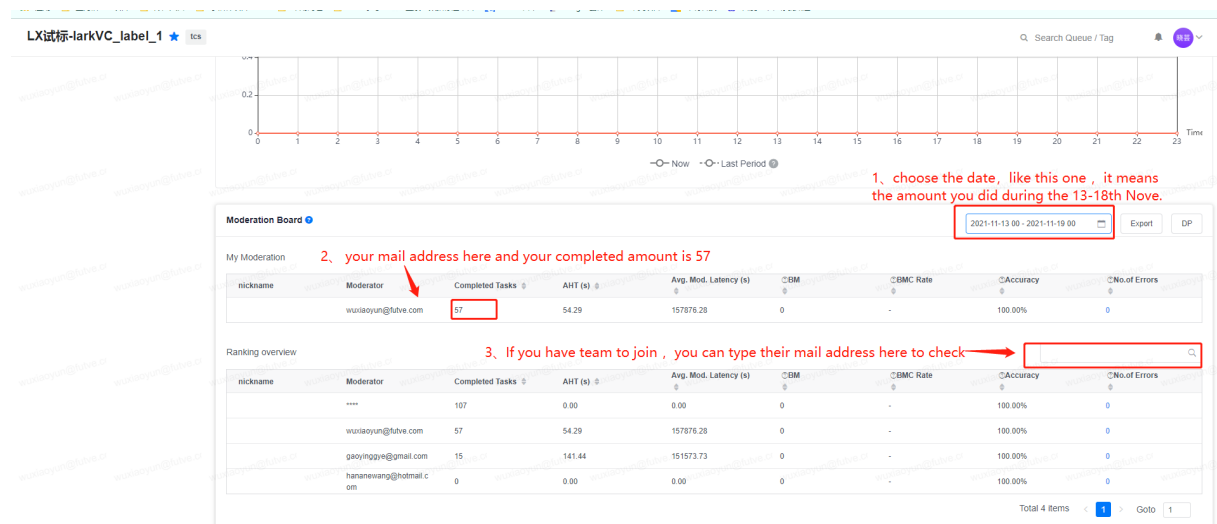
5. How to check the amount you did

Move to the home page and hit "...", and then **"dashboard"**, then you will see the page as the following.

2) Choose the date you worked and type your mail address on **Moderation Board**, then you can get the amount you did

Queue Title	To be Moderated	Mod.Latency	To be QA	QA Latency	To be Evaluated	Label	Operating
LX试标-larkVC_label_1 7029231329807958542	861	0s	0	0s	0	tcs	Moderate QA Evaluate View 1

2 Dashboard



● Transcription Standard

1. Transcribe English audio part only, the non-English audio needs to be marked "discarded"
2. Transcribe all English numbers into **English TEXT only**, Arabic numerals are not allowed;
3. Transcribe according to the actual pronunciation (applicable to slang, numbers, symbols, website addresses, mailboxes...)

For example, punctuation marks should be transcribed into English text:

123@hotmail.com should be transcribed as One two three at hotmail dot com

4. Punctuation marks should be transcribed
5. Only Non-English part / Singing / Overlapping part need to be cut
6. Write abbreviations normally, do not add spaces, such as: ITV, which belongs to English abbreviations, please does not add spaces
7. If the whole audio is singing (non-slow rap or not similar to speech), DISCARD the audio
8. Regarding to overlap:

Scenario 1: The main speaker's voice is loud and clear, and it's not interference from other speakers' voices;

Scenario 2: The main speaker is interfered with by other speakers, and you can't clearly hear the main speaker and other speakers' parts.

If it is Scenario 1, then transcribe the content of the main speaker; if it is Scenario 2, just cut off this part;

If there is no overlap between multiple speakers, please transcribe each part of the speaker's voice in sequence. For example, if the speaking order is A first half sentence + B half sentence + C half sentence + A second half sentence, then the content of the transliteration is: A+B+C+A

9. When a word is cut off (at the end of the audio):

- ① If you don't know clearly what the truncated part is, then just cut it and no need to transcribe;
- ② If the cut sound can have a word fit, please transcribe it, but do not make up word that is not existing.
For example, the person in the audio wants to say semifinal, but at the end of the video, only semi is pronounced and the other part is cut out. This is a word that can be transcribed because semi can be used as an independent word, so you can transcribe *semi* at the end of the sentence.
- ③ If the rest part is not an English word, please cut it off
For example: The speaker just says **organi**, then the speaker changed the topic, and did not continue to say organization. But actually there is no English word **organi** if you transcribe according to the pronunciation, and there is no word close to the pronunciation.

● Transcription Details

Type	Problem	Standard
Writing	Common Mistakes	- No need to transcribe the part if the main speaker is singing. If it is a slow rap which is similar to speech, then it can be transcribed - Stuttering words need to be transcribed according to the actual pronunciation. If possible, transcribe half of the word as well, such as <i>re re restaurant</i>. If the stuttering sound is too short or vague and you can't transcribe it, then just leave it - If multiple people are having a normal conversation without speaking at the same time, please transcribe the conversation in normal order - Transcribe the Slang according to actual pronunciation, <i>gonna/wanna/gotta</i> is not equal to <i>going to/ want to/got to</i> - If the whole conversation is clear and non-overlapping, but the transcriber can only understand 1-2 words, we'll stop cooperation with this transcriber - Do not randomly spell words based on speaker's pronunciation, if you transcribe the audio into a non-searchable English word, then it's incorrect - Do not mark any information unrelated to the audio in the text box
	Words with similar pronunciation	- No need to correct tense and grammar, just try to keep the transcription matches with pronunciation - If the pronunciation is ambiguous, please transcribe according to the context. But if the pronunciation is more biased towards another word, then please transcribe according to the closer pronunciation rather than the context.
	Punctuation	- Please use English punctuation only, Chinese or other punctuation is incorrect - I'm, don't, there's etc. those need to add with ['] indeed - Punctuation marks that are pronounced (e.g. [. -dot] [@-at]) should be transcribed according to the actual pronunciation. The general rule of transcription is that the better it matches the pronunciation, e.g. 50\$, the better it is to mark fifty dollars. - Do not add - hyphen/separator, numbers like 59 fifty nine. No need to add -, if it is stuttering like <i>re re restaurant</i>, then transcribe as <i>re re restaurant</i> - Symbols such as @ is also marked with at if pronounced [at], & symbols such as [Q & A] are marked with and if pronounced with [Q and A]
Writing	Number	- The English URL should be transcribed according to the actual pronunciation - All numbers should be transcribed into English words, with or without dashes in the middle. Such as 59 can be transcribed as fifty nine. - Phone numbers should also be transcribed into English words, for example, 911 should be written as nine one one - Email should be transcribed according to actual pronunciation
	Capsules and lower case	- The beginning of a sentence can be used without capitalisation and is not mandatory. If a problem with the case at the beginning of a sentence is found during inspection, it should not be counted as an error - If a capital letter suddenly appears in a word, e.g. [this] is written as [tHis], it will be counted as an error in careless inspection - Proper nouns such as names of people and places are capitalized or all-capitalized as appropriate; I, etc. need to be capitalized, but ordinary words are not considered wrong in terms of case - Capitalisation is not mandatory, but recommended, but if pronounced [as soon as possible], do not write [ASAP] ; if pronounced directly as an abbreviation such as OKR, then transcribe as OKR - ok okey okay uniformly written as ok【critical】 - Case is not mandatory for book and film titles - Do not use abbreviations in Month or Week, eg. January should be transcribed as January instead of Jan.
	Modal Particle	No strict requirement for modal particles

		- Try to transcribe the number of [ha] in laughter. The incorrect number of [ha] in laughter will not be counted as incorrect in quality inspection, but please try to correct it - For words like [um], just write [um] instead of [umm], see the following list of modal particles.
	Onomatopoeia	- Short coughs or short noises can be left unmarked - Onomatopoeia, e.g. someone simulating the sound of an animal, honking a horn etc. or kissing [mwah] These do not need to be transcribed if it appears during a conversation
	Space	Spelling Letters, for an instance [A B C], need a space in between words

• Modal Particle

ah	eh[er]	hee	jeez	shh	wow
ahah	eww	hey	jesus	uh	ya
bah	gee	hmm	nah	uh-huh	yeah
dah	gosh	ho	oh	uh-oh	yep
darn	haha	huh	oops	um	

When transcribing, please use the standard writing format!!

The general principle of this project is: transcribe what the speaker actually says as much as possible, please refer to the [Transcription Details]

• More details about [Punctuation transcription standard]:

. Full stop (Period)

- The full stop (period) is used when a sentence is completely finished.
- Full stops (period) can also be used for abbreviations of English words, such as *Mrs.*, *Dr.*, *P.S.*, etc. But be careful not to use full stops when the letters of the abbreviation form a word. Such as *IBM*, *DNA*, etc.

? Question Mark

- The question mark should be used in a direct question, not in indirect questions.
For example,
How will you solve the problem? Is the correct usage of question mark.
I wonder how you will solve the problem? is an incorrect usage of question mark. You should mark full stop in this sentence.
- In addition, in polite language, use full stops instead of question marks. Such as Will you please give me a call tomorrow.

! Exclamation mark

1. The exclamation mark is used in the statement of exclamations. Pay attention to the application of the exclamation mark in business writing, because inappropriate use will appear abrupt and unstable.

; Semicolon

1. Semicolons are used to separate independent clauses with equal status. In some cases, the use of a semicolon is more obvious than the use of a period to show the close relationship between clauses. In addition, the semicolon is often used with the connecting adverb **thus, however, therefore** (before these words). Such as *I realize I need exercise; however, I'll lie down first to think about it.*
2. If a comma has been used in the sentence, in order to avoid ambiguity, use a semicolon to separate similar content. Such as: *The employees were Tom Hanks, the manager; Jim White, the engineer; and Dr. Jack Lee.*

NOTE: A complete sentence starts with a capital letter and ends with a period. When writing English, please also pay much attention to other punctuation marks. Please compare the following example sentences:

(Note: The cross-lined part in the below sentence is two different subjects, and the sentence before and after the comma is complete-----Each part represents a complete meaning when taken out separately. Therefore, using a comma violates English grammar.)

Incorrect:

- *It was raining hard, they could not work in the fields.*

Correct:

- *It was raining hard; they could not work in the fields.*
- *It was raining hard. They could not work in the fields.*
- *It was raining so hard that they could not work in the fields.*
- *They could not work in the fields because it was raining hard.*
- *It was raining hard, so they could not work in the fields.*
- *As it was raining hard, they could not work in the fields.*

Incorrect:

- *The essay is poorly organized, there is no central idea.*

Correct:

- *The essay is poorly organized; there is no central idea.*
- *The essay is poorly organized: there is no central idea.*

: Colon

1. A colon is used to introduce or explain what follows, e.g. *This is her plan: go shopping.*
2. A colon is used before a list, especially a vertical one.
 - We transferred three employees to new branches:
Tony Wang to New York City
Mike Jackson to Tokyo
Mark Foster to Paris

When the list is laid out horizontally, the colon is used after a complete sentence, eg. *We need seven people: three students, three engineers and a professor.*

3. A colon is used before a formal reference. eg. *The professor said: "It was horrible."*
4. A colon may also be used after the title of a business or formal letter, eg. *Dear Mr. Lee:* (In American English, a colon is used after a salutation in a letter or speech, whereas in British English a comma is mostly used.)
5. Colons are used to indicate numerical time, eg. *16:45* or *4:45 p.m.*
6. A colon is used between the main heading and the subheading, eg. *Web Directory: World and Non-U.S. Economic Data.*

, Comma

Note: If the main speaker stammers, do not mark a comma, such as: I was I was, don't mark the comma in the middle.

1. A comma is used to separate a series of simple content, eg. *I will go to Shanghai, Beijing and Shenzhen.*
2. A comma is used between multiple adjectives modifying a noun, eg. *a small, fancy bike.*
3. A comma is used to join two longer independent clauses, each with a different subject, eg. *The Grizzlies were out of timeouts, and Miller missed a desperation 3-pointer as time expired.*
4. Commas are used between associated clauses, eg. *Since he's your younger brother, please take care of him.*
5. A comma is used after a longer modifying phrase, eg. *In the middle of the coldest winter on record, the pipes froze.*
6. A comma is used before a directly quoted sentence, e.g. *Mary said, "Let's go fishing."* (Notice: what is said here is not the same as the use of a colon in a direct quote mentioned above. If you are quoting a more formal speech address you should use a colon, in general you use a comma.)
7. If the sentence contains an indirect reference there is NO need for a comma, eg. *Mary said we should go fishing.*
8. Use a comma before a rhetorical question, For instance :*He worked very hard, didn't he?*

The above are the more commonly used punctuation, the following is a list of some of the less commonly used punctuation:

-Hyphenation (hyphens and separators are not written in this project)

() Parentheses

1. Mark the numbers and letters in table order (eg.: (1) , (2) etc.)
2. It is used to indicate an inserted or additional explanatory element. This inserted element may be a word, phrase or sentence. Note, however, that parentheses can weaken emphasis, so use a (em) dash if you want to emphasise the inserted sentence element.

They might take a walk together(remember feet) and see the neighborhood with fresh , new eyes.

" " ' ' Quotation marks

Inverted commas are divided into single quotation marks and double quotation marks. Single quotation marks are only used for direct quotations contained in one direct quotation.

1. Indicates a direct quotation.

The foreigner said to him, "Well, you look like an engineer."

Full stops and commas must be placed inside inverted commas (double and single inverted commas).

Colons and semicolons must be placed outside inverted commas.

Question marks, exclamation marks and dashes are sometimes placed inside inverted commas,sometimes outside them. If the quotation is itself a question or exclamation point or has a dash, the question mark, exclamation point or dash is usually placed inside the inverted commas. Otherwise they are placed outside the inverted commas.

2. Title of a short publication, such as a magazine, newspaper article, short poem, short story and a chapter of a whole book.

Have you read "The Old Man and the Sea"?

3. This indicates that the words used have a particular meaning. In addition, when slang appears in more formal texts, it is also used in inverted commas to indicate a deliberate change in style.

The report contained the "facts" of the case.

4. Used to indicate words that catch the reader's attention, or special words that are unfamiliar to the reader.

“SOS” is a message for help from a ship or aircraft when in danger.

... Ellipses

This apostrophe, whether it appears at the beginning, middle or end of a sentence, indicates the omission of a word.

1. Indicates omission in direct quotation.

Max wrote, “...it changes with every change in the conditions of his material existence....”

The ellipsis and the full stop after a sentence should be written as four black dots in a straight line. The first three black dots indicate an ellipsis and the last one a full stop.

2. Indicates hesitation in speech.

"If that's the way you think...just go back to school," he said.

3. A full line of black dots must be used to indicate the omission of a paragraph or a whole line of words or phrases.

' Apostrophe

1. Forming the noun possessive.

Peter's books.

2. In addition to the contracted form of the verb, it also indicates the omission of one or more letters and numbers.

I've got it. “Yes ,ma’ am ,”the waiter said.

Note: A significant proportion of ellipses are used in spoken language and should not appear in written language. For example: I'd like to (in the written language you should write I would like to)

Differences in punctuation between British English and American English

1. Usage of inverted commas: (1) commas and full stops belonging to quotations are located inside inverted commas in American English, while in British English they are mostly located outside inverted commas; (2) when quotations are then applied inside quotations, double inverted commas are outside single inverted commas inside in American English, while single inverted commas are outside and double inverted commas are inside in British English.
2. Usage of apostrophes: In American English, if the apostrophe happens to be at the end of a sentence, four dots are used, as in I'd like to... .that is... .if you don't mind....
3. Usage of the colon: ① between hours and minutes, American English mostly uses a colon, British English mostly uses a full stop; ② in American English, a colon is used after the address words of a letter or speech, while in British English a comma is mostly used.

(尽量看英文版哈, 中文版更新少)

Lark VC 英文训练集标注(中文版)

● 标注标准

- 只需要标注英语语音对应的文本, 非英文内容需要标注“丢弃”;
- 数字都用英文, 不能使用阿拉伯数字;
- 按实际发音标(适用于俚语、数字、符号、网址、邮箱...), 例如标点符号被念出来(如“dot”)则标注对应的英文;
- 需要标注标点符号;
- 仅非英语部分、唱歌部分、多人对话严重重叠部分, 需要截掉
- 缩略词的话咱们正常写, 中间不要加空格, 比如:ITV, 属于英文缩略词, 是不加空格的
- 整条case如果都是唱歌(非慢速rap类似speech的话), 就丢弃case
- 有关重叠:
场景一:主说话人的声音又大又清晰, 不受其他说话人声音的干扰;
场景二:主说话人受到其他说话人的干扰, 导致主说话人和其他说话人全都互相听不清。
如果是, 场景一, 那就还是直接转写主说话人的内容;如果是场景二, 就直接截掉不要这一段;
补充一点, 多个说话人没有重叠, 依次发音, 需要按序转写。比如说话顺序是A半句+B半句+C半句+A后半句, 那么转写的内容就是:A+B+C+A
- 当单词被截断变成切音(在音频结束时)时:①遇到无法贴合的切音就不用转写;②如果切音可以有单词贴合, 就转写, 但是单词不能瞎拼, 不要无中生有。
举个例子:比如音频里的人想说semifinal,但是视频结束切掉一半只发了semi的音, 这就属于可以转写的词, 因为semi可以作为独立单词。

● 标注细节

类型	问题点	标准
书写	常见问题	*主说话人在唱歌不用标注, 如果是很慢的rap, 跟speech差不多可以标注 *结巴单词需要依照实际发音来标注, 如果可能的话把半个词也标注出来, 如re re restaurant, 小于1s的结巴音如果标不出结巴那半个词不判错 *如果多人正常对话, 没有同时说话情况, 则正常按顺序标注 *俚语, 直接按照发音标, gonna/wanna/gotta不等于 going to/ want to/got to *若全程清晰且非重叠语音, 但【个人能力问题】只能听懂1-2个单词, 请押后, 不可只转写1-2个单词 *不可根据发音瞎拼单词, 若非正常的可搜到的英语单词质检算错 *文本框内不要备注任何和音频无关的信息
	近音词	*不需要对时态和语法做纠错, 只需要尽量保持发音正确 *发音员出现模棱两可的读音, 根据语境可获知发音员想表达哪个单词, 但是其发音更偏向于另一个单词, 这时根据更接近的发音而非语境去写
	符号	*标点符号仅限于英文标点符号, 中文标点符号算错 *I'm, don't, there's等需要加'的一定要加' *发音念出的标点符号(如".-dot" "@-at")需按照实际发音标注。标注的总体原则是和发音越匹配越好, 如念出50\$,则标注fifty dollars *不需要-连字符/分隔符, 数字如59 fifty-nine 无需添加-, 若发音结巴, re re restaurant即可 *@符号如发音则也需要标注at, &符号如Q&A若念出and发音则标注and
书写	数字	*英文网址按照实际读法标注 *数字都标注英语单词, 中间可加/可不加短横线, 质检时不做要求, 如59 fifty-nine *电话号码也按英文字母写, 比如911应写nine one one *邮箱依照发音标注
	大小写	*句首可以不用大写, 不做强制规定。如质检时发现句首大小写的问题不应算错 *单词中若突然出现大写字母如this写成tHis, 属于粗心质检时会算错 *专有名词如人名、地名依照实际情况首字母大写或者全大写, 我I等需要大写, 普通单词大小写问题不算错

		*缩略词大小写不做强制规定, 推荐大写, 但若念出as soon as possible, 不可写ASAP; 若直接念出缩写如OKR, 则标注OKR *ok okey okay统一写成ok【重要】 *书名、电影名大小写不做强制规定 *星期月份不能缩写, 如January必须写成January而不是Jan
	语气词	*语气词不做严格质检要求 *笑声尽量转写出“ha”的个数, 笑声“ha”个数不对质检时不应算错, 但需尽量改正 *um之类的语气词, 写成“um”即可, 不要写成“umm”, 具体写法参考下面语气词表
	拟声词	*短咳嗽或者短噪音可以不用标注 *拟声词如有人模拟动物、鸣笛等声音或亲吻mwah这些如在对话期间, 则无需转写
	空格	*拼读字母, 如“A B C”, 需要在字中间加空格

语气词

ah	eh[er]	hee	jeez	shh	wow
ahah	eww	hey	jesus	uh	ya
bah	gee	hmm	nah	uh-huh	yeah
dah	gosh	ho	oh	uh-oh	yep
darn	haha	huh	oops	um	

标注时, 请采用规范的书写格式！！但是本项目的总原则是:标注是和发音越匹配越好, 详见上文的【标注细节】

针对【标点符号】的书写规范, 具体如下:

. 句点

1. 句点用于当一句话完全结束时。
2. 句点也可以用于英文单词的缩写, 如 Mrs., Dr., P.S. 等。但要注意的是当缩写的字母形成了一个单词的时候就不要使用句点。如 IBM, DNA 等。

? 问号

1. 问号要用于在一个直接的问句, 而不是间接的。如 How will you solve the problem? 是正确的用法, 但用在 I wonder how you will solve the problem?就不对了, 应该使用句点而不是问号。
2. 另外, 在客气的用语中, 也是用句点而不是问号。如 Will you please give me a call tomorrow.

! 感叹号

1. 感叹号用于感叹和惊叹的陈述中, 在商业写作中要注意感叹号的应用, 因为不恰当的使用会显得突兀及不稳重。

; 分号

1. 与中文一样, 分号用于分隔地位平等的独立子句。在某些情况下, 使用分号比使用句点更显出子句之间的紧密联系, 另外分号也经常与连接副词 thus, however, therefore一起使用(放在这些词语之前)。如 I realize I need exercise; however, I'll lie down first to think about it.
2. 在句子中如果已经使用过逗号, 为了避免歧义的产生, 就用分号来分隔相似的内容。如 The employees were Tom Hanks, the manager; Jim White, the engineer; and Dr. Jack Lee.

需要注意的是:一个完整的句子以大写字母开始,以句点结束。写英文时用逗点代替句点、分号、冒号或破折号叫“逗号错”,这正是中国学生所要避免的。请比较下列例句:

误: It was raining hard, they could not work in the fields.

(注意:上面句子中划横线的部分是两个不同的主语,而且逗点前后的句子是完整的-----单独拿出来都能代表一个完整的意思。因此,用逗号违反了英文规定,即一个句子只能有一套主干。)

正: It was raining hard; they could not work in the fields.

It was raining hard. They could not work in the fields.

It was raining so hard that they could not work in the fields.

They could not work in the fields because it was raining hard.

It was raining hard, so they could not work in the fields.

As it was raining hard, they could not work in the fields.

误: The essay is poorly organized, there is no central idea.

正: The essay is poorly organized; there is no central idea.

The essay is poorly organized: there is no central idea.

: 冒号

1. 冒号用于对后面内容的介绍或解释,如 This is her plan: go shopping.

2. 冒号用于名单之前,特别是一个竖排的名单。

We transferred three employees to new branches:

Tony Wang to New York City

Mike Jackson to Tokyo

Mark Foster to Paris

当名单横排的时候,冒号要用在一个完整的句子之后,如 We need seven people: three students, three engineers and a professor.

3. 冒号用于一个正式的引用之前。如 The professor said: “It was horrible.”

4. 冒号也可用于商业或正式信函的称谓后面,如 Dear Mr. Lee:(美国英语中,信件或演说词的称呼语之后用冒号,而在英国英语中多用逗号。)

5. 冒号用于数字时间的表示,如 16:45 或 4:45 p.m.

6. 冒号用于主标题和副标题之间,如 Web Directory: World and Non-U.S. Economic Data

, 逗点

注:如果主说话人结巴时,不标逗号,比如: **I was I was**, 这中间就不要标符号了

1. 逗点用于分隔一系列的简单内容,如 I will go to Shanghai, Beijing and Shenzhen.

2. 逗点用于修饰名词的多个形容词之间,如 a small, fancy bike

3. 逗点用于连接两个较长的独立子句,而且每个句子的主语不同,如 The Grizzlies were out of timeouts, and Miller missed a desperation 3-pointer as time expired.

4. 逗点用于关联的子句之间,如 Since he's your younger brother, please take care of him.

5. 逗点用于一个较长的修饰短语之后,如 In the middle of the coldest winter on record, the pipes froze.

6. 逗点用于直接引用的句子之前,如 Mary said, “Let's go fishing.”(注意:这里说的和上面提及的冒号在直接引语中的使用不一样。如果是引用比较正式的发言讲话就要用冒号,一般情况下就用逗点。)

如果句中含有间接引用就不需要逗点,如 Mary said we should go fishing.

在反问句之前要使用逗点,如:

He worked very hard, didn't he?

以上是比较常用的标点,下面列出一些次常用的标点:

-连字号(本项目中不写连字符和分隔符)

0 圆括弧

1. 标出表顺序的数字和字母(如: (1)、(2)等)

2. 用来表示其中插入的或附加的解释成分。这个插入成分可以是单词、词组或句子。但要注意,括号会削弱强调作用,因此,如果要强调插入的句子成份,则要用破折号。

They might take a walk together(remember feet) and see the neighborhood with fresh, new eyes.

" ' 双单引号

引号分单引号(single quotation marks)和双引号(double quotation marks)。单引号只用在直接引语中所含的另一个直接引语上。

1. 表示直接引语。

The foreigner said to him, "Well, you look like an engineer."

句号和逗号必须置于引号(双引号和单引号)之内。

冒号与分号必置于引号外。

问号、感叹号和破折号有时置于引号之内,有时置于外号之外。如果所引内容本身是疑问句或感叹句或带有破折号,问号、感叹号或破折号一般放在引号之内。否则放在引号之外。

2. 标明短篇出版物的标题,诸如杂志、报纸上的文章、短诗、短篇故事和整部书的某一章节。

Have you read "The Old Man and the Sea"?

3. 表示所用的词语具有特殊意义。另外,当俚语出现在较正式的文章中,也用引号引起来,以表示文风的有意转变。

The report contained the "facts" of the case.

4. 用于表示引起读者注意的词语,或读者不熟悉特殊词语。

"SOS" is a message for help from a ship or aircraft when in danger.

... 省略号

此省略号无论出现在句首、句中、还是句尾,都是表示单词的省略。

1. 表示直接引语中的省略

Max wrote, "...it changes with every change in the conditions of his material existence..."

句后的省略号和句号的写法应是在一条直线上的四个黑点。前三个黑点表示省略号,后一个黑点表句号。

2. 表示说话中的犹豫或迟疑

"If that's the way you think...just go back to school," he said.

3. 表示段落或整行词句的省略,须使用一整行黑点。

' 撇号或省字号

1. 构成名词所有格

Peter's books.

2. 除表动词的紧缩形式外,还表一个或几个字母和数字的省略。

I've got it. "Yes, ma'am," the waiter said.

注意:有相当大一部分省略词是口语中的用法,不宜出现在书面语中。例如:I'd like to(在书面语中要写作I would like to)

英国英语和美国英语标点符号的差异

1. 引号的用法:①属于引语的逗号、句号在美国英语中位于引号内,而在英国英语中多位于引号外;②引语内再套用引语时,美国英语中双引号在外单引号在内,而英国英语中的单引号在外、双引号在内。
2. 省略号的用法:在美国英语中,如果省略号恰好在句尾,就用四个点,如I'd like to...that is...if you don't mind....
3. 冒号的用法:①在小时与分钟之间,美国英语多用冒号,英国英语多用句号;②美国英语中,信件或演说词的称呼语之后用冒号,而在英国英语中多用逗号。

问题:

1. 平台加载慢怎么解决

平台这个可能目前解决会有点吃力, hhh

2. 多人说话重叠,需要丢弃还是如何处理

重叠的话需要判断:

如果是A人声和B人声重叠了,一种场景是主说话人又大又清晰,不受另一人的干扰;第二种场景是A人声和主说话人音量持平,清晰度两方都收到干扰听不清。这两种场景的处理方式分别是:场景一可以保留,转写主说话人,场景二截掉不要。

3. 音频中前后切音无法识别的单词如何处理?

前后的切音,无对应的英文单词,就不转写(不需要截掉)

4. 主说话人唱歌(非可转写的rap)是否做如下处理:1、若唱歌只是音频的一部分,则忽略唱歌部分,转写其余可转写部分;2、整个音频都是说话人在唱歌,直接丢弃。
唱歌部分截掉不要;整个case如果全都是唱歌(非慢rap)就丢弃