

Transcription

Knowing three things about transcriptions may vastly improve the quality of your life (as far as transcriptions go). First, phonetic symbols represent approximate ranges and not absolute values; second, transcribing two segments is exponentially harder than transcribing one segment; third, good transcription is like getting to Carnegie Hall.¹

The meaning of symbols

The symbols of standard transcriptional schemes such as the IPA are widely presumed to have well-defined meanings: after all, that is the point of a standard system of transcription, that symbols have agreed-upon values: $\pi = 3.1415926_{\text{blahblah...}}$ and $[y] = \text{“high front rounded vowel”}$. The possibilities for positioning the tongue and lips in the production of a vowel are infinite, and infinitely many physical configurations result in infinitely many output signals. Yet there are not just not an infinity of phonetic symbols, there are actually very few phonetic symbols. So really, the values of phonetic symbols are quite approximate.

The underlying assumption in phonetic transcription is that there should be enough symbols in an adequate inventory, so that any phonemic distinction encountered in some language can be represented, but it is not necessary to represent every phonetically possible distinction that exists across human languages with a symbol. Because there are languages that phonemically contrast $[p]$ and $[b]$, or $[p]$ and $[p^h]$, then $[p]$, $[b]$ and $[p^h]$ should all be represented in a theory of transcription. But this does not mean that every case of phonetically transcribed $[p]$, or $[p^h]$, is exactly the same across languages.

As an example, Korean and Hindi phonemically contrast $/p/$ and $/p^h/$; there is also a non-contrastive phonetic difference between $[p]$ and $[p^h]$ in English. If you carefully listen to these sounds in Hindi, English and Korean, you can hear that the physical value of Korean phonetic $[p]$ is not the same as it is in English (appearing after $[s]$ or in the context $\acute{v} _ \check{v}$) — Korean $[p]$ is more aspirated (has a greater VOT lag) than supposed $[p]$ of Hindi or English. The aspirated allophone $[p^h]$ of English has less aspiration than Korean $[p^h]$ (which is unusual in the languages of the world in having very heavy aspiration on their aspirated stops).

Vowels are clear examples of the relativity of phonetic symbols. There is a noticeable phonetic difference between the pronunciation of the vowel $[i]$ in American English and in German — in German $[i]$ is much higher and further front. The vowel transcribed as $[a]$ in Norwegian and (Fenno-)Swedish are phonetically different even though they are usually transcribed by the same symbol, being further back in Norwegian than in Swedish. Many African languages have a contrast between the vowels $[i,i]$ and $[u,u]$: finding even two languages where the fine-grained phonetic values of these vowels is identical is quite hard.

This doesn't mean that phonetic transcription is hopeless: it actually makes the job a bit easier, because it means that the demands of precision in phonetic transcription are not as onerous as would have been the case if transcriptional symbols had very precise physical values and

¹ A visiting violinist from the Boston Pops Orchestra is appearing in a concert at Carnegie Hall. He gets lost and doesn't know the New York subway system, so he asks an elderly gentleman in the subway “Excuse me, can you tell me how to get to Carnegie Hall?”. The old man replies “Practice, practice”.

had to be carefully matched to actual phonetic qualities. This doesn't mean that [biljøn] is just as accurate a transcription as [biljøn] or [buljøn] (I will leave it up to you to decide what it's a transcription of, or which transcription is better). It means is that there are limits to how accurately phonetic properties can be identified just by listening. After a certain point, precision is achieved by measuring the physical properties of sounds.

The right way to pin down the *exact* phonetic value of a sound is to measure its acoustic properties, e.g. determine its mean duration in some context, measure F1, F2 and F3, check F₀, and so on. It suffices to say that in both Norwegian and Swedish, there is a low back unrounded vowel, even though the objects given the same general description in the languages are not literally the same. If you want to get more specific about that exact difference, find the mean formant values from a sample of Norwegian speakers and Swedish speakers; go somewhere and take x-ray pictures to see exactly what the tongue is doing.

Contextual influence

Usually, sounds don't just stand alone — there are other sounds before and after them. This makes transcription harder in two ways. First, it's somewhat easy to identify an isolated sound because you only have to pay attention to one thing, whereas when you've got a sound mixed in with a bunch of other sounds, you have to pay attention to a lot of different things. Not only do you have to identify many more things in phonetic [q^hrawiʃtnuχɑðo] than in [a], but you have to remember the order of each of them, which is a huge drain on your cognitive resources, especially when you're trying to write the word down after someone has pronounced it. Even worse, the duration allotted to [a] in the utterance [a] is vastly greater than it is in [q^hrawiʃtnuχɑðo], so again more segments makes it harder to parse. At a certain point, there is just too much stuff to hold in your head and it went by too quickly, so that all you get is a blur of sound. What do you do??

Write down what you did hear (a vowel or two and a few consonants), then ask the speaker to say it again, and listen to just the first part of the utterance. After you have heard the initial handful of segments a few times, the repeated auditory experience will render that part sufficiently familiar that you will recognise it automatically, and that portion of the word won't tax your auditory computational ability, so that you can focus your attention on the latter parts of the word. Keep listening and focusing attention on just a part of the utterance (once that part is solved, move on to the next part of the utterance).

The second problem is that the phonetic value of sounds can be “insignificantly” different according to what's in the neighborhood. An example is that a vowel might sound slightly different before a nasal consonant than before an oral consonant. You may have initial transcriptions such as [dæs] and [dæm], but find upon rechecking that the vowels sound a little different, and you're wondering whether the right transcription should be [dæs] versus [dam], or [dæs] vs. [dæm]. The answer is that it depends in part on what kind of transcription you're making — phonetic or phonemic. Once you know what the phonemes of the language are, the question could be answered more easily by comparing supposed [dæm] with some known form such as [tan] or [kæm]. If the vowels [a], [æ] and [ɛ] phonemically contrast before nasal consonants, the question is whether in the word you're looking at the vowel is more like the quality found in words known to have the phonemes /a/, versus /æ/, versus /ɛ/. To decide that, you have to *know* that these vowels are all phonemes in the language, and you have to have examples which you

are certain have each of these phonemes. That is possible only after you have done a fair amount of analysis and you have a mental image of what these phonemes generally sound like.

It's much harder making this decision when you don't know what the phonemes of the language are. A natural inclination is to economize on symbols, and to ignore little differences that you think you hear, like [ɪ] versus [i], or [æ] vs. [a]. Since there are huge numbers of phonetic varieties of sounds arising from basic physical constraints on articulation, you can't expect [dæm] and [dæn] to have exactly the same phonetic vowel. If you had a good enough ear, you could probably detect 300 subtle variants of [æ] (Henry Higgins could, according to the movie). That would be an obsessive number of distinctions to make, and you would probably be right to disregard some of those supposed phonetic differences. On the other hand, you should not automatically assume that you don't have [ɛ], [æ] and [a] in the language, even if it seems like a small distinction to make. Since the difference [ɛ] ~ [æ] ~ [a] is a difference that languages can make contrastively, you should transcribe those vowels as different, if you can reliably hear them as phonetically distinct.

More often, you may hear a word as [dɛɪ] one time and as [dɪɪ] the next time. This kind of contradiction can usually be resolved by having the speaker repeat the word a few more times and going with whatever impression you get 2/3 of the time (given at least 4 samples). This implies, of course, that if you get contradictory transcriptions of a word many times, you need to come back to that word, and also consider the possibility that there actually are two pronunciations (a fact that is often revealed by the speaker noticing your puzzlement and saying "We can pronounce that as [dɛɪ] or [dɪɪ]").

It is okay (and often desirable) to switch to a phonemic transcription later in the analysis, once you have enough data to be certain that the difference is non-contrastive, and once you have a sufficiently large corpus of data that justifies assuming that the phonetic difference between trilled [r] and tapped [ɾ] can be predicted (if it can). We will phonemicize the phonetic distinctions later; in the first few weeks of transcriptions where you are first becoming familiar with the sounds of the language, you should err on the side of writing more information, and making more distinctions than are perhaps strictly needed. The basic rule of thumb should be, if you can reliably hear a difference, write it. You can always remove superfluous information once you're certain that there is a rule that allows you to predict it. You can't supply missing information that you ignored because you thought it wasn't important.²

The experience factor (external variety)

Like getting to Carnegie Hall, the ability to associate phonetic symbols and acoustic events takes practice. The abstract knowledge that IPA [ɑ] is a low back unrounded vowel and [ɒ] is a low back rounded vowel isn't a lot of practical use if you have no idea what such vowels sound like. What you need is experience with the sounds. Where do you get that experience?

² There are pretty much no guarantees in life. After working on Kikerewe for a number of years, it was really clear to me that the phoneme /b/ was pronounced as [b] after a nasal and as [β] otherwise. Quite reasonably, I therefore stopped writing [β] in transcriptions. Then one day, I ran into the verb [kubomoka] "to knock down a house", pronounced with [b] and not [β]. Apparently, there are a very small number of words which have a stop [b] — this necessitated reviewing all of the lexical data to see if I had missed any other cases of [b] not after a nasal (I hadn't).

Alas, that is largely done by apostolic tradition, by being carefully trained by Daniel Jones (who “defined” the cardinal vowels, the standard reference points for vowel symbols). However, Jones is quite dead, and so apparently are all of his students, and his grandstudents are retired or on the verge of retiring. Most practicing fieldworkers have gained their knowledge of phonetic values of symbols via the technique of by guess and by gosh, and relatively few have received formal training from a Jones descendant. Typically when you are first introduced to phonetic symbols, the instructor pronounced them for you, and I hope you were paying close attention and can remember what they sounded like.

Fortunately, if you are in need of a little refresher (I, for example, constantly need help with those blasted central vowels), there are examples available online. Peter Ladefoged has made available some “phonetic reference” recordings, in the form of his expert renditions of the symbols of the IPA. This is an extremely useful resource. His book *Consonants and Vowels* has an accompanying CD, the contents of which are also available at:

<http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants>

The structure of the website sometimes changes, so check around for something like “index” or even “home”. The IPA chart is currently at:

<http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/course/chapter1/chapter1.html>

and you can click on vowels and consonants to hear him pronounce them.³

The experience factor (internal variety)

Listening to a bunch of canned “standard samples” of transcriptional symbols won’t address the practical problem of how to write down data from an informant. What you need is a set of experiences in the target language, which will allow you to differentiate things which are different, and integrate things which are the same. Here are some suggestions about how to do this. Start with a distinction that you’re wondering about, for example are there both [e] and [ɛ] in the language. The first thing to do is gather together all of the words which you think have either vowel. Review those words (re-elicite them), and listen to see if some sound like they have [e] and some have [ɛ]. You may need to do this twice, depending on whether the distinction becomes crystal clear when you recheck, or, conversely, obviously non-existent. Whether you are convinced by a single re-checking will depend on the number of examples that you have, and the extent to which the vowels appear in comparable phonetic contexts. When you organise a word-list for re-elicitation for the purpose of re-checking vowel qualities, it may also be useful to write down your examples without carrying over the distinction that you are testing which you marked in your earlier notes.⁴

³ His performances should be seen as “reference pronunciations”, not actual sounds of a language pronounced by a speaker of the language. I’ve never hear a language with pharyngeals that sound the way Ladefoged is pronouncing them, and there is no [ʃ]-like sound in his recordings that actually sounds like English [ʃ].

⁴ The reason is that if you organise a list of words that you want to recheck — [pes], [lɛk], [yet], [pɛr], [ken] — and write down your target words on the elicitation list as [pes] and [lɛk], then if you re-listen and agree with your previous transcription, it’s possible that you are being influenced by what you wrote before, and the apparent re-

The size of your corpus and the variety of contexts represented influences your ability to correctly categorize the sounds of the language. With only three examples of supposed [e] and three of supposed [ɛ], listening to those examples 100 times won't make matters clearer. Having 100 examples of those vowels is what you need — of course you also want to make sure that you don't have most examples of [e] coming before [ʃ] and [ʒ]. This is why gathering a largish lexicon is the first step in doing fieldwork.

Technology can be usefully thrown at the problem of figuring out if sets of segments are the same or different. Using the case of [e] versus [ɛ], surrounding segments often interfere with segment perception. In SA, the impact of of this problem can be reduced by selecting just the segment of interest and playing that part of the word. By putting the cursors around the vowel [e] or [ɛ] and eliminating the distraction of the surrounding segments, you can focus on the phonetic value of just the vowel. This technique tends to strengthen the impression that the sounds you are comparing are phonetically different, which is not entirely a good thing. Any two tokens of exactly the same word will be physically different, and that difference may be large enough to perceive. Needless to say, the vowels [ɛ] in [dɛt] and [dɛp] will, when removed from their consonantal context, sound different. The technique of segment-excising should therefore be used cautiously because it tends to increase the number of phonetic units. It becomes more useful when you look at a relatively large collection of examples, say a dozen or two of supposed [ɛ]'s, and watch for a clear two-way distinction. If you hear twelve different types of [ɛ] vowel which are equally different, you have good ears (if your performance is replicable), but you probably aren't hearing a distinction that needs to be recorded. If comparison of a dozen excised vowels leads to a clear partition into two types (or three), then you should consider the possibility that there is a distinction needing to be captured in your transcriptions.

It can also be a problem that it is hard to judge the quality of sounds because they are quite short. This problem can be addressed in SA by playing the word (or fraction of a word) at a reduced speed, perhaps 50% speed. This introduces a small distortion in the quality of the sound, but reductions between 40% and 80% can be useful for clarifying what the sound is. Finally, differences can be clarified by doing side-by-side comparisons of words, even when the words were not said next to each other in the original recording, by using the editing ability of SA to create a file with the two words next to each other. You could collect all of the words with the vowel [e] into one file and play those words; when you find two words that seem to have different vowels, you can cut-and-paste the words next to each other and see whether they still sound different when they are next to each other.

Of course, the best technique for checking on the values of sounds is to re-elicite them, to get a new auditory impression. The most valuable experience with the phonetics of a language is being exposed to many examples of pronunciation in the language. Finally, a propos the cultural transmission of phonetic experience, bring your puzzles to class and ask “what the heck is that sound?”. You may not be the only person wondering.

hearing isn't genuine re-hearing. If you write these examples as [pes] and [lek] on your elicitation list, and upon re-checking you hear them as [pes] and [lek], that makes a stronger case that you actually heard a distinction and weren't just influenced by a previous idea that you had. This isn't foolproof, of course, because you might also simply have remembered “I thought that was [lek] before”.