All the Sounds of All the World's Languages

Melissa A. Epstein
melissae@ucla.edu

Infants are born with the ability to discriminate between sounds that are phonemically contrastive in any language. As they grow older, though, they learn to discriminate only between those sounds that are in the language (or languages) spoken by their caretakers. Assuming that for a child to learn a sound he or she would have to learn a language containing that sound, how many languages would a child have to learn in order to be able to discriminate between all the sounds of all the worlds languages?

This question requires both a definition of the term "language" and the term "sound". *Ethnologue* catalogues more than 6,700 languages using mutual intelligibility as the criteria for distinguishing between languages and dialects. Traditional nomenclature is often an accident of history. For example, Chinese "dialects" are less mutually intelligible than the Scandinavian "languages" (Comrie 1987). The International Phonetic Association in its 1993 edition of the International Phonetic Alphabet gives symbols for 112 sounds along with an assortment of suprasegmentals and diacritics which are used to further define the sounds, For example, the alveolar stop, /t/, can be further categorized as labialized, /t^{w}/, lengthened, /t:/, breathy, /t̤/, etc. Gaps are left in the list of symbols for sounds that are felt to be empty by chance, such as a bilabial tap; other regions are blacked out because the sound is judged to be impossible, such as a velar trill. According to Ladefoged and Maddieson (1996), new sounds are likely to be created by "re-arrangements of properties of sounds that have been previously observed in linguistic usage" and are unlikely to be innovations in the use of articulators, such as whistles or tongue wags.

Unfortunately, there is no phonetic database of all 6,700 plus languages in the world. The UCLA Phonetic Segment Inventory Database (UPSID), however, does contain phonological inventories for 317 genetically distinct languages. Although the languages were not chosen for variety within their phonological inventory, the selection does cover a wide range of the languages in the world. The phonological inventory for each language in UPSID was determined by listing the sounds which were phonemically contrastive in each language. In other words, each sound is not only phonetically distinct, but also not predictable from context. For example, in Spanish (and many other languages) there will be a velar nasal, "ŋ", before velar stops. Since this sound does not contrast with the non-velar nasal it will not be listed in the phonological inventory in UPSID. Furthermore, tone, although contrastive in languages such as Mandarin Chinese, was considered suprasegmental, and not included in the inventory. All 770 actually occurring phonological segments were classified according to combinations of 58 phonetic attributes. The "smallest" languages were Rotokas and Mura with eleven contrastive sounds each. The largest was !Xũ with 141 contrastive sounds. Thus, according to UPSID's sample, there are 770 sounds in the world's languages; now our question is how many languages would a child need to learn in order to master these 770 sounds?

To answer this, the database was analyzed by first finding which sounds were "unique" – segments which were represented by only one sound in the UPSID database. One hundred and sixteen of the languages in UPSID have at least one sound that no other language in the database has. In fact, 359 of the 770 sounds in the database (47%) are

unique, i.e. occur in only one language. The remaining sounds were then scanned to find if they were used by one of the languages having a unique sound. Three of the sounds in the database had no principled way to choose a "representative language" – they were not spoken in any of the 116 languages with unique sounds, nor did they share any languages between them. Two other sounds were not spoken in any of the languages with unique sounds, but were shared by a common language. Assuming that in order to learn a sound a child would have to learn a language which contains it, a child would then have to learn 120 languages in order to learn the 770 sounds in the UPSID database.

Sounds are not distributed evenly among the languages in UPSID. If the 120 languages were to equally account for all the sounds, they would each have an inventory of 6 to 7 sounds. Instead, according to Maddieson (1994), languages tend towards 20 to 27 sounds in their inventories. Furthermore, the unique sounds were not distributed evenly. If they were, one would expect the 116 languages containing unique sounds to have approximately 3 unique sounds each. Instead, !Xũ contributed 70 sounds, Irish 14, Nama 10, Arabic 10 (29% of the unique sounds from these four languages alone), French 1, etc. The number of unique sounds was also related to the size of the language's phonemic inventory. Although languages with only 20-30 phonological segments could contribute 1-3 unique sounds, for the most part, languages that contributed 4 unique sounds had at least 30 segments, and languages that contributed 5 or more unique sounds had more than 40 segments.[1]

Finally, this survey does not indicate how the number of unique sounds would change if more languages were added to the UPSID database. On one level, we would think there would be a decrease in the percentage of unique sounds because it is likely that many of these unique sounds would be part of the additional languages. On the other hand, adding more languages is also likely to add more sounds, and also unique sounds, to the inventory. We can get some idea of this by comparing our results so far with a larger sample. In the computerized version of UPSID, 451 languages are surveyed yielding an approximately 900 phonological segments. Of these, 421 are unique. This is the same 47% as the smaller database, despite the addition of 134 languages and 130 phonological segments (in fact, this is approximately one new unique sound per additional language!).

An alternate task would be for a child to learn to speak with the greates number of people in the world. If a child were to learn only the ten most common languages (those with the largest numbers of native speakers), how many sounds would he or she learn through natural exposure? The ten most populous languages are:

[1] There were three exceptions to this trend. Tamang and Aranda, with 29 and 30 phonological segments respectively, each contributed 6 unique sounds. Russian, with 38 phonological segments, contributed 8 unique sounds.

Language	Population
Mandarin	885,000,000
English[2]	322,000,000
Spanish	266,000,000
Bengali	189,000,000
Hindi	182,000,000
Portuguese	170,000,000
Russian	170,000,000
Japanese	125,000,000
German	98,000,000
Wu (Chungchow)	77,175,000
Total	2,484,175,000

Learning these languages would allow a child to communicate with roughly half of the world's population in their native tongue. These languages encompass 192 different sounds (76 vowels and 116 consonants and semivowels), using 30 manners and 20 places of articulation. However, if a child learned these ten languages, he or she would have learned only 25% of the sounds in the UPSID database; although this number is impressive, considering the child has learned fewer than 01% of the world's languages.

In conclusion, it is perhaps an impossible task to learn "all the sounds of all the world's languages" through natural exposure. According to both the UPSID databases, approximately 47% of the sounds are unique, and this number is stable with an increase in languages in the survey. Therefore, the 120 languages originally proposed would merely be a starting point for an enormous endeavor.

References

Comrie, B. 1987. Introduction. In B. Comrie (ed.), *The World's Major Languages*. Croom Helm, London.

Crabtree, M. and Powers, J. (eds.). 1991. *Language Files*, 5th ed. Ohio State University Press, Columbus.

Grimes, B. F. (ed.). 1996. *Ethnologue: Languages of the World*, 13th ed. http://www.sil.org/ethnologue

Ladefoged, P. 1993. *A Course in Phonetics*, 3rd ed. Harcourt Brace College Publishers, Fort Worth.

Ladefoged, P. and Maddieson, I. 1996. *The Sounds of the World's Languages*. Blackwell Publishers, Oxford.

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge University Press, Cambridge.

[2] Neither English nor Portuguese have phonological descriptions in UPSID. Combined, they had nine vowels (but no consonants) that were not described in UPSID. However, the number and types of sounds in English and Portugues is somewhat uncertain because the phonological transcriptions were not standardized to UPSID.

Parkinson, S. 1987. Portuguese. In B. Comrie (ed.), *The World's Major Languages*. Croom Helm, London.