

IGBO CONSONANT CONFUSION MATRIX: ISSUES ON PHONETIC SIMILARITY

*Cecilia Amaoge Eme
Department of Linguistics
Nnamdi Azikiwe University
Awka, Anambra State, Nigeria
ca.eme@unizik.edu.ng
08035611899*

*Linda Chinelo Nkamigbo
Department of Linguistics
Nnamdi Azikiwe University
Awka, Anambra State, Nigeria
lc.nkamigbo@unizik.edu.ng
07036415787*

*Benjamin Ifeanyi Mmadike
Department of Linguistics
Nnamdi Azikiwe University
Awka, Anambra State, Nigeria
b.mmdike@unizik.edu.ng
08065029941*

Abstract

This paper is on Igbo consonant confusion experiment aimed at producing an Igbo consonant confusion matrix. One hundred Igbo-speaking subjects were used for the experiment at the ratio of 1:1 for the control and experimental groups. Whereas the 28 consonants of standard Igbo were presented to the control group in VCVC frame in a noiseless environment, same material was presented to the experimental group under the condition of masking noise. The experiment and matrix reveal that the greatest confused consonant is [j]. All the consultants in the experimental group confused this consonant with another sound; hence it bears 0% in the confusion matrix. Conversely, the least confused of all the Igbo consonants is [kw]; followed by [f]; and [k m z] in that order. In the confusion matrix, they scored 94%, 92% and 88% respectively; only 3 people confused [kw] but 4 people confused [f] while 6 consultants confused each of [k m z]. The segments [j] and [Y] have the closest perceptual similarity and therefore often confused perceptually, scoring as high as 56% across themselves in the confusion matrix. This result is surprising since they are not all that phonetically similar. However, we see the phonetic similarity between these confused sounds shown in the matrix - [ɟ] is confused with [t], as much as [l] is confused with [r]; both scoring 40% in the matrix. Also, [kp] is confused with [p] as much as [n] is confused with [l]; each of the pairs scored 28% in the matrix. On the other hand, many consonants were never confused for each other. For instance, none of the fricatives or approximants was ever confused for the voiceless bilabial plosive [p]. We discovered, however, that some consultants could have confused some segments not necessarily because the segments are phonetically similar but because they tended to consider meaningfulness of syllables rather than the segment they actually perceived. This calls for further investigation on segment confusion in tone languages.

IGBO CONSONANT CONFUSION MATRIX: ISSUES ON PHONETIC SIMILARITY-----Cecilia Amaoge Eme, Linda Chinelo Nkamigbo and Benjamin Ifeanyi Mmadike

Background

When there is a close phonetic similarity between two or more consonants, there is usually a close similarity in the perceptual quality of the segments. Their close perceptual quality leads to their being confused when they are not clearly perceived, such as when the stimuli are given under conditions of masking noise. This is to say that one can find out how phonetically and perceptually related certain consonants are by matching their level of confusion; better in a confusion matrix (cf. Laver, 1994; Clark, Yallop & Fletcher, 2007). The phonetic similarity of consonants causes their auditory quality to be similar. This is a language universal phenomenon. Thus, when the phonetic quality of two or more consonants is similar, the perceptual quality of the segments is very likely going to be close. Apart from using the phonetic similarity of segments to determine the degree to which they are articulatorily, acoustically or perceptually similar, phonetic similarity or its inverse, phonetic distance, could also be used, according to Laver (1994:391-2),

... to locate all segment-types in multidimensional auditory space in such a way that a statement could be made of the nearest neighbours of any given segment-type, together with a description of their respective positions and mutual distances. The greater the phonetic similarity between any two segment-types, the greater their proximity in this multidimensional auditory space.

On this basis, Laver (1994:392) displays a matrix of English consonants showing their phonetic similarity. He uses the inverse notion of phonetic distance such that, using a percentile scale, every two segments with zero means that they are identical whereas those with 100 means that they do not have any common phonetic feature. From the matrix, one sees that, for example, the segment [p] is more phonetically similar to [k] than to [t]; as the phonetic distance between [p] and [k] is 15% while it is 25% for [t]. The segment [t] is, however, more phonetically similar to [k] than to [b] because the distance between [t] and [k] is 25% while it is 45% for [t] and [b].

A confusion matrix of consonants is presented in Clark, Yallop & Fletcher (2007:315). The phonetic similarity of the segments is specified in this matrix in form of the number of subjects that failed to perceive a specified consonant correctly and the segment they confused the specified consonant with. The consonants, presented in a CV frame, appear in the rows of the matrix while the actual segments perceived appear in columns. The number of subjects that perceived the actual consonant produced is used to calculate the percentage confusion of the consonants. The segments, presented in a masked natural speech, are all English phonemes except [μ]. All the 15 subjects perceived [v] and [n] correctly, hence the two segments scored 100% each. This is unlike the segments [b], [g] and [□] that indicated that only one of the subjects perceived the sounds accurately.

According to Mbah and Onah (2015:241), one can account for the changes that occur in the learning of L2 by using Speech Learning Model (SLM), which “generates predictions relating to the accuracy with which a highly experienced learner produces and perceives L2 sounds.” Whenever the language learner detects differences between a speech sound in his L2 and its closest L1 sound, he creates new phonetic categories. For SLM hypothesis, the differences are easily detected if the perceived phonetic distance among these segments is high, which ultimately leads to setting up new phonetic categories.

Preparing confusion matrices for languages based on the phonetic similarity (or its inverse, phonetic distance) of the segments is a necessary step towards a better understanding of the languages in the area of segment-type identification either in natural speech or in speech recognition systems. The uses of a confusion matrix for a language are explained by Laver (1994:393). According to him, a confusion matrix “predicts the perceptual confusions that human subjects are likely to make when asked to identify segment-types presented as stimuli in perceptual labeling experiment.” This is because the greater the phonetic similarity among two segments the more likely the listeners would confuse the two segments. A confusion matrix can, therefore, play a predictive role during experiments that concern consonant confusion. Moreover,

A confusion matrix [is a good instrument] for use in the design of automatic speech recognition systems. When these computer-based systems attain a

performance which is at all comparable to human performance, then the greatest number of false identifications made should be those of the acoustically nearest neighbours of the true segments... it is hence important that the mistakes made by such systems should not be random, but related in a principled way to the properties of the segments concerned (Laver, 1994:393).

We are, however, aware that there are many other issues that must come into the design of automatic speech recognition machines for them to be workable, apart from issues arising from phonetic similarity (cf. Ladefoged, 1962; Lieberman & Blumstein, 1988; Nolan, 1992; Coppens, 1997; Eme, 2007; O'Callaghan, 2014). For instance, in the words of O'Callaghan (2014), "The articulatory consequences of phonemic context change the acoustic features of the signal and compound attempts to map phonemes to signals (which presents the difficulty for artificial speech production and recognition)."

Most of the experiments on phonetic similarity and discussions on consonant confusion matrices have been based on languages that are not tonal; and using the basic syllable structure CV frame. Can phonetic similarity reliably be the only issue in the accuracy of perception of consonants in tone languages such as Igbo? Can a confusion experiment be performed in tone languages using a VCV frame with a specific tone pattern, and a confusion matrix presented based on the consultants' confusion of the consonants? These beg for answer.

Aims of the experiment

The experiment basically aims at providing relevant data that would be used to prepare a confusion matrix for Igbo consonants. It would find out whether the degree of phonetic similarity among the Igbo consonants was the major issue in segment confusion. How much a segment is confused will, to some extent, be provided. This is to be achieved by calculating the percentage of the subjects that perceived each consonant correctly; which would invariably reveal the percentage that failed to get it right. All these are to be displayed in the Igbo consonant confusion matrix.

Methodology

One hundred Igbo-speaking subjects were used for the experiment. Fifty of the students formed the control group, and they were undergraduate students of the Department of Linguistics, Nnamdi Azikiwe University, Awka. A classroom was the venue where information was elicited from the control group at a period when most students have gone home. This was to ensure not only that the subjects were not distracted by their peers but also that the environment was as noise-free as possible. Fifty undergraduate students of the Department of Igbo Language, Chukwuemeka Odumegwu Ojukwu University, Igbariam Campus were used as the experimental group for the experiment. They were taken into the language laboratory of their university for the recording of the research proceedings.

The stimuli comprised 28 consonants of standard Igbo, presented to the consultants in the following order - /p b t d k g kp gb kw gw m n ɲ ɲw f v s z ʃ ʎ h ɸ dʒ l r j w/. Each is presented in a VCV frame, the vowel and tone being [á]; thus [á _ á] frame. This frame is used so that both the vowels and the tones of the resultant words remain constant, even though some of the resultant words may not be Igbo words because of tone differences. This is intentional as it reduces the incidence of correct guessing of the sounds by the consultants since, according to Harley (1995:31), "knowing the word may help us to identify their constituent sounds... We may not even need to hear all the sounds of a word before we can identify it."

The stimuli were presented to the control group in an almost noise-free environment at normal hearing intensities (cf. Lawrence, 1970; Crystal, 1997; Malmkjaer, 2002). For the experimental group, however, the stimuli were given under conditions of masking noise generated by a familiar music from a cassette recorder.

IGBO CONSONANT CONFUSION MATRIX: ISSUES ON PHONETIC SIMILARITY-----Cecilia Amaoge Eme, Linda Chinelo Nkamigbo and Benjamin Ifeanyi Mmadike

The researcher presented the stimuli at normal hearing intensities irrespective of the masking noise emanating from the cassette recorder.

We are not ignorant of Piston’s (1997) findings on phoneme identification, cited in Ingram (2007:152), that for a listener’s identification of phonemes as produced in words by certain speakers, “being familiar with their speaking characteristics will facilitate the task of identifying the phonological content of their speech...” In order for us to avoid giving any clues, the second researcher whose speech characteristics are not familiar to the consultants was used. Also, we acted in line with the advice of Akandu (2015:196), citing Hibbs et al (1965), that the presenter’s “attire must not distract attention from [the] message. Dangling earrings, flashy jewelry, illuminated bow ties and other extreme styles can defeat [the presenter’s] speaking efforts”

Each research subject was given a sheet of paper. She was expected to number the paper serially beginning from 1 down to 28. This corresponds to the number of Igbo consonants to be perceived and identified. She then wrote out the ‘word’ she heard for every number as it was being presented by the researcher. This is irrespective of whether the word is tonally an Igbo word or not. The researcher presented the research material to the consultants at normal hearing intensities under conditions of the masking noise. Before presenting the material to both the control group and the research group, the students were made to understand that the interest of the research is not on the meaningfulness of the resultant material but rather the research wants to find out the segments they could correctly identify. For those in the control group, they were to listen attentively and write down the materials they perceived, even though the materials were presented under masking noise. At the end of the presentation, the sheets of paper were collected and their responses analyzed.

For our analysis, we shall adopt a system different from that adopted by Laver (1994). For him, a matrix of the English consonants showing their phonetic similarity is better displayed using the inverse notion of phonetic distance where two segments with 0% means that they are identical whereas those with 100% means that they do not have any common auditory feature. In our own system, we tilt towards Clark, Yallop & Fletcher’s (2007:315) approach where the confusion matrix for consonants shows a segment that was not confused by any of the consultants as having scored 100% whereas that which was confused by all the consultants scores 0%. In other words, what is shown in our confusion matrix is that the higher the score (in %) for any consonant, the more the number of consultants that perceived it correctly; conversely, the lower the score, the fewer the number. The 28 Igbo consonants, presented in VCV frame, appear in the rows of the matrix while the actual segments perceived appear in columns. The number of our consultants in the experimental group that perceived the actual consonant produced is used to calculate the percentage confusion of the consonants.

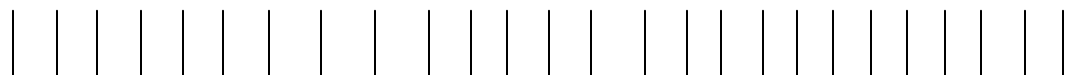
Results and findings

Results

The researchers collected a total of one hundred response documents from the consultants. Whereas half the number is from the control group, the remaining half is from the experimental group. Apart from three subjects from the control group that did not indicate any response for items 14, 21 and 28 respectively, all the 28 items were responded to and correctly too by all in the control group. We take the non response for a number each by three consultants to be a result of their not being fast enough in their writing; not as a result of non perception of the produced segment.

As for the experimental group, the materials we collected show that some of the consonants were actually confused. The table below shows the number of the consultants that confused each segment. The number is indicated under the consonant that is confused.

p	b	t	d	k	g	k	g	k	g	m	n	ɲ	ŋ	f	v	s	z	ʃ	ɣ	h	ʧ	ɗ	l	r	j	w	
1	1	1	1	6	1	2	8	3	2	6	2	1	3	2	4	1	1	6	3	2	1	3	2	2	1	5	8
2	4	9	3	6	8	3			0		6	8	7	6	2	6		6	4	7	5	1	9	4	0		



p	b	t	d	k	g	kp	gb	kw	gw	m	n	ɲ	ɲw	f	v	s	z	ʃ	ʎ	h	ɣ
ɔ̃	l	r	j	w																	
12	14	19	13	6	18	23	8	3	20	6	26	18	37	26	4	12	16				
6	36	24	17	35	21	29	14	50	8												

Table 1: Number of the consultants that confused the indicated consonants

The results also show that at one point or the other one consultant confuses one segment for the other. Thus, we see that each of the 28 consonants has been confused for another consonant somewhere in the cause of the perceptual experiment. This could be an indication that each segment has another or others with which it shares many phonetic similarities; hence their close perceptual qualities, which invariably leads to segment confusion if not quite clearly perceived. The confusion could also be as a result of the consultants’ attempt at perceiving meaning, such that where the emergent word is a nonsense one, the consultant tries to fix a meaning to it by replacing the presented segment with another that would make the word meaningful. This is a very clear possibility. In the words of Clark, Yallop & Fletcher (2007:314), experiment shows that, “when segments are excised from the stream of speech and replaced by noise, listeners will report hearing the correct missing segment. They presumably restore the segment by top-down contextual prediction.” Thus, a segment perceived under conditions of masking noise could be wrongly dropped for another if the former appears in a nonsense word and replacing it with the predicted choice makes the word meaningful.

Findings

We present below some of the findings derivable from our experiment:

One of our findings is that none of the 28 consonants was correctly perceived by all the consultants. We, therefore, did not have any 100% figure in the confusion matrix; implying that no segment was not confused by even a single consultant.

Confusion of segments is peculiar to the consultants in the experimental group. All the consultants in the control group perceived all the segments accurately. This is expected because all the consonants were familiar to the consultants, and the VC̣Ṿ syllables were presented in a noise-free environment and therefore clearly perceived. This does not rule out the fact that certain Igbo consonants have close perceptual qualities. Their close perceptual qualities lead to their being confused when they are not clearly perceived; such as when they are said in a noisy place and the listener is expected to sieve them out from the noise. This is achieved in this experiment by our use of masking noise. No wonder some segments were confused.

Table 1 and the confusion matrix show that the greatest confused consonant is [j]. All the consultants in the experimental group confused this consonant with another sound, hence none of the 50 consultants perceived the segment. It scores 0% in the confusion matrix. Conversely, the least confused of all the Igbo consonants is [kw]; only 3 people confused it and it scored 94% in the matrix. Closely following [kw] is [f] which scores 92% and confused by 4 people; and then, [k m z] with each scoring 88% and confused by 6 consultants. According to our confusion matrix, the segments [j] and [ʎ] have the closest perceptual similarity and therefore often confused perceptually. The confusion score across themselves is as high as 56% in the confusion matrix. This result is surprising, since the segments are not all that phonetically similar; though both are oral high voiced consonants requiring open spread lips for their production. We feel that many of the consultants ‘forced’ themselves to perceive [áʎá], the standard Igbo word for ‘war’; as against perceiving the presented [ájá] ‘war’, which is dialectal.

IGBO CONSONANT CONFUSION MATRIX: ISSUES ON PHONETIC SIMILARITY-----Cecilia Amaoge Eme, Linda Chinelo Nkamigbo and Benjamin Ifeanyi Mmadike

Furthermore, we see from the matrix that [tʃ] is confused with [t], as much as [l] is confused with [r]; both scoring 40%. Also, [kp] is confused with [p], as much as [n] is confused with [l]; each of the pairs scored 28%. The phonetic similarities between these confused sounds seem more glaring. On the other hand, many

	p	b	t	d	k	g	kp	gb	kw	gw	m	n	ɲ	ɲw	f	v	s	z	ʃ	ɣ	ɦ	tʃ	dʒ
--	---	---	---	---	---	---	----	----	----	----	---	---	---	----	---	---	---	---	---	---	---	----	----

consonants were never confused for each other. For instance, none of the fricatives or approximants was ever confused with the voiceless bilabial plosive [p]. This is as expected because they are not phonetically similar.

An Igbo consonant confusion matrix based on the results and findings

Below is a confusion matrix showing which consonant has been confused, the segment with which it is confused and the number of consultants involved in such confusions. Meanwhile, the segments presented by the researchers appear in the rows of the matrix while the actual segments perceived appear in columns.

IGBO CONSONANT CONFUSION MATRIX: ISSUES ON PHONETIC SIMILARITY-----Cecilia Amaoge Eme, Linda Chinelo Nkamigbo and Benjamin Ifeanyi Mmadike

VCV frame consciously presented under the condition of masking noise. Conversely, phonetic distance and meaningless or nonsense structures greatly contribute to wrong perception of the sound segment presented if not clearly perceived. Thus, some consultants could have confused some segments not necessarily because the segments are phonetically similar but because the consultants tended to consider meaningfulness of syllables rather than the segment they actually perceived. One should actually not be surprised at this development, as Holt and Lotto (2010) have earlier shown that “speech is perceived through the lens of native language categories.” There is a need for further investigation on segment confusion in tone languages, possibly using more frames, in accordance with the phonotactics of the tone language to be investigated.

References

- Akandu, U.L. 2015. Strategies for speaking effectively in the English language. In L.C. Nkamigbo and F.O. Asadu (eds.). *Current issues in linguistics, language and gender studies: A festschrift in honour of Professor Cecilia Amaoge Eme at 50*. Nkpor: Brystevand Publishers.
- Clark, J. E., C. Yallop and J. Fletcher. 2007. *An introduction to phonetics and phonology*. Oxford: Blackwell Publishing.
- Coppens, A. B. 1997. Sound. In Parker, S. P. (ed.) *McGraw-Hill encyclopedia of science and technology*. New York: McGraw Hill Book Company. 17: 5-16.
- Crystal, D. 1997. Crystal, David. *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.
- Eme, C.A. 2007. Parameters of intersegmental co-ordination in Igbo speech patterns: A perceptual analysis. PhD Thesis, University of Nigeria, Nsukka.
- Harley, T.A. 1995. *The psychology of language: From data to theory*. Hove: Psychology Press.
- Holt, L.L. and A.J. Lotto. 2010. Speech perception as categorization. www.ncbi.nlm.nih.gov/.../PMC2921848. Retrieved on 22 August, 2015.
- Ingram, J. C. 2007. *Neurolinguistics: An introduction to spoken language processing and its disorders*. Cambridge: Cambridge University Press.

- Ladefoged, P. 1962. *Elements of acoustic phonetics*. Chicago: The University of Chicago Press.
- Laver, J. 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.
- Lawrence, A. 1970. *Architectural acoustics*. London: Applied Science
- Lieberman, P. and S. Blumstein. 1988. *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
- Malmkjaer, K. (ed.). 2002. *The linguistics encyclopaedia*. London: Routledge.
- Mbah, E.E. and I.B. Onah. 2015. Effect of podcasts on the Igbo learners of English vowels. In R.I. Okorji, C.A. Eme & G.S. Omachonu (eds.) *Contemporary issues linguistics and language studies: A festschrift for Prof. Clara Ikekeonwu at 65*. Enugu: Rossen Publications Ltd. 240-256.
- Nolan, F. 1992. Acoustic phonetics. In W. Bright (ed.). *International encyclopedia of linguistics*. Oxford: Oxford University Press. 1: 3-9.
- O'Callaghan, C. 2014. Speech perception: Empirical and theoretical considerations. *Stanford encyclopedia of philosophy*. www.plato.stanford.edu/entries/perception-au. Retrieved on 22 August, 2015.