An Analysis of Minimal Pairs in Igbo Using a Multimodal Approach to Speech Perception

Chinenye Esther Onuoha*
&
Ebele Deborah Uba*
https://dx.doi.org/10.4314/ujah.v25i1.2

Abstract

Speech perception is the process by which the sounds of a language are heard, interpreted and understood. The traditional assumption that language is an auditory phenomenon was over ruled by the McGurk effect which proved by the use of experiment that visual signal also assists the auditory signal in the perception of speech sounds of a language. This work aims at examining minimal pairs in Igbo to determine if this theory of multimodal (audio-visual) perception of speech is applicable. The minimal pairs used in this work are extracted ten Igbo native speakers comprising of six females and four males, and from existing Igbo literature on minimal pairs while some are by introspection of the researchers as native speakers of the language under study. The variety of Igbo used in this work is the standard Igbo which is the variety of the variety that is officially recognized. For data analysis, the Multimodal Theory of speech perception was applied. Findings indicate that the native speakers of Igbo do not use visual information in order to properly perceive the sounds of their language, even those that are minimal pairs. The work concludes that the need for visual information for adequate identification, perception and

comprehension of speech sounds cannot be said to be a general phenomenon since native speakers of a language could without the visual information, perceive and identify their speech sounds when audio recorded not minding the confusion that usually arises with minimal pairs. This work suggests that the Multimodal Theory of speech perception may only be applied when working with non-native or non-fluent speakers of a language.
Keywords: Multimodal, Speech Perception, Language, Speech Sounds, Minimal Pairs.

1.Introduction

Speech perception aims at understanding how language users properly recognize speech sounds and use this information to understand spoken language. It is mainly concerned with how language is understood and not how it is produced. Research in speech perception has applications in computer systems that are capable of recognizing speech, improving speech recognition for hearing and language-impaired listeners, and in foreign-language and second language teaching.

In speech perception, listeners focus attention on the sounds of speech and notice phonetic details about pronunciation that are often not noticed at all in normal speech communication. Our auditory system as speakers of a language shapes our perception of speech. That does not also mean that all speech perceptions are solely determined by our auditory abilities. Our perception of speech is also guided by our personal phonetic knowledge of speech production, and the knowledge of the linguistic structures of our native language.

The previous methods of investigating the process by which human beings understand speech has been to focus on the auditory perception of speech signals alone. However, current research has proven that the production and perception of speech may not only involve the auditory organs but the optical signals also.

That is to say that a different approach which believed in the interaction of the audio and visual phenomena developed was established. McGurk and MacDonald (1976:746) in their publication "Hearing lips and seeing voices" claim that it had become clear that visual modality also plays a role in the perception of speech.

McGurk and MacDonald (1976) played a video of someone saying "ba" when the audio was of a voice saying "ga". Participants reported hearing "da", a blend of visual and auditory information. This demonstration is referred to as the McGurk effect. In these cases, subjects sometimes report hearing neither exactly what was seen nor exactly what was heard, but some different syllable in between. For example, if subjects see a speaker producing a /ga/ and hear simultaneously a /ba/, they may report hearing a /da/. This McGurk effect is taken as one of the major pieces of evidence for gestural approaches because such theories give detailed account of how the auditory and visual information are integrated during perception. It is understood that they are integrated because both the auditory and visual signals provide the language user and the observer with information about the articulatory process. The visual cue gives information about the place of articulation (e.g. bilabial, alveolar or velar), and the acoustical cue, about the manner of articulation (Summerfield 1987:64) the visual cue from a speaker's mouth play a role in speech perception in everyday life. Another example was provided by Massaro 1998 (cited in Massaro

2002:46) where he asserts that 'a perceiver's recognition of an auditory-visual syllable reflects the contribution of both sound and sight. For example, if the ambiguous auditory sentence, *Mybab pop me poo brive* is paired with the visible sentence, *My gag kok me koogrive*, the perceiver is likely to hear, *my dad taught me to drive*. Two ambiguous sources of information can be combined to create a meaningful interpretation'. Although this was not duly experimented, however, it has provided sufficient proof that the process of speech perception is a multimodal rather than a unimodal process. The purpose of this paper is to apply this multimodal approach in the perception of minimal pairs in the Igbo language with the aim of determining whether the visual information is solely as mandatory as the auditory information in the perception of minimal pairs by Igbo native speakers. The reason is to test the multimodal theory of speech perception using a natural language like Igbo, either to support or refute the notion.

Methodology

The data for this work were extracted from ten Igbo native speakers comprising of six females and four males, and existing Igbo literature on minimal pairs. Also, some of the data were got through the introspection of the researchers as native speakers of the Igbo language. The variety of Igbo used in this work is the standard Igbo which is the one that is officially recognized by all the Igbo native speakers. Using the Multimodal Theory of speech perception, the researchers conclude that it is best applied when working with non-native or non-fluent speakers of the Igbo language.

2. Speech Perception

Speech perception does not involve only phonological knowledge of sounds and prosody rather it includes syn tactic and pragmatic knowledge (Tatham & Morton 2011:127). The process of speech perception begins at the level of the sound signal and the process of auditory signals.

According to Barlow and Gierut (2002:58), "a minimal pair is a set of words that differ by a single phoneme whereby the difference is enough to signal a change in meaning". They can also be defined as "pair of words that are identical in all respect except for the sounds in question" (Dirven and Verspoor 2004:114). The features of minimal pairs include: i) they must have the same number of; ii) they must differ in a single sound in a corresponding position in the two words; iii) they must differ in meaning (Akmajian et al. 2004:94). Mbah and Mbah also point out that there is the existence of minimal pairs for tone contrast in tonal languages which also distinguishes meaning (2010:154).

Hearing is the first component of perception process. It is merely the vibration of sound waves on the ear, which is one of the five senses of a person. Hearing occurs involuntarily and effortlessly. The ear simply picks up noise without giving attention to interpret or understand the noise heard. Hearing can even occur as one sleeps. What the ear does is to convert what it hears (sounds) into nerve impulses and send it to the brain but the brain does not always react to the sounds. Tatham and Morton (2011:126) distinguish between hearing and perception thus; 'Hearing is a peripheral process which is largely passive, in the sense that we do not influence its workings by thinking about it either tacitly or with some awareness while Perception is a central process which is mostly active, in

the sense that thinking, or cognitive processing, dominates in the process rather than anything physical.' It is essential to make this distinction between the passive physical process of hearing (which they referred to as tutorial-hearing), which takes place in the ear, and the active, cognitive process of perception, which takes place in the mind. The authors further posit that "speech perception is about sorting out which symbols the speaker had in mind when they produced the speech the listener can hear". They illustrate this using the diagram below.

Speaker		Listener
The speaker has ideas expressed as linguistic symbols, and encodes these as speech sound waves.	The sound waves travel between speaker and listener	The listener hears the signals, and perceive them by working out the original symbols the speaker had in mind when the speech was produced.

Table 1 (Culled from Tatham and Morton 2011: 127)

In studying perception, there are a few observations that can be made about perceptual behavior which Tatham and Morton (2011:130) present as:
1. It seems that our perception is categorical; that is, we categorize what we hear into predefined groups based on the phonological segments that match the specific language we are listening to. Thus, the theory goes that we need to be aware of the phonology of our language. These are acquired

as we learn a language, especially our native tongues, as speakers of different languages employ distinct sets of categories.

2. Understanding prior to perception entails at least:

a) some of the general properties of sounds and how they are used in language in general;

b) the language's usage of sound categories;

c) the sound features that must be met in order for them to be classified;

d) the phonotactics of the language – that is, the rules for combining the sound features and the rules for sequencing sounds in syllables (and words).

Agreeing with the theory of speech as a multimodal process, Massaro (2002:46) states that ‘‘speech as a multimodal phenomenon is supported by experiments indicating that our perception and understanding are influenced by a speaker’s face and accompanying gestures, as well as the actual sound of the speech which we hear’’. According to Massaro, many communication environments involve noisy ear canals, which impair speech perception and recognition. Visible speech from the speaker's face (or from a reasonably accurate synthetic speech head) improves comprehension in these situations. The number of words understood by an impaired auditory message can often be doubled by associating that message with the speaker's facial speech. The combination of auditory and visual abilities in speech perception is called super-additive because their combination can yield much greater accuracy than that of either modality alone.

There are three theories of speech perception, which Tatham and Morton have outlined as:

The Motor Theory of Speech Perception - An Active Theory: This active theory of speech perception (Liberman *et al*. 1957), involves the participation by the listener in interpreting the incoming sound wave in terms of basic units – sound segments. Perceivers ask themselves what motor control they would have to perform to make the particular sound. When the articulation is identified, the next question to ask is what phonological unit would underlie the attempt. It is then hypothesized that it is that same phonological unit that underlies the speaker's production. The Motor Theory of speech perception focuses on the idea that speech sounds are renderings of underlying phonological units and that a perceiver is trying to identify what these underlying phonological units are. The theory sees speech production in terms of three levels: abstract phonological planning, physical articulation in terms of motor control, and the resultant acoustic signal due to aerodynamic effects.

The Analysis- by- Synthesis Theory of Speech Perception – An Active Theory:

This theory of speech perception (Stevens and Halle 1967), involves the active participation by the listener in interpreting the heard sound wave. Perceivers ask themselves whether they can make the sound they hear, and if they can, go on to ask what phonological unit would underlie the attempt. It is then hypothesized that it is that same phonological unit which underlies the speaker's production, as shown below. The theory recognizes that speech sounds are renderings of underlying phonological units, and that a perceiver is trying to identify these phonological units. The theory omits reference to articulation or motor control – a level that is

important in the Motor Theory of speech perception, which was contemporary with the Analysis- by- Synthesis theory.

The Associative Store Theory of Speech Perception – An Active Theory:

This theory (Levinson 2005; Tatham and Morton 2006) is a comprehensive, active theory which accounts for a number of observations about speech perception. These include the following:

The continuous acoustic signal is interpreted in terms of an underlying sequence of abstract phonological units; abstract cognitive labels are assigned to the acoustic signal.

ii. A device is present which detects and traps errors of interpretation, and causes a reappraisal of that portion of the signal which has been wrongly interpreted – interpretation error correction.

iii. Mechanisms exist to repair signals damaged before the interpretation process – production and transmission error correction.

The Associative Store Model thus focuses on the ability to become aware of error, and repair it.

3. Minimal Pairs in Igbo

Akmajian et al. (2004:94) assert that for a pair of words to constitute a minimal pair, it must – i) have the same number of phonemes ii) differ in a single sound in a corresponding position in the two words and iii) differ in meaning.

They use the English words ‘fine’ and ‘vine’ to exemplify this fact. The words have the same number of phonemes, differ in meaning, but

phonologically, they differ only in contrast between initial /f/ and initial /v/. Thus, /faɪn/ and /vaɪn/ constitute a minimal pair. This study agrees with Akmajian et al (2004) because it is the contrastive sounds in the same position that are considered in determining a minimal pair not the letter symbols. In the opinion of Barlow and Gierut (2002:58), "a minimal pair is a set of words that differ by a single phoneme whereby that difference is enough to signal a change in meaning." Dirven and Verspoor (2004:114) submit that 'a simple way of deciding whether two sounds in a language belong to one phoneme or to two different phonemes is to look for minimal pairs.' They regard minimal pair as a pair of words that are identical in all respect except for the sounds in question. They illustrate with the words 'pat', 'bat' and conclude that the pair confirms that /p/ and /b/ constitute separate phonemes in English, while the impossibility of a contrast between [sp^{h}ai] and [spai], or between [p^{h}ai] and [pai] confirms that [p^{h}] and [p] do not belong to different phonemes in English. The definitions of 'minimal pair' given by Barlow & Gierut (2002), and Dirven & Verspoor (2004) are very clear. They give a clearer picture of what is a minimal pair and one can easily deduce from these scholars' view that minimal pair is the juxtaposition of related words of a language to find out the sounds of the language that occurs in the same environment, thereby giving different meanings to the words.

Mbah and Mbah (2010:154) argue that 'the writing system of a language is used to determine minimal pairs only when the written form coincides with the spoken form as in [pet. bet].' They further illustrate with /k/ sound in *knee* and *gh* in *nigh*, pointing out that because no one produces /k/ in *knee*, it cannot be used to judge whether minimal pair exist in a set of words or not. Mbah and Mbah also contend that since the glide /aɪ/ is what appears in the pronunciation of the word *nigh* and not *gh*, the glide /aɪ/ should equally be considered in its forming of a minimal pair or not rather

than *gh*. This study also agrees with Mbah and Mbah's (2010) view because in minimal pairs test, what are considered very vital are sounds and not the letter symbols of the language.

Furthermore, before any meaningful contrast is arrived at, Mbah and Mbah (2010) have identified two levels of analysis for generating minimal pairs. The two levels are the prosodic and the segmental levels (Nartey and Arishi, 1989; Essien, 1990; Roach, 2009). Mbah and Mbah note that the phenomenon of tonal minimal pair is rampant in African languages. This means that there is equally minimal pair for tone contrast in tonal languages and that tone distinguishes meaning in tonal languages. Some scholars such as Pike (1948) and Roach (2009) refer to this as a contrast involving the tonemes. Based on the two tiers identified for generating the minimal pairs, Pike (1948) concludes that contrastive lexical unit of sounds are phonemes, while contrastive sound units especially in tone languages are tonemes. He defines a tone language as a language that has a lexically significant contrastive but relative pitch on each syllable.

Igbo is a tone language where tone contrasts meaning. Katamba (1989:186) defines a tone language as one which has morphemes that are at least in part realized by pitch modulation. He identifies the various pitch modulations and their representations:

(3) [ˊ] high tone
[ˋ] low tone
[˗] mid tone
[ˆ] falling tone
[ˇ] rising tone

Fromkin, Rodman & Hyams (2011:213) describe tone languages as languages that use the pitch of individual vowels or syllables to contrast meanings of words. They made a distinction between the two types of tones: the register tone and the contour tone. A register tone is identified if the pitch is level across the syllable; whereas a contour tone is identified when the pitch changes across the syllable, whether from high to low or from low to high. Igbo language typically operates a register tone system. It exhibits three tones: high (◌́), low (◌̀) and downstep (◌̄) tones as can found in the following Igbo words:

(4) éké ‘python
(5) ísí ‘head’
(6) áká hand’
(7) ìsì ‘blindness’
(8) àlà ‘land’
(9) ùgò ‘dove’
(10) égō ‘money’
(11) ọ́nṵ̄ ‘mouth’
(12) ézē ‘teeth’

In the next section, we shall examine the perception minimal pairs in Igbo by native speakers.

3. Data Presentation and Analysis

In this work, a number of the selected minimal pairs are distinguished by tone but majority of them are within the same tone.

Selected Igbo minimal pairs

No.	Minimal Pairs	Difference
13.	ágū‘lion’ árū ‘abomination’	Consonantal
14.	ọ́rụ́ ‘work’ ọ́kụ́ ‘fire’	Consonantal
15.	íkwú ‘relation’ ígwú ‘lice’	Consonantal
16.	ọ́gwụ̀ ‘medicine’ ọ́gụ̀ ‘hoe’	consonantal
17.	ọ́chị̀ ‘laughter’ ọ́jị̀ ‘iroko tree’	consonantal
18	ènwè ‘monkey’ èfè’ ‘dress’	consonantal
19.	ụ́zọ̀ ‘door’ ụ́kọ̀ ‘reverend’	consonantal
20.	íkū ‘to plant’ ítū ‘to throw’	Consonantal
21.	ímò ‘name of Eastern state in Nigeria ísò ‘to follow’	Consonantal
22.	óké ‘male’ óché ‘chair’	Consonantal

Table 2: Consonantal Minimal Pairs 1

No.	Minimal Pairs	Difference
23.	ényì ‘friend’	Consonantal

	ézì ‘pig’	
24.	ákwá ‘cry’ áfá ‘divination’	consonantal
25.	ńkwụ́ ‘palmwine’ ńkụ́ ‘firewood’	consonantal
26.	ìtè ‘pot’ ìhè ‘light’	Consonantal
27.	íwé ‘anger’ íhé ‘thing’	Consonantal
28.	árá ‘breast’ áká ‘hand’	Consonantal

Table 2: Consonantal Minimal Pairs 2

No.	Minimal Pairs	Difference
29.	ákà ‘bead’ ụ́kà ‘church’	Vocalic
30.	ísā ‘to wash’ ísū ‘to cut grass’	Vocalic
31.	ìgbé ‘box’ ìgbó ‘indian hemp’	Vocalic
32.	ísè ‘to draw’ ísò ‘to follow’	Vocalic
33.	ákà ‘bead’ ụ́kà ‘church’	Vocalic
34.	ọ́nwá ‘moon’	vocalic

No.	Minimal Pairs	Difference
35.	áká ‘hand’ ákà ‘bead’	Tonal
36.	áfọ̀ ‘year’ áfọ́ ‘stomach’	Tonal
37.	òké ‘rat' óké ‘male’	Tonal
38.	íré ‘tongue’ írē ‘to sell’	Tonal
39.	énYí ‘elephant’ énYì ‘friend’	Tonal
40.	ígà ‘chain ígā ‘to go’	Tonal
41.	ímà ‘to measure’ ímā ‘to know’	Tonal
42.	ísì ‘smell’ ísī ‘to cook’	Tonal
43. 44.	íkè ‘bottom’ íkē ‘to tie’ ákwà ‘cloth’ ákwá ‘cry’	Tonal Tonal
	ọ́nwụ̀ ‘death’	

Table 3: Vocalic Minimal Pairs

Table 4: Tonal Minimal Pairs

It is observed that in all these examples of minimal pairs above, the difference for every pair is in respect of one element in the same

environment, whether segment, vocalic or tone. The differing element is not more than one for each pair because if it exceeds one, it ceases to be described as a minimal pair. Through the audio recording of all the minimal pairs by the researchers, the native speakers exhibited a proper identification and comprehension of the sounds in the pairs both those with the same tone and the ones that differ in tone. The native speakers proved to possess a perfect understanding of the phonetics and the linguistic structures of the Igbo language by identifying the sounds of the language easily without any visual knowledge of the speaker. From the analysis, the native speakers of Igbo do not use visual information in order to properly perceive the sounds of their language. Unlike the McGurk's video where someone perceives /ba/ as /ga/, this research shows that even without visual information, Igbo native speakers can identify, perceive and understand speech sounds adequately. This goes to say that linguistic experience plays a major role in speech perception among Igbo native speakers more that visual information does.

4. Summary and Conclusion

This work has studied speech perception in the Igbo language. The McGurk effect of audio-visual (AV) perception was the basis for the study with the purpose of examining the minimal pairs in Igbo using Igbo native speakers to find out if the McGurk effect is also applicable to the Igbo language and most importantly, to native speakers of the language. The study was carried out using thirty-four (34) minimal pairs which included a few which could only be differentiated by tone. The highlights of the findings indicate that native (fluent) speakers of Igbo do not need the

information on a speaker's face to enable their quick perception of the sounds not minding the tonal changes.

This work therefore, concludes that the McGurk effect which claims the need for both visual and auditory information for speech perception may not be applied to native speakers of Igbo language (rather only be applied when working with non-native or non-fluent speakers of a language) because if the native speakers of Igbo could easily identify minimal pairs that differ only in sound (consonants), identifying other varieties of sound will not pose much difficulty. That is to say that the multimodal nature of speech perception remains under investigation in various aspects of language as it cannot be generalized.

Chinenye Esther Onuoha, PhD*
Department of Linguistics,
Nnamdi Azikiwe University.
cokoye@unizik.edu.ng
&
Ebele Deborah Uba, PhD*
Department of Linguistics,
Nnamdi Azikiwe University.
ed.uba@unizik.edu.ng

References

Akmajian, A., Dermers, R., Ann, K., and Robert, M. (2004). *Linguistics: An Introduction to language and communication* (5th Ed.). New Delhi: Prentice Hall of Indian Private Limited

Barlow, J.A. and Judith A.G. (2002). *Minimal pair approach to phonological remediation*. www.speechlanguagetherapy.com/Pdf/barlowgierut2002.pdf.

Dirven, R., and Marjolijn V. (2004). *Cognitive exploration of language and linguistics*. 2nd Ed. Amsterdam: John Benjamins Publishing Company.

Fromkin V., Rodman R. & N. Hyams. (2011). *An Introduction to Language*, 9th edition. Canada. Wadsworth Cengage Learning.

Katamba, F. (1989). *An introduction to phonology*. New York: Longman Group.

Levinson, S. (2005). Mathematical Models for Speech Technology. Chichester: John Wiley.

Liberman, A.M., Hams, K.S., Hoffman, H.S., Griffith, B.C. (1957). *The discrimination of speech sounds within and across phoneme boundaries*. Journal of experimental psychology, 54 (5), 358 https://doi.org/10.1037/h0044417

Massaro D.W. (2002). *Multimodality in Language and Speech Systems.* B. Granstrom et al (eds.). Nertherlands, Academic Publishers, 45-71.

Mbah, B. M. and E.E. Mbah. (2010). *Topics in phonetics and phonology; contribution from Igbo.* Nsukka: A P Express Publishers.

McGurk, H. and. MacDonald, J. (1976). *Hearing lips and seeing voices*. Nature, 264, p. 746-748.

Stevens, K.N., and Halle, M. (1967). Remarks on analysis by synthesis and distinctive features, in Models for the Perception of Speech and Visual Form. Edited by W. Wathen- Dunn. M.I.T. Press. Cambridge, pp. 51-66.

Summerfield, Q. (1987). *Some Preliminaries to a Comprehensive Account of Audio-Visual Speech Perception.* In Hearing by eye: The psychology of lip-reading. Dodd B and Campbell R (editors). London: Erlbaum.

Tatham Mark and Morton. K. (2011). *A Guide to Speech Production & Perception*. Edinburgh University Press.

Tatham, M. and Morton, K. (2006). *Speech Production and Perception*. Basingstoke: Palgrave Macmillan.