A Model for Speech Processing in Second Language Listening Activities

Wafa Shahada Zoghbor[1]

[1] Department of English Language and Writing Studies, Zayed University, United Arab Emirates

Correspondence: Wafa Shahada Zoghbor, Department of English Language and Writing Studies, Zayed University, United Arab Emirates. Tel: 971(0)2-5993270. E-mail: wafa.zoghobr@zu.ac.ae

Abstract

Teachers' understanding of the process of speech perception could inform practice in listening classrooms. Catford (1950) developed a model for speech perception taking into account the influence of the acoustic features of the linguistic forms used by the speaker, whereby the listener 'identifies' and 'interprets' these linguistic forms based on the association between them and the context of speech. This paper critically reviews Catford's model and proposes an alternative one distinguishing between two levels of perceiving speech: word recognition and utterance comprehension. Smith and Nelson (1985) refer to these as 'intelligibility' and 'comprehensibility', respectively. The proposed model could inform classroom practice as well as curriculum and material design.

Keywords: Catford (1950), intelligibility and comprehensibility, top-down and bottom-up processing

1. Introduction

Teaching resources on practising listening for second language learners subscribe primarily to Howatt's and Dakin's (1974) definition of listening ability, in which the successful completion of the listening process relies on the listener's ability to identity and understand what is being said. Catford (1950) provides a model for speech perception which focuses not only on how utterances are pronounced and heard but also on how the listener may cognitively receive and interpret speech.

The critical review of Catford's model in this work is based on a discussion of two main groups of concepts. The first is the two contrasting processes for perceiving speech: *bottom-up* and *top-down* (Brown, 1990). The first process assumes that speech is perceived in a series of phases starting from the phonemes (e.g., /b/, /ɒ/, /g/) as the smallest unit of speech, then moving gradually to larger units which can cover an utterance and the message it carries (Anderson & Lynch, 1988). The second process is the *bottom-up* process, which contrasts with top-down processing in the sense that the listener interprets a message through investigating its context and employing his/her background knowledge to grasp the possible meanings of an utterance (Pinker, 1994).

These two processes focus on how the receiver (or listener) might perceive and process speech, thus marginalizing the role of the speaker in a setting. This leads to introducing the second group of concepts, which have long been associated with the speaker (rather than the listener): these are 'intelligibility' and 'comprehensibility'. Linguists have provided several definitions for the concept of 'intelligibility', which is more commonly seen in the literature than 'comprehensibility'. This work gives equal weight to both of these two terms, and subscribes to the distinction between them given by Smith and Nelson (1985) in which 'intelligibility' refers to recognition of individual utterances, while 'comprehensibility' refers to understanding the meaning of the utterance. The concept of 'intelligibility' has been widely appealed to as an important criterion for any pronunciation model. Recently, this has even been the case more often than ever before, with the emergence of literature on English as a lingua franca and the argument that 'intelligibility' (more often than 'comprehensibility') is the main concern in cross-cultural communication. Although the term 'intelligibility' has been present in much of the literature that discusses pronunciation models, and despite the informative relationship between speaking and listening skills, the influence of 'intelligibility' on listening (not only speaking) remains passive.

This work will start by detailing Catford's (1950) model and then the two groups of concepts in order to provide

an overview of their practicality in second language classrooms. Catford's model will then be revised and an alternative one will be proposed based on integration with the other two groups of concepts.

2. Catford's Model of Speech Perception

Speech perception is not only closely linked to the speaker's pronunciation of utterances but also to the listener's cognitive psychology (Clark & Yallop, 1995). Catford (1950) discusses two types of context that might, possibly, increase or decrease speech understanding thresholds: *linguistic* and *situational*. While the former is limited to the given words or other linguistic forms, the latter broadly includes everything else in the situation relevant to the speech-act, including the hearer's and speaker's linguistic and cultural backgrounds and experience.

For Catford (1950), the speaker must *select* the linguistic forms which are deemed appropriate to the situation. This involves selecting appropriate words and deciding the possible structure and sounds. Next, the speaker should *execute* the linguistic forms he/she has selected in an appropriate manner that will approximate to the norm obtained in the speech-community within which the speaker is operating. At this stage, execution may fail if sounds are mispronounced. Execution is followed by *transmission* of sounds through a physical medium. Some loss of speech recognition and understanding may occur due to defective transmission.

The hearer must correctly *identify* the linguistic forms he/she hears. This involves the hearer's ability to discriminate between the heard sounds and to associate them correctly with his/her private 'mental images' of these sounds. For example, failure during identification might occur if a hearer cannot distinguish between /o/ and /ʌ/, so that *collar* might be misheard as *colour* (Catford, 1950).

Finally, the hearer is expected to associate the heard linguistic forms with the elements in the setting.   In doing this, the hearer is then expected to respond to the utterance in accordance with the shared set of norms among the people in the speech-community within which he/she is operating. Failure to do this may result in *failure in interpretation*.

Catford summarized the stages mentioned above as follows:

a.　　Speaker's *selection* of linguistic forms.

b.　　Speaker's *execution* of linguistic forms.

c.　　*Transmission* from speaker to hearer.

d.　　Hearer's *identification* of linguistic forms.

e.　　Hearer's *interpretation* of linguistic forms.

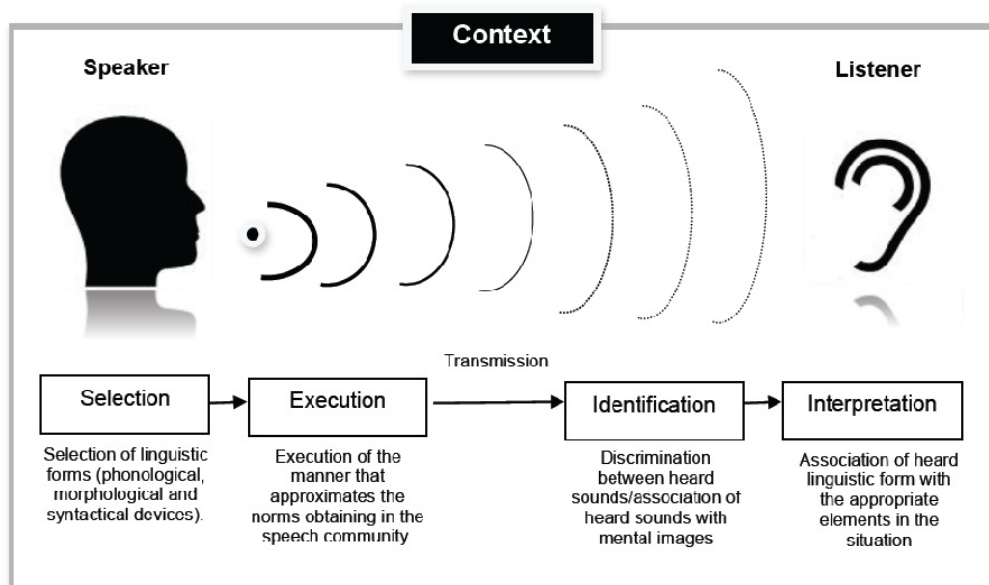Figure 1 below was developed to visualize Catford's (1950) model of speech perception.

Figure 1. A visualization of Catford's (1950) model of speech perception

3. Top-Down and Bottom-Up Processing

In listening and speaking in second language classes the discussion of top-down/bottom-up processing is relevant in two senses. Firstly, as it relates to the differences between the approach used by native speakers (NSs) and non-native speakers (NNSs) in perceiving speech (as well as the differences among NNSs according to their level of command in English), and secondly the classroom practices that are connected with them.

Literature suggests that NSs and NNSs perceive speech differently. Both Brown (1990) and Jenkins (2000) report that NSs are more able to use a top-down process even with limited phonological input due to their background knowledge of the language. In everyday situations, even if NSs do not hear all the details at the phonemic level of the utterance, they still have the potential to guess what could have been said. In contrast to NSs, NNSs and second language learners are more likely to rely on 'bottom-up' processing (Brown, 1990; Jenkins, 2000), especially in the early stages of learning the target language. Learners at this stage depend on given cues (or phonemes) in the language provided by the speaker, rather than employing background knowledge about the language. Listeners who are able to use the phonological code competently have a good chance of recognizing most of the words intended by the speaker (Brown, 1990).

While both Brown (1990) and Jenkins (2000) emphasize that it is bottom-up processing that is connected with the phonological code, and with identifying which phoneme is being used, what seems to be negotiable in employing the above processes is the effect of the proficiency level of second language learners in employing top-down processing. While Brown (1990) mentions that NNSs of English with high proficiency might exploit the context and use top-down processes, Jenkins (2000) seems to believe that NNSs, even at relatively high levels of competence, still predominantly process speech using bottom-up strategies. Jenkins (2000) attributes this to the complexity of the top-down process, which requires the employment of both linguistic and extra linguistic levels, causing the top-down process to be rarely applied at the same level of efficiency as that employed in NSs. In listening classes, it might be expected that learners will process utterances by relying on their recognition of their phonological code but they are also encouraged to infer what the components of individual utterances are from their understanding of the context. Teachers also need to distinguish between what is expected of students and how they might actually process the heard speech in listening activities. Figure 2 was developed to explain the differences as well as the relationship between top-down and bottom-up processing.
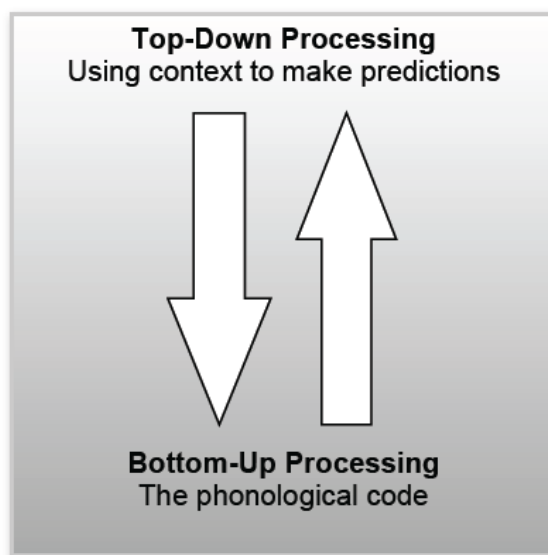

Figure 2. Top-down and bottom-up processing

4. Intelligibility and Comprehensibility of Speech

Kenworthy (1987, p. 13) identifies 'intelligibility' as "being understood by a listener at a given time in a given situation". It is viewed as being the same as 'understandability'. For Kenworthy (1987), intelligibility correlates positively with successful identification of the words in speech, even though intelligibility can still be successful when words are not fully identified.

Catford (1950) offers a broader definition of 'intelligibility' that covers the identification stage which Kenworthy talked about but goes past this stage into the hearer's response. For Catford (1950), an utterance is considered 'intelligible' if it is 'effective', where 'effectiveness' is an appropriate response from the hearer that is in line with the semantic habits of the speech-community in specific communication settings.

Both Munro and Derwing (1995) and Derwing and Munro (1997) identify 'intelligibility' as the extent to which a speaker's utterance is understood. They emphasize the importance of distinguishing this notion from 'comprehensibility', which refers to the listener's estimation of the difficulty or ease with which he/she understands an utterance. Similarly to Munro and Derwing, Smith and Nelson (1985) distinguish between these two concepts but in association with different entities: 'intelligibility' refers to the ability of the listener to recognize individual words or utterances, while 'comprehensibility' refers to the listener's ability to understand the meaning of the word or utterance in its given context.

In this way, the range of work by Munro and Derwing, and Smith and Nelson elucidates the importance of the distinction between intelligibility and comprehensibility because, to them, being able to do well with one component does not ensure doing well with others (Munro & Derwing, 1995). Nelson (2008, p. 302) says that "comprehensibility can fail even when the degree of intelligibility between participants is high". The idea of discrepancies between recognising words and understanding the message is also supported empirically by Zielinski (2004), who found that listeners who could identify words accurately also puzzled over the whole message (cited in Yang, 2009). Matsuura et al. (2009) found that, although Japanese listeners could easily understand utterances in the varieties of English in their study, they could not transcribe the words correctly.

This relationship between intelligibility and comprehensibility sounds more reciprocal in the definition by Smith and Nelson than in that by Munro and Derwing. The latter suggest only a 'one-way' relationship, where the speech might be intelligible despite poor understanding (which is equivalent to understanding the speech with difficulty) but there is no route back in this relationship. In contrast, the definition of Smith and Nelson better explains the phenomenon of the message of speech possibly being understood despite drawbacks in identifying many of its individual words. The following quotation, in which Smith is speaking as an invited respondent to a paper given by Nelson in the early 1980s, sheds some light on this idea:

"We may find an argument intelligible but not comprehensible because of the way it was structured. It is not uncommon to hear people complain, 'What was he trying to say?' I don't think that refers to intelligibility of the speaker to the hearer but to the comprehensibility of the speaker's presentation." (Nelson, 2008, p. 301)

The definition of these terms by Smith and Nelson (1985) places these concepts at two different levels: intelligibility is limited to recognition of the individual words by which the speaker conveys his/her message, while comprehensibility is the ability to understand the message being delivered. At this level, comprehensibility acts beyond the boundaries of individual words by drawing in neighbouring words in the same utterance. In other words, the comprehensibility of the overall message can be enhanced through using the linguistic context to recognize words that might have been missed by the listener.

In listening and speaking classes the definitions of intelligibility and comprehensibility by Smith and Nelson (1985) may be more functional for three reasons. The first reason is the ability of their definitions to reflect the reciprocal relationship between recognising words and understanding the utterance (as mentioned above). This could explain how a learner might grasp the meaning of an utterance despite missing segmental features employing non-linguistic aspects (e.g., context, tone of speaker, and learners' expectations and knowledge). Secondly, the distinction between two levels of understanding (within and beyond word boundaries) facilitates error analysis in classroom teaching and makes instructions more directive and targeted. For example, the teacher could focus on individual phonological features when the goal is improving intelligibility, whereas when the goal is comprehensibility, more communicative activities and instructions for improving accommodation skills could be targeted. Nevertheless, some teachers might still prefer to integrate work at these two levels. Thirdly, Smith and Nelson's definitions of intelligibility and comprehensibility are commensurate with top-down and bottom-up processing. That is, in intelligibility, the learner is expected to recognize individual words relying on the words' phonological codes and employing bottom-up processing. If the required phonological input is insufficient for word recognition (so intelligibility is not achieved), the listener starts to investigate neighbouring words and linguistic context by implementing top-down processing and using their overall understanding of the utterance to predict what the missed word could have been.

Based on the above discussion considering the literature on top-down/bottom-up processing and the definitions of intelligibility and comprehensibility by Smith and Nelson (1985), Figure 3 was developed to visualize the relationship between these two contrasting processes and the intelligibility and comprehensibility of speech. In

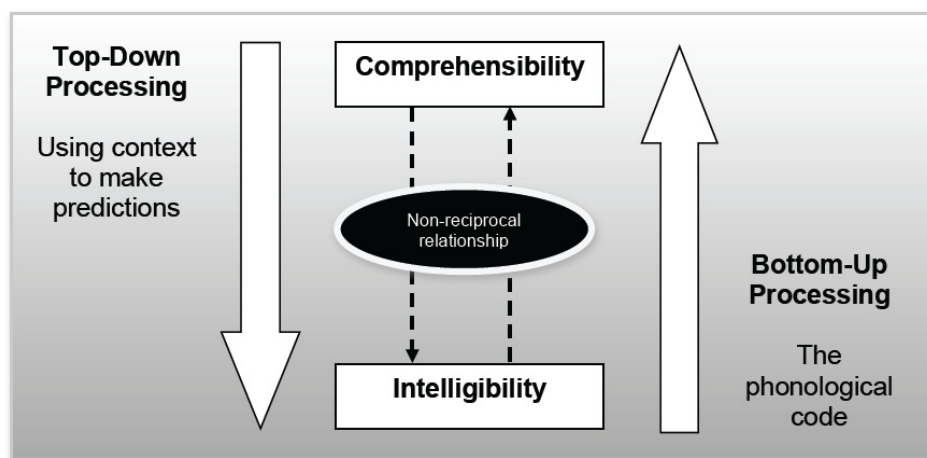this figure, the dotted arrows indicate the non-reciprocal relationship between intelligibility and comprehensibility.

Figure 3. Proposed relationship between intelligibility/comprehensibility and approaches to listening

5. Integration between Catford's Model and Top-Down and Bottom-Up Processing

Catford's model has been successful in providing a comprehensive overview of how speech is perceived by balancing acoustic and non-acoustic features. It also draws a clear distinction between two levels of understanding speech: recognition of words and comprehension of an utterance within its context. Although Catford's model only uses the term intelligibility to describe the successful completion of the identification and interpretation process (see Figure 1), it still distinguishes between recognition of acoustic features (or identification) and processing these acoustic features in relation to factors eventually leading to comprehension of the message within a specific context. Within these features there are two aspects that should be rethought in this model. In its current form, this model does not reflect the non-reciprocal relationship between intelligibility and comprehensibility, in which it is possible that words in speech might be individually recognizable but the listener might still hesitate over the utterance's meaning. In other words, it does not indicate that identification of words is not necessarily a prerequisite of understanding speech. Additionally, it does not introduce intelligibility and comprehensibility as two different notions but considers intelligibility the terminal point which describes the extent to which speech has been communicatively successful, and identification of speech has to precede successful interpretation. In this sense, Catford's model can incorporate the two types of listening processing, top-down and bottom-up processing, with the latter being given favourable consideration due to its importance in teaching listening to NNSs. Based on this logic, Catford's model is revised and presented below in Figure 4.
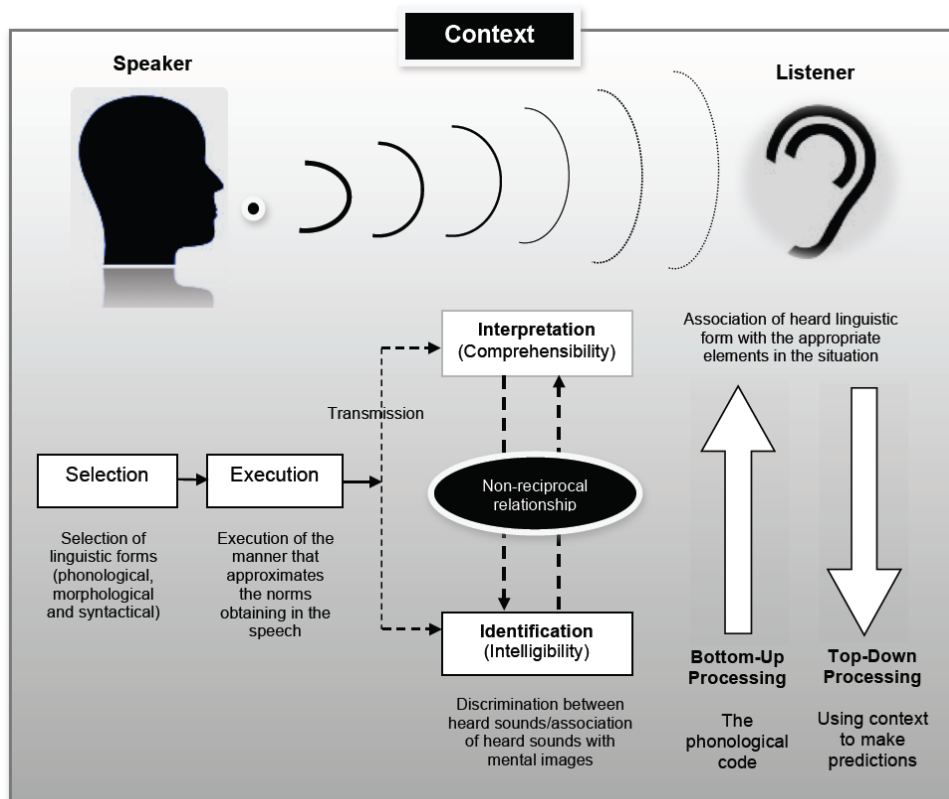
Figure 4. Revised version of Catford's model

In the revised version of this model the speaker *selects* the linguistic features and then *executes* them in a manner that is expected to approximate to what is considered appropriate in a specific context. After transmission of speech, the hearer receives the utterance and processes it in one of two ways. The first involves recognition of individual words, and with this the listener starts employing bottom-up processing by looking at segmental features and then moving up gradually to process these features in order to understand the meaning of the larger message. The second possibility is comprehension of an utterance in a way that may not necessarily mean that individual segments were recognized (or intelligible). Through comprehensibility the listener can enhance recognition of an individual utterance by employing top-down processing, which facilitates anticipation of what has been missed in utterances. During classroom teaching, the focus can be on the phonological code and the pronunciation of individual utterances when the target is 'intelligibility' and employing bottom-up processing, whereas the focus can be on context and on employing top-down processing when the purpose is comprehensibility of speech.

6. Conclusion

The purpose of this work was to rethink Catford's model and propose an enhanced model while providing a theoretical basis for speaking and listening classes, taking into consideration two main areas in the literature about speech perception. The two ways of processing speech, bottom-up and top-down processing, were considered. These are bordered by two entities which also incorporate two levels of understanding: intelligibility and comprehensibility. Bottom-up processing is associated with intelligibility, and this refers to the listener's attempts to recognize individual utterances or words by relying on the phonological code of the utterance This contrasts with top-down processing, which is associated with comprehensibility, in which the listener may not recognize individual utterances but might still grasp the meaning of the utterance by relying on context and background knowledge rather than the phonemes of individual words. The revised version of Catford's model provides an explanation for the functional role of the top-down and bottom-up processes in perceiving the intelligibility and comprehensibility of speech in speaking and listening classes. The model also has implications for the design of activities to help students practise these two modes of processing.

References

Anderson, A., & Lynch, T. (1988). *Listening*. Oxford: Oxford University Press.

Brown, G. (1990). *Listening to Spoken English*. Harlow: Longman.

Catford, J. (1950). Intelligibility. *ELT Journal, 1*(1), 7-15. http://dx.doi.org/10.1093/elt/V.1.7

Clark, J., & Yallop, C. (1995). *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.

Derwing, T., & Munro, M. (1997). Accent, Intelligibility, and Comprehensibility. *Studies in Second Language Acquisition, 19*(1), 1-16. http://dx.doi.org/10.1017/S0272263197001010

Howatt, A. P. R., & Dakin, J. (1974). Language laboratory materials. In J. P. B. Allen, & S. P. Corder (Eds.), *Edinburgh Course in Applied Linguistics: Techniques in Applied Linguistics*. London: Oxford University Press

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.

Kenworthy, J. (1987). *Teaching English Pronunciation*. London: Longman.

Matsuura, H., Chiba, R., & Fujieda, M. (1999). Intelligibility and Comprehensibility of American and Irish Englishes in Japan. *World Englishes, 18*(1), 49-62. http://dx.doi.org/10.1111/1467-971X.00121

Munro, M., & Derwing, T. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning, 45*(1), 73-98. http://dx.doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nelson, C. (2008). 'Intelligibility Since 1969' *World Englishes,* 27(3), 297-308. http://dx.doi.org/10.1111/j.1467-971X.2008.00568.x

Pinker, S. (1994). *The Language Instinct: The New Science of Language and Mind*. London: Penguin. http://dx.doi.org/10.1037/e412952005-009

Smith, L., & Nelson, C. (1985). International Intelligibility of English: Directions and Resources. *World Englishes, 4*(3), 333-342. http://dx.doi.org/10.1111/j.1467-971X.1985.tb00423.x

Zielinski, B. (2004). Measurement of speech intelligibility: What are we actually measuring? Paper presented at the annual meeting of the American Association for Applied Linguistics, Portland, OR. In J. Yang, (2009). *English Phonological Intelligibility: A Phono-Numerical Measure.* Retreived May 25, 2015, from http://teacher.yuntech.edu.tw/yanght/cv/phono-measure-reference.pdf