Construing negotiation
The role of voice quality features in American-Filipino business telephone conversations

Jenny Yau-ni Wan
Hong Kong Shue Yan University (China)

The call centre conversation is a telephonic exchange of voices between the customer and the customer service representative (CSR). Both lexicogrammatical and prosodic features are used to construe emotional and attitudinal recognition. Studying these features can investigate how the call centre discourse is construed, and how the interpersonal meaning takes shape through the text. The spoken data are constructed by Filipino CSRs and American English-speaking customers. The findings show that participants tend to make specific paralinguistic voice quality choices to express their emotions in dialogue. This article first discusses the voice quality framework for its semiotic features in relation to interpersonal meaning, reviews previous voice quality studies and later delineates how voice quality relates to interpersonal meaning in the calls.

Keywords: voice quality features, spoken discourse, business telephone conversation, negotiation, interpersonal meaning

1. Introduction

Call centre conversations may include calling a telephone centre to enquire about a service or product, or may involve a wide range of customer service activities and goals. The customer service representative (henceforth CSR) must maintain a positive attitude at work (Lee 2006). Apart from positive lexicogrammatical choices, good interpersonal meaning cannot be achieved without the contribution of voice quality. However, in the literature the meaning-making process of voice quality is often interpreted as less systematic. Some call centre trainers often ask the CSR to sound “positive and sincere” and to avoid annoying and offensive expressions. However, voice quality features that result in creating interpersonal meaning of “positive and sincere” have yet to be defined and remain unexplored.
Often in such training or communication, voice quality terms are used with little or no definition. A common sense interpretation is left for the CSR to understand and act upon. Unfortunately, the call centre trainers focus more on pronunciation and accent trainings. Thus there is a strong need to seek a reliable and systematic analytical method for analysing voice quality.

Interpersonal meaning is a key resource in dialogic call centre conversations. Interpersonal meaning is defined as intersubjective meaning between speakers, and it correlates with tenor (Halliday 1994; Martin 2001). In the study of interpersonal meaning, I draw on Systemic Functional Linguistics (SFL) theory as SFL allows the researcher to understand how intersubjective meaning is developed in a text within the clause and across the text and how such meanings are related to social activity. Our participants in call centre interactions are mainly Filipino CSRs and American customers. The present study has a research question What relative changes in voice quality features can be identified as key resources for construing negotiation in call centre conversations? The aim of the present study is to find typical interpersonal features which are used to make meaning at points of negotiation by conducting a voice quality analysis.

2. Literature review

Voice quality features are paralinguistic and nonverbal resources (Martin 2007; see Leijssen 2006). Crystal (1969) suggested that the listener can recognise the speaker’s identity through his/her voice quality. Voice quality analysis is interpreted as a conceptual analysis (van Leeuwen 1999). Meanings can be conveyed through the nonverbal behaviour of the face, body and voice quality (Hall and Friedman, 1999). Voice quality is usually described and classified by qualitative descriptions such as rough, warm, creaky, dull and so on (Titze and Story 2002). Voice quality has been studied in different fields and has been viewed to have significant values. Many studies establish a strong association between voice quality and attitudinal and interpersonal meaning. The voice provides clues to the speaker’s emotions (Jones and Jones 1990; Ko, Judd, and Blair 2006; Yogo, Ando, Hashi, Tsutsui, and Yamada 2000). Some previous research focused on the emotion of speech sound (phonology) by first creating emotion word lists, for example, word lists expressing the feelings of happiness, sadness, anger and fear (Dellaert, Polzin, and Waibel 1996; Gorin 1995; Plutchik 1994), then searching for words in the data according to the word lists. However, this kind of emotion categorisation was carried out for frequency distributions. These fail to generate the unfolding prosody development and the dialogic interaction in detail. To date, there are no published studies which examine voice quality in call centre discourse.
2.1 Voice quality framework

The present study attempts a semiotic analysis. In the past, semiotics was commonly understood as a code with clear definite meanings but not as having meaning potentials (van Leeuwen 1999, 10). However, sound analysis should not aim to create a code book with correct usage. This is because sound and voice quality are “semiotic resources offering its users a rich array of semiotic choices” (van Leeuwen 1999, 6). The significant value of the semiotics of sound is the interpretation of sounds in various situations. Describing semiotic resources provides a way to describe and explain the use of these resources (van Leeuwen 1999). Voice quality is a part of the semiotic resources of sound, and thus creates different meaning potentials in different contexts. Voice quality offers a set of semiotic resources for meaning-making (van Leeuwen, 1999). Voice affects the interpretation of interpersonal meaning. For instance, when a speaker talks in a slow, clear voice with a level tone, his voice can create calmness and stability in the listener (Jewitt 2002, 9). On the other hand, a loud, quick and breathless voice can signal a sense of instability and remove clarity (Jewitt 2002, 9). In addition, an emphatic and deep voice can be used to signify threat or authority (Esling and Wong 1983).

If a person changes his or her voice when speaking according to different situations, the interpersonal meaning construed will vary. The “variant in pitch and loudness and changes in voice quality can fulfil different communicative goals” (Eriksson 1994, 48). For example, television anchors adjust their voices, microphone distance and broadcasting style according to different programmes (van Leeuwen 1982, 1984). A higher and tenser voice is used by the newsreader because it is a more formal genre while the host of a musical programme tries to energise the audience via the way s/he talks (van Leeuwen 1999). Differences in voice quality can signal differences in language functions and speaker roles. Therefore, a change in voice quality can lead to a change in interpersonal meaning between the speaker and the listener. Van Leeuwen (1999) developed a system network of voice quality. The seven key features of voice quality posited by van Leeuwen are tension, loudness, pitch register, roughness, breathiness, vibrato and nasality, as shown in Figure 1. Van Leeuwen (1999) stated that in the human voice tension is recognised by the tense and lax aspect of the throat muscles. “The voice becomes higher (lower overtones are reduced, higher overtones increased), sharper, brighter and above all, more tense” (van Leeuwen 1999, 130). The voice can create meaning potentials such as “aggression”, “repression” and “excitement” (van Leeuwen 1999, 131). On the contrary, a lax voice is produced when the speaker opens the throat and relaxes the voice (van Leeuwen 1999). Soft and Loud voice are critically associated with social distance (van Leeuwen 1999, 133). Loud voice functions to extend more
boundaries while soft voice can be used to present intimacy or confidentiality (van Leeuwen 1999, 133). Loudness and softness are frequently used in call centre spoken data. For example, negative lexicogrammatical choices combine with loudness to form what the present study has termed “hot anger”; with softness they form what can be called “cold anger” in interpersonal meaning. In addition, pitch register relates to sound quality as high and low (van Leeuwen 1999, 134). Meaning potentials created by different pitch levels are significantly related to gender, age, context and the overall purpose of the talk (van Leeuwen 1999). Meaning potentials created by pitch register can be very rich, for example, a low voice associated with softness, such as mumbling, may function to indicate lack of dominance; while a low voice associated with loudness can demonstrate power and authority to the listener (van Leeuwen 1999). In the call centre data, there are examples of a female customer who speaks low and soft in order to project intimidation. Lastly, breath is defined as “an extraneous sound mixes in with the tone of the voice itself” (van Leeuwen 1999, 133). This voice quality feature is very soft and always associated with intimacy in informal situations such as sensual advertisements (van Leeuwen 1999, 133). However, in call centre exchanges, breathy voices cannot be found, but the sound of sighing, such as audible breaths, is frequently found, associated with pressure being releasing by the CSR or the customer. The present study builds on previous studies, aiming to provide a comprehensive interpretation of the creation

Figure 1. *System Network of Voice Quality* (van Leeuwen 1999, 151)
of interactive meaning through voice and lexicogrammar in call centre discourse. Voice quality features are defined as: (1) categories of volume, pitch, tension and rhythm; (2) comprised of multiple categories, where a voice quality feature can be soft, low and lax; and (3) not being limited to syllables, but also including clauses and turns. By working with the above definition of voice quality, I am able to systematically analyse the impact of voice quality in complex complaint calls.

2.2 Phonology – intonational system in SFL

Patterns of intonation and rhythm are called the melody and measure of speech respectively (Halliday 1989; Steele 1975). Intonation is realised by tone groups. The unit below the tone group is the Foot which is the basic unit of rhythm (Halliday 1989, 50), while the structure of the unit below the Foot is the syllable (Halliday 1994, 292). Rhythm is the syllable pattern which imposes the “beat” of language (Halliday 1989). English is a “foot-timed” or “stress-timed” language (Halliday 1989, 50). Feet in English are of similar length (Halliday 1989). Foot structure is realised by Ictus and Remiss: Ictus (^Remiss) (Halliday 1994, 9). Here ^ represents sequence and ( ) means optional component. A slash / is put at the beginning of each foot (Halliday 1994). “The syllable immediately following the slash is the SALIENT Syllable, the one carrying the beat (the ‘ictus’, in metric terminology)” (Halliday 1994, 293). Therefore, one foot must contain an Ictus (salient syllable) which is realised by at least one syllable, while Remiss (non-salient) is optional (Halliday and Greaves 2008).

An example / why are there / more / floods in / houses in the / basement / is taken from Halliday’s (1989) book to illustrate differences in between “no. of syllables in foot”, “relative duration of feet” and “actual timing for utterance” (pp. 50–51). Halliday (1989) suggests that if strict tempo is followed, all feet will have same length. The number of syllables of the example is:

/ why are there / more / floods in / houses in the / basement /

3 1 2 4 2

In ordinary spoken language, the timing of each syllable can be faster on average. The relative duration of feet is:

/ why are there / more / floods in / houses in the / basement /

1.4 1 1.2 1.6 1.2

However, the actual duration of feet in seconds was:
Here, the syllables can contract or expand based on the weak or strong arrangement (Halliday 1989, 51). That means the speakers can speed up slow down the utterance according to the meanings of weak or strong they intend to make. Rhythm is one of the important categories in the call centre voice quality system. The customer and CSR frequently employ different rhythmic patterns in the negotiation process. When a customer speaks loud and fast, hot anger can be formed while when a customer speaks a turn very slowly, sometimes even separating and expanding every syllable, he or she may wish to project his or her cold anger by creating social distance. Lastly, a computer software program was also used to facilitate the identification of voice quality features in the present study. *Praat* (means “talk” in Dutch) was the computer program adapted to identifying a relative change in the sounds used in the present study. It is a free software program, and widely used by linguists and phoneticians to analyse speech sounds (Halliday and Greaves 2008). *Praat* can be downloaded and installed from [http://www.fon.hum.uva.nl/praat/](http://www.fon.hum.uva.nl/praat/). Selected examples of sound files were supported by using *Praat*. *Praat* was able to indicate, for example, decibel (dB), intensity, frequency, and the range of energy in spectrogram.

### 3. Methodology

The data for the present study were collected from call centres in Manila (The Philippines). The spoken data are audio-recordings of conversations in inbound commercial customer-service telephone enquiries between Filipino CSRs and American English-speaking customers. An inbound call centre mainly deals with phone-in calls, and the CSRs answer enquiries (Jones 1999). The present study uses a corpus of 20 complex call centre conversations, comprising approximately four hours of talk, which were selected from about 2,000 English-language calls to an insurance call centre. Such a focused study has enabled a detailed investigation of specific voice quality interpersonal features which were realised in the negotiation process of the call centre conversations.

### 4. Findings and discussion

Van Leeuwen’s (1999) framework was used as a starting point in the present study to look at human voice quality in the call centre discourse. In this section, I
identify voice quality realisations in the text and then discuss how these function in the data. Four voice quality features, namely volume (Loud/Soft), pitch (High/Low), tension from the muscle of the throat (Tense/Lax), and rhythm (Slow/Fast) emerged in the call centre data. Transcription conventions used to describe voice quality and related features are as follows:

Volume (Loud / Soft, dbs)
- "word" a passage of talk that is softer than surrounding talk
- WORD a passage of talk that is louder than surrounding talk

Pitch Register (Low / High)
- ↓ marked falling shifts in pitch that is lower than surrounding talk
- ↑ marked rising shifts in pitch that is higher than surrounding talk

Rhythm (Slow / Fast)
- > word < talk is faster than surrounding talk (fast rhythm)
- < word > talk is slower than surrounding talk (slow rhythm)

- :: an extension of a sound or syllable

Tension (Tense / Lax)
- ~ word ~ talk is laxer than surrounding talk by relaxing the muscle of your throat
- + word + talk is tenser than surrounding talk by relaxing the muscle of your throat

[pause – x secs] timed intervals in seconds show length of silence

4.1 Voice quality feature – volume (loud/soft) in call centre conversations

Volume refers to loudness and softness in the voice of the CSR or the customer. Volume is measured by decibel (db). Figures 2 and 3 illustrate examples of Loud and Soft as shown by Prat. Prat measures the db level of the speaker by importing the sound track. The db level can be referenced by measuring the intensity, indicated by the circled areas in Figures 2 and 3 for the intensity (green line) for the dbs. The black arrow indicates the highest dbs along this sound track.

The green line refers to the intensity, shown in the circled area, and variants in the green line illustrate the relative change in the loudness. This example is of an explicit change in loudness in the call centre conversation. The range of the dB is indicated by the numbers on the side, where 0dB indicates absence of loudness, such as silence, and 100 dB being very loud. Figure 2 shows that the customer says No, I don’t (Transcript 6, turn 29), and her corresponding db is 81.17.

The sound track Soft was also imported in Prat. The CSR says I’m trying to listen (and) everything that you are trying to say (Transcript 6, turn 9). As shown in Figure 3, the highest dbs in this sound track is 65.47, indicated by the black arrow. The green line varies in a wavy way which means the speaker varies the decibels
along the utterances. Volume has a link with social distance (van Leeuwen 1999). Loudness is a factor essential for transmitting message (Möller 2000). A loud voice can expand the territory (van Leeuwen 1999). As shown in Example (1), turns 46–51 provide examples of higher levels of loudness and faster speeds found in the data.
Example 1. **Voice quality analysis: Faster speed and Loudness** (Audio file: Call 1_4:17-4:48)

R = Female CSR  
C = Female Customer  

[Generic stages: Servicing and Objection]

46 C1: because I \*ASKED+ “do I \*NEED+ to sign the form, do I NEED to+ write the letter” and the lady I \*SPOKE+< with, her name is Kelt.

47 R1: ↓°Mhm°↓

48 C1: SAID “NO” she would do it for me and THAT’S ALL

49 R1: That’s right all you need to do = = is

50 C1: = = >WHAT HAPPEN WHAT HAPPEN IF NEXT MONTH YOU GUYS DRAW MONEY< again because that’s unacceptable

51 R1: We \*WON’T DRAW+ any money out, maam, now that your request is already placed on the system

As shown in turn 49, when the CSR (R1) is describing the action which the customer (C1) needs to take, the customer (C1) interrupts in turn 50, by saying *what happen what happen if next month you guys draw money again* in a louder voice at a faster speed than the surrounding discourse. The pace here is faster than in previous turns, and C1’s voice also becomes louder. In the call centre discourse, rhythm differences are frequently noted between the CSR and the customer. I observed that a faster rhythmical voice is often associated with loudness. Sacks et al. (1974) indicated in their seminal study that the combination of faster rhythm and loudness functions to obtain speakership and that sometimes these result in self-selection at the time of turn taking. In Example (1), when the customer speaks Fast and Loud, the CSR immediately gives up speakership in turn 49. The customer is very keen to occupy speakership which sounds more powerful.

Loudness can be associated with repetition to form hot anger. Hot anger refers to the overheated conversation in which a speaker dominates with active emotion (Scherer 1986; Wallbott, 1998; Wehrle, Kaiser, Schmidt, and Scherer, 2000). Hot anger realised in voice quality features of high pitch, tension and loudness can be identified in arguments in the 20 complex calls, as shown in the Objection stage in turns 44 to 48, in Example (2).

Example 2. **Voice quality analysis: hot anger** (Audio file: Call 8_5:50-6:14)

R = Female CSR  
C = Male Customer  

[Generic stage: Objection]

44 C8: what happens to insurance that you pay the same amount every month for the rest of your life. +I DON’T UNDERSTAND WHY DON’T HAVE THAT KIND OF = = [POLICY+

45 R8: = = Ok, sir
The customer (C8) in this call is very frustrated and angry about his insurance payment. In turn 44, he uses loudness and tension which are reinforced and realised through negative polarity, *I don’t understand why don’t have that kind of policy*. He also chooses to use Loud and Tense voice in turn 46, *gone up several times* and in turn 48, *thousands of thousands of thousands of dollars* which *several times*, and *thousands of thousands of thousands of dollars* can be categorised as intensifier. Repetition itself is already a quantification item building up prosody of force. In this example, together with the change in loudness and tension, voice quality force is added to the repetition of lexicogrammatical items, all combining to construe a clear realisation of frustration. This kind of combination doubly graded up complaint, and I have termed this occurrence “hot anger” in the present study. In the data, repetition examples are frequently found. Through using repetition of complete phrases, clauses or clause complexes, the speaker aims to generate longer turns and to reinforce the negative attitude being constructed. In Example (3), C6* referring to the customer’s (C6) wife, says in turn 39, *years years years and years ago* in a raised tone which forms a prosodic pattern. The lexicogrammatical feature *years years years and years ago* can be categorised as repetition. This quantification resource where the repeated prosody grades up the intensity.


R = Female CSR  
C = Female Customer  
[Generic stage: Servicing]

38 C6: under EIC = =  
39 C6*: = = Wait ma’am. Here I’m gonna let you speak to my husband ah again. But I don’t know John Smith and I that I’m certainly would like, eh, I mean that the person in the family whose name was John Smith guys, ah ↑YEARS YEARS YEARS AND YEARS AGO↑ when he was closed 100 years old. So I don’t know what you are doing there. But I’m going to let you speak to my husband, but you certainly don’t give, ah, me any confidence in your company.

Another example of repetition can be found in Example (4), Transcript 6 in turns 27 and 29, *No, I don’t*. This short expression, *No, I don’t*, closes up the space for modality, and no space is available for negotiation. The customer shuts down the opportunity of prolonging the conversation following the two questions asked, *do you know John Smith? and You don’t have any recollection with Mr. Smith? No I don’t* is a monoglossic emphatic answer. A monoglossic answer is interpreted as
a single voice which leaves no room or opportunity for further negotiation, and increases the intensity. The inscribed meaning potential of No, I don’t refers to Let’s move on to the next topic.


\[\text{R = Female CSR} \]
\[\text{C = Female Customer} \]
\[\text{[Generic stage: Servicing]} \]

26 R6: to know John Smith?
27 C6*: ↑ NO, I DON’T ↑.
28 R6: You don’t have any recollection of Mr. Smith?
29 C6*: ↑ NO, I DON’T ↑.

On the field visits, I found that the call centre trainers were often labouring under a misconception, believing that the customer tends to shout and use foul language to express their dissatisfaction in a complex complaint call. However, this is not always the case. The data show that the customer may use other resources, such as using voice quality of Softness, to express their frustration. If the conversation becomes overheated, hot anger is created (Scherer 1986; Wehrle et al. 2000), with the speaker dominating by what is interpreted as an “active emotion” (Wallbott 1998, 887). It is possible to identify softness features in the argument in Example (5). In Example (5), turn 33, the CSR (R1) raises the volume of the voice to create a counter expectancy, such as but you already, was informed you. At this particular moment, the customer (C1) gives up Loudness but chooses to use softness in her turn to continue to express her anger. This different way of expressing anger is termed cold anger. Cold anger is marked as a contrast to the active negotiation (Banse and Scherer 1996; Scherer 1986). Sometimes when a speaker, for example C1, is disappointed or angry, a soft voice may be chosen. This can be interpreted as cold anger which can sound more threatening than if they were expressing their emotion directly through loudness and shouting. This is because when a speaker expresses cold anger, s/he may be seen as having a greater negotiation power due to prior careful consideration and having corresponding solutions. Cold anger is a marked choice as the customer is inferring, for example, “I know I can shout. You know I am angry, but I’m keeping myself cool, stating this clearly. So that you cannot say that I am an irrational customer. I am in control of my emotions well.” By selecting cold anger, customers aim to enhance the persuasive nature of their arguments.

Example 5. Voice quality analysis: Soft voice graded by repetition (Audio file: Call 1_2:42-3:00)

\[\text{R = Female CSR} \]
\[\text{C = Female Customer} \]
\[\text{[Generic stage: Objection]} \]
The customer used a soft voice to express cold anger for only a few seconds. It is the contrast of relative pace, or a high loud voice that impacts the importance of the realisation of the message. It would be almost impossible for the customer to use a soft voice to express dissatisfaction throughout the whole call from the beginning till the end. On the contrary, the CSR can adopt softness for a longer period of time in the call because he or she is trained to maintain a professional image and to calm the angry customer as part of her/his work. The CSR will become frustrated too. However, the CSR cannot show this frustration explicitly as it would impact his/her work, especially if it occurred on a regular basis. The emotion of the CSR may also be realised via changes in the voice quality level, such as volume, tension, higher pitch or even through audible breaths. Audible breaths, only air voices, are frequently produced by CSRs in the conversation. The function of the audible breaths is to express their frustration and pressure implicitly.

The constant interplay between volume (Loud and Soft) indicates that the interaction between the CSR and the customer is not simply seen as an individual or single expression but the text unfolds as an interactive and dynamic process. Friginal (2007, 2010) suggests that studying prosodic development can enable the CSR to improve the clients’ perception of the service encounter and to enhance competence to handle the conversation. Example (6) shows how the use of loud and soft voice quality features can result in actualization.

Example 6. Voice quality analysis: interaction between Loud and Soft (Audio file: Call 6_0:59-1:19)

R = Female CSR
C = Female Customer

[Generic stage: Legitimization and Clarification]

8 C6: … And tut… uh…I don’t know I don’t know what you guys are doing? (You don’t know what you’re) doing either [laugh, along the clause]... Are you THERE?
9 R6: I’m °still° here sir. I’m °trying to listen° (and) °everything° that you are °trying to say.
10 C6: I’m °NOT TRYING°+ to say, I’m °SAYING°+ it. You sent this to °US°+.

As shown in Example (6), in turn 9 the CSR (R6) says, I’m trying to listen (and) everything that you are trying to say. However, the customer (C6) replies in turn 10, I’m not trying to say, I’m saying it. You sent this to us. In the above case, this is not simply a problem created by low or high modality. This is a problem related to fulfilment and realisation. In turn 9, the CSR says trying to listen to... which refers
to the description of her mental process. It is acceptable for her to describe her own ability of handling the problem. However, when the conversation continues, she says ... *what you are trying to say* indicating a problem. The expression *trying to say* refers to the customer saying something in an incomplete manner. This phrase is equal to a critique of the other’s ability. The CSR is downplaying the customer’s ability. As a result, it is a critique of the customer’s lack of fulfilment. However, it could also be possible that the CSR here simply makes an error and does not fully understand the meaning she has made. Hence the customer (C6) immediately replies in turn 10, *I’m not trying to say* in order to repair this critique of lack of fulfilment. When the conversation continues, the customer says, *You sent this to us*. The voice used in the sentence is Loud, High and Tense. The customer neglects the modality and voice quality features to make it definite, and to become more monoglossic for closing down further space of negotiation. A complex call has rich resources of High and Loud while a general call has rich resource of Low and Soft.

4.2 Voice quality feature – pitch (high/low) in call centre conversations

Pitch refers to High and Low voice quality features in the present study. Generally, humans’ audible frequency range is within $20Hz < f < 20000\ Hz$. The upper and lower frequency preset in *Praat* is $50Hz$ to $500Hz$. Figures 4 and 5 are examples of High and Low pitch as displayed by *Praat*. The Hertz (blue line) represents the frequency of the pitch.

**Figure 4.** Example of High Pitch, as displayed by “Praat”
The level of Hertz is circled, and the black arrow indicates the highest Hertz on this sound track. Figure 4 shows that the High voice is of 284.6Hz. The Hertz line is also the pitch of the corresponding syllable. The two black lines represent the sound track of the audio file.

![Image 64x328 to 360x540]

**Figure 5.** Example of Low, as displayed by “Praat”

In Example (7), a Low voice of 158.5Hz is indicated by the black arrow in Figure 5. Example (7) shows a male customer mimicking another’s voice through changing his pitch to hide his self-identity and masquerade as his wife in the conversation. The flow of the interaction in Example (7) is related to the customer’s disclosure of sensitive information of the registered client to provide identification for the CSR. The husband is pretending to be the registered client, i.e. his wife. The CSR (R17) insists that she can only provide the information to the wife of the customer (C17), such as in turn 77, *yes, but I am not allowed to give out any policy value other than to your wife*. However, the caller’s wife is not with him at that moment. The caller feels extremely frustrated and inconvenienced.

**Example 7. Voice quality analysis: High and Lax voice to mimic** (Audio file: Call 17_6:38-7:25)

R = Female CSR  
C = Male Customer

[Generic stage: Servicing and Objection]

77 R17: *yes, but I am not allowed to give out any policy value other than to your wife*

78 C17: ↓Jesus Christ↓
In turn 77, the CSR (C17) states that disclosing sensitive policy information to a third party, i.e., someone other than the policy holder, is not permissible in her line of work. She closes down the space of modality and stresses that the only person from whom he can receive the information is his wife. In turn 78, the caller (C17) expresses his disappointment. He uses a direct explicit marker expressing anger Jesus Christ. The CSR emphasises the policy again and tries to seek alignment from the customer through choices of high modality need and capacity understand we have this laws passed cannot in turn 79. However, the customer immediately closes down her intention, and reduces any possibility for discussion Hang on just a second. The customer totally refuses to listen to the explanation. Instead, the customer completely disregards the policy in the next turn. He tries to lie to the CSR by mimicking a female voice in a very obvious manner (High and Lax voice quality features), yes, this is Mrs. O’Connor, can you give me that information, are you going to give me the information? (turn 80). The customer tries to falsify his identity by mimicking a female voice, but in a rather obvious manner. This creates a serious level of dissatisfaction. Generally, male voices have a lower pitch than female voices. To mimic the voice of the opposite sex, men tend to use a higher and more lax voice feature. The male customer, C7, disregards the law and shows complete disrespect to the listener. As discussed above, the CSR uses the short expression, No, I’m sorry in turn 81 to reject the request firmly by closing down the space of modality. This implies high obligation and stresses that the only person who can receive the information is his wife. After seven seconds of silence, the customer gives up by using an interpersonal metaphor would you do me a favour? There is also a change from a high mimicking voice back to his normal pitch. To interpret his meaning from lexicogrammatical and voice quality features, would you do me a favour?, is not an optional request but a command which carries a sarcastic meaning. This is a phrase one usually asks of a close friend. Indeed the situation as well as the way it is said in this call suggests that the customer has relented and can no longer insist; instead he requests a favour from the CSR. Do me a favour hints at
sarcasm. Both the lexicogrammar and voice quality features of the customer (C17) are consistent in constructing a strong negative emotion signifying dissatisfaction.

A low voice can be used to show despondency such as when a customer uses a low voice, without much positive emotion, to say thank you in a monotonous manner. Such an utterance is devoid of intonation and is seen to construe irony in call centre conversations. It is a marked choice used to downgrade the satisfaction level. As shown in Example (8), the customer uses a monotonous low voice to say thank you very much to the CSR in turn 104 at the end of the whole conversation. The voice totally offsets the gratitude that is usually carried by the expression thank you.

Example 8. **Voice quality analysis: Low voice** (Audio file: Call 5_12:44-13:43)

<table>
<thead>
<tr>
<th>R = Female CSR</th>
<th>C = Female Customer</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Generic stages: Objection, Servicing and Closing]</td>
<td></td>
</tr>
<tr>
<td>102  C5: This is not right hahaha he’s just going through a divorce. And he has no money, erm, now &gt;who’s supposed to pay his taxes&lt; My husband or him?</td>
<td></td>
</tr>
<tr>
<td>103  R5: (audible breaths) Ok, hold on just a second. [pause – 10 secs] ok maam I did verify that one in and since Carl is the owner Central he is the owner I mean ↓he will be the one who’s going to shoulder the taxes. Carl Carl is the owner so he should be the who’s one going↓ to pay for the taxes [pause – 9 secs]</td>
<td></td>
</tr>
<tr>
<td>104  C5: (audible breaths) ↓Thank you very much↓</td>
<td></td>
</tr>
<tr>
<td>105  R5: You are welcome, maam.</td>
<td></td>
</tr>
</tbody>
</table>

Generally, a normal and fruitful thank you with a raise in tone and pitch is categorised as positive interpersonal meaning. However, in Example (8), the meaning potential of this monotone expression of gratitude is superficial. This realises sarcasm, and what the speaker really means is that the CSR or the company has not provided “real” help at all. The customer is not sure if the CSR deserves to be thanked. In the data, I found that when the customer feels that there is no possible further action that can possibly be carried out, they will close the conversation by using a low voice to indicate they are not satisfied and are basically ready to give up. The low pitch of thank you actually means “I’m despondent”. This analysis reveals that when a Low voice combines with gratitude, the level of positive inscribed attitude between speakers will be graded negatively.

A similar case can be found in Example (9). The CSR (R8) provides information to the customer that 35,000 dollars will be the minimum price set for his policy. In turn 105, the customer (C8) immediately discovers a very difficult fact that he may need to give up one policy if he cannot afford it. C8 says oh boy in a Low, Slow, and Lax voice in order to show his despair. He continues by saying, so I guess if I can’t afford both I m gonna drop one of them, right?
Example 9. Voice quality analysis: Low and soft voice (Audio file: Call 8_11:50-12:19)

R = Female CSR
C = Male Customer

[Generic stage: Servicing]

104  R8: >Thank you< thank you for staying on the line. My apologies to inform you this BUT 35 dollars that is the minimum that >I mean< 35 000 dollars that is now the <MINIMUM> that we can give you. <WE can no longer LOWER that death benefit >...

105  C8: <~↓↓Oh boy↓↓~> so I guess if I can’t afford both I m gonna ↓drop one of them↓, right?

106  R8: (audible breaths) Well, ↓that would be your decision sir↓.

In the next turn, R8 produces audible breaths and picks up a Low voice to say that would be your decision sir. In this case, it is apparently reflected in their voice quality features, Low voices, that both the CSR and the customer are forced into a bad decision or situation that they are not happy about.

4.3 Voice quality feature tension (tense/lax) in call centre conversations

The voice quality feature tension is created when one tenses the muscles of one’s throat during speaking (van Leeuwen 1999). Figures 6 and 7 give examples of Tense and Low as displayed by Praat. The spectrum (black area) displays the energy level, representing tension. The more energy that is released from the vocal band, the darker the colour shown. Different levels of spectrum form black lines, and more lines means stronger tension.

As shown in Figure 6, in the circled area of this example of Tense there are black four lines formed. Within this spoken utterance of aware of, several layers are formed. These layers mean the tension existed at different Hz.

In Figure 7, two lines are formed in the circled area. The grey and black areas represent the energy level of tension released from the vocal band along the Hertz scale.

The tense voice can create tension for the listener and make the listener become nervous, fearful and angry (van Leeuwen, 1999). Example (10), turns 15 to 16, illustrate Slow rhythm and tension found in the data. Generally, the customer and the CSR may speak slower to achieve clearer transmission of information in telephone communication due to the absence of visual cues. However, when a customer intends to emphasise a particular piece of information, he may also use a Slow voice; sometimes he even expands the syllables or words. The pronunciation of particular syllables and words is more elongated than usual, such as I was not aware of it in turn 16. In this turn, C2 pronounces each word very clearly and as an independent unit, with each unit being extended beyond the normal word length found in previous turns by the speakers.
Example 10. Voice quality analysis: Rhythm and Tension (Audio file: Call 2_1:03-1:28)

R = Female CSR
C = Female Customer

[Generic stages: Servicing and Legitimization]

15 R2: The policy ♦HAS LAPSED♦, Elle?
16 C2: Yes, but I ♦was not aware♦ of it. And I ♦got this notice♦ yesterday and is ♦there anyway♦ I can, what is that I have to do? I had this a long time and ♦I was not aware ♦I had to pay any more money on it.

In fact, the lexicogrammatical phrase of *I was not aware that* only includes negative polarity. If realisations are only investigated at the lexicogrammatical level, we are limited to seeing a distinct attitude embedded in this choice. However, when I listened to the sound file, I found an explicit tension associated with this phrase. This can be coloured to create a negative affect such as a defensive and self-protective retort to the CSR. The customer emphasises that she was not to blame. As a result, slow rhythm and tension of the speaker’s voice quality level grades up the negative affect of this particular turn.

Lax voice in call centre conversations carries a significant function, namely, to seek empathy. Example (11) is an example of the interplay of Lax and tension to form alignment in the negotiation process. In turn 40, C2 tells the CSR about her undesirable health *I am practically blind number one*. She uses a Low voice to say *and I can just barely make out these numbers so then*, and she continues to
Construing negotiation

remove the tension in her voice by using a very Soft and Lax voice I don’t read everything. These voice qualities are frequently used by the customer when providing an emotional and negative personal recount. This aims to seek understanding and solidarity from the CSR.

Example 11. Voice quality analysis: Lax (Audio file: Call 2_03:03-03:24)

R = Female CSR
C = Female Customer

[Generic stage: Objection]

40  C2: <Okay>, ah I I am practically blind number one ↓ and I can just barely make out these numbers so then ~°I don’t read everything°~ ↓.
  ↓WHEN I BOUGHT THIS↑+ I was +ASSURED+ it was going to be
  ↑FOREVER↑+. I had paid up policy.

However, when the conversation continues, the customer changes her voice by using a Loud, High and Tense voice to say when I bought this I was assured it was going to be forever. She makes some key terms such as assured (a Tense and Loud voice) and forever (a High, Loud and Tense voice) become prominent. This particular turn begins with a Low, Soft and Lax voice and grows into a Loud, High and Tense voice in the latter part. I believe the earlier part, the Soft and Lax voice, is a prelude to build up the focus at the end. As a result, the latter part, which is under focus, has become powerful, and carries stronger objections. The use of a range of voice quality features is clearly manipulated by the customer in Example (11).
4.4 Voice quality feature – rhythm (slow/fast) in call centre conversations

Rhythm is the study of the unit foot which is realised in patterns of syllables (see Halliday, 1989, 1994). In this section, feet in the utterances of CSR and customer in the negotiation are studied. The following figures are the examples of fast and slow rhythm as visualized by Praat. The intensity and rhythmic patterns are shown in the blue line (bigger circle) and the separate blue lines correspond to syllables uttered. The blue line represents the changing pitch. The smaller circle near the bottom of Figure 8 indicates the duration of the sound track analysed, so the sound track for this image is 3.248866 seconds (visible part) and 3.288 seconds (total duration). The shorter duration of 3.248866 seconds is the selected area that corresponds to each syllable uttered.

Figure 8. Example of Fast Rhythm, as displayed by “Praat”

The duration of the sound track which can be found in the smaller circle in Figure 8 is 3.248866 seconds (visible part). The present study suggested that the rhythmic patterns can be calculated as syllables per second. The duration of the sound track (in seconds) will be divided by the number of syllables. The maximum duration for Praat to process is 15 seconds each time. In Figure 8, the CSR in Transcript 10 says well one was already mailed out on January 2 have you gotten that yet in 3.288 seconds. There are 22 syllables in this utterance, and the speaker speaks about 6.7 syllables per second.
However, in Figure 9, the customer in Transcript 10 says *showing who the beneficiary is* in 2.976 seconds. There are 10 syllables, and this customer speaks more slowly than the CSR in Transcript 10, at a rate of only 3.4 syllables per second.

Rhythm measures the flow of time in the conversation (van Leeuwen 2005). The scope of rhythm can apply to syllables and feet. Changes in loudness, pitch and tension combined in different rhythmic patterns can result in a more notable voice in the conversation. Different rhythmic patterns are frequently used in call centre conversations to express various meanings. For example, the consumer and the CSR believe that it is possible to communicate with patience by using slow rhythm in speech (Dalton 2006). CSRs often resort to asking customers to repeat their unclear utterance as slowly as possible. In the opening stage of a call, the CSR may request the caller to repeat the name of the policy holder slowly or to spell out their social security number. In a complex call, slow rhythm may be used by the caller to express anger. Example (12) is a good example to illustrate how the CSR and the customer speed up or slow down the utterance in a complex call. Turn 10 is an event recount made by a customer who is grumbling to the CSR about a piece of delayed written information that was requested some time ago.

Example 12. *Rhythm example* (Audio file: Call 10_00:37- 01:16)

R = Male CSR  
C = Female Customer  

[Generic stages: Purpose and Servicing]

9  R10: How may I help you today?

10  C10: Peter I have been calling for almost a year now, trying to get someone to send me in writing < ::showing:: who the beneficiary is > on my policy. It had me as the owner and also the beneficiary. And that supposedly has been changed in your record. But I want something in writing to put with my um < ::information:: not policy > == that

11  R10: == > well one was already mailed out to you on January 2 have you gotten that yet? <

12  C10: No, it has not been == so

13  R10: == > perhaps it’s already on its way to you maam <

The customer varies her rhythmic pattern in saying *showing who the beneficiary is* in turn 10 and *information not policy*. She speaks in a very slow voice. Each word or syllable is almost separated. The number of syllables in the foot are:

<table>
<thead>
<tr>
<th></th>
<th>in writing</th>
<th>showing</th>
<th>who</th>
<th>the beneficiary is</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of syllables</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>proportionate duration</td>
<td>1.4</td>
<td>1.2</td>
<td>1</td>
<td>2.6</td>
</tr>
</tbody>
</table>
The foot, marked with a slash /, begins with a strong syllable in, show (in showing), who and the. However, in ordinary conversation, the proportion duration of the feet are calculated as “a two-syllable foot will be about one fifth longer than the one-syllable foot; a three-syllable foot will be longer again by a little bit less than a fifth” (Halliday 1994, 293). However, the actual duration of this utterance is as follows:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>0.7</td>
<td>1.3</td>
</tr>
</tbody>
</table>

The customer expands the foot, making it longer than the surrounding feet. A similar example is found where the speaker slows down the word information:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>(no. of syllables)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(proportionate duration)</td>
<td>1.6</td>
<td>1.9</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>(seconds)</td>
<td>1</td>
<td>2.1</td>
<td>1.3</td>
<td></td>
</tr>
</tbody>
</table>

“Information” is interpreted as the focus of her claim. As discussed above, commonly a caller will prolong their utterances to voice requests and to exaggerate the seriousness of the matter. However, the strategy of the CSR may well be just the opposite. The CSR in this example repairs the slow rhythm exaggeration by using the total duration of this utterance is 2.976 sec (s). The visible part is 2.8251496 sec(s) which is the selected area of this figure.

Figure 9. Example of Slow Rhythm, as displayed by “Praat”
a quick voice. In turn 11, *well one was already mailed out to you on January 2 have you gotten that yet?* and in turn 13 *perhaps it’s already on its way to you maam.* The voice quality features of these turns have a fast rhythm with some shortening. These are typical voice quality features in the call centre conversation at points of negotiation. The syllable patterns of these examples were analysed and showed the number of syllables in feet and their proportionate duration to be as follows:

| / well one was / already mailed out / to you on January 2 / have you / gotten that / yet |
|---|---|---|---|---|---|
| 3 | 5 | 8 | 2 | 3 | 1 |
| 1.4 | 1.9 | 2.6 | 1.2 | 1.4 | 1 |

The actual duration of the CSR’s utterance is much faster than its proportionate duration. The actual timing in seconds of his utterance is as follows:

| / well one was / already mailed out / to you on January 2 / have you / gotten that / yet |
|---|---|---|---|---|---|
| 0.6 | 0.7 | 1.1 | 0.2 | 0.5 | 1 |

In addition, there are no pauses between the words. This shows the eagerness of the CSR to provide an explanation. One more example of a fast rhythmic pattern is the following:

| / perhaps it’s / already on its way / to you maam/ |
|---|---|---|
| (no. of syllables) | 3 | 6 | 3 |
| (proportionate duration) | 1.4 | 2.2 | 1.4 |
| (seconds) | 0.2 | 0.8 | 0.5 |

The resources available for slow and fast work are frequently used in every call. Generally, fast rhythm will be used in a general call because the flow of information goes smoothly and fast with plenty of expected responses. In addition, fast rhythm can also be found in a complex complaint call when the customer and the CSR are both eager to explain. Sometimes, these two parties will compete for the speakership. Cold anger of speaking softly and slowly is found in a small proportion of calls. This is because the speaker cannot speak slowly for too many turns.

To summarize, when a CSR or a customer construes their attitudinal standpoint, they not only express their attitudes, but also dynamically negotiate and share their views and feelings. As demonstrated in the examples above, the interaction between the CSR and the customer is dynamic in terms of both lexicogrammatical and voice quality features, which are co-developed throughout. Lexicogrammar and voice quality play essential roles in construing interpersonal meaning in the call centre conversation.
5. Conclusion

In this article, I have attempted to demonstrate the ways in which I study how meaning unfolds and that interpersonal meaning patterns exist in voice quality. I have identified how lexicogrammar and voice quality features interact to construe the attitudinal and emotional flow of the call centre discourse between speakers throughout the complete text as it unfolds. The present study argues for the importance of the voice quality feature, which is found to be indispensable to construing interpersonal meaning in call centre conversations. Voice quality is very similar to phonology in several ways; however, they are not identical. Traditionally, call centre trainers make a serious effort to improve CSRs’ accent and pronunciation. However, it is clear that a successful call consists of a range of meaning-making features and is not dependent solely on what is commonly seen as knowledge about accent and pronunciation. Voice quality features are an innovative area which could be systematically taught in call centre language training sessions. Voice quality should be incorporated into recruitment, training and quality assurance measures within the industry. The findings from the present study hope to offer insights into the world of the global phenomenon of call centre discourse.

References


Jones, Peter and Peter A. Jones. 1990. “Stress: Are you serving it up to your restaurant patrons?” * Cornell Hotel and Restaurant Administration Quarterly* 31(3): 38–44.


Lee, Chipongian. 2006. “Poor English threatens Philippines outsourcing”. *South China Morning Post, Hong Kong*, p. 11.


Author’s address

Jenny Yau-ni Wan
Hong Kong Shue Yan University
LG500B, 10 Wai Tsui Crescent, Braemar Hill, North Point
Hong Kong
China
ynwan@hksyu.edu

Biographical notes

Dr. Jenny Yau-ni Wan is currently a Lecturer in Department of English Language and Literature at Hong Kong Shue Yan University. She graduated from the Department of English in The Hong Kong Polytechnic University. Her research interests include discourse, English for Academic Purpose and Systemic Functional Linguistics.