



- (51) **International Patent Classification:**  
*G10L 21/003* (2013.01) *G10L 25/15* (2013.01)  
*G10L 25/63* (2013.01) *G10L 25/90* (2013.01)
- (21) **International Application Number:** PCT/SG2015/050519
- (22) **International Filing Date:** 31 December 2015 (31.12.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:** 10201500040U 5 January 2015 (05.01.2015) SG
- (71) **Applicant:** CREATIVE TECHNOLOGY LTD [SG/SG]; 31 International Business Park, #03-01 Creative Resource, Singapore 609921 (SG).
- (72) **Inventors:** SIM, Wong Hoo; 5 Chestnut Terrace, Singapore 679123 (SG). LEE, Teck Chee; 84 Corporation Road #01-07, Singapore 649821 (SG). LIU, Xiaoting; Blk 656B #16-303 Jurong West Street 61, Singapore 642656 (SG). SOFIANOS, Efstratios; 9 Lauderdale House, 155 Gresham Road, Staines-upon-Thames Middlesex TW18 2AQ (GB).
- (74) **Agent:** POH, Chee Kian Daniel; Marks & Clerk Singapore LLP, Tanjong Pagar, P.O. Box 636, Singapore 910816 (SG).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) **Title:** A METHOD FOR SIGNAL PROCESSING OF VOICE OF A SPEAKER

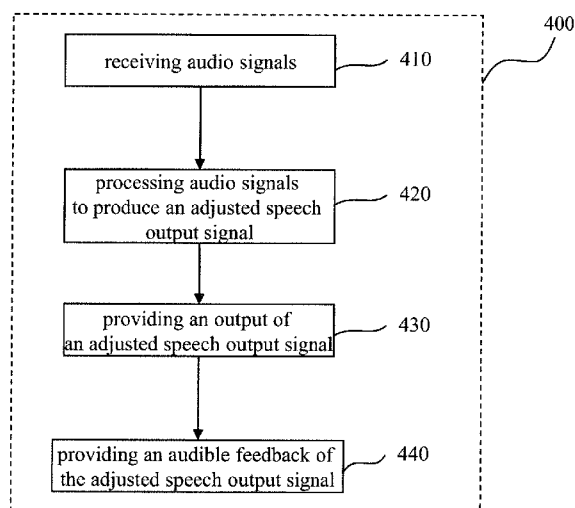


Fig. 4

(57) **Abstract:** A method for signal processing of voice of a speaker is provided. The voice of the speaker can correspond to audio signals including one or more voice parameters associable with emotion of the speaker. The method can include receiving the audio signals and processing the audio signals by adjusting one or more of the one or more voice parameters to produce an adjusted speech output signal. The method can further include providing an output of the adjusted speech output signal so that the adjusted speech output signal is audibly perceivable and providing an audible feedback of the adjusted speech output signal to the speaker. The speaker is capable of receiving the audible feedback of the adjusted speech output signal and further adjusting the one or more voice parameters in a manner so as to accentuate speaker emotion.



---

**Published:**

— *with international search report (Art. 21(3))*

## A METHOD FOR SIGNAL PROCESSING OF VOICE OF A SPEAKER

### Field Of Invention

The present disclosure generally relates to method for signal processing of voice of a speaker so that emotion of the speaker can be audibly perceived to be adjusted.

### Background

Conventional research relating to speech emotion (i.e., emotion(s) expressed in speech) is based on off-line processing. Additionally, speech features such as pitch contour and speech rate are typically used on classification.

Examples of the aforementioned research are:

[1] An acoustic study of emotions expressed in speech, S. Yildirim, M. Bulut, C. M. Lee et al., Proceedings of the International Conference on Spoken Language Processing (ICSLP '04), vol. 1, pp. 2193–2196, 2004; and

[2] Vocal expression of emotion. R. J. Davidson, K. R. Scherer, H. Goldsmith (Eds.). Handbook of the Affective Sciences (pp. 433–456). Scherer, K. R., Johnstone, T. & Klasmeyer, G. (2003). New York and Oxford: Oxford University Press

Therefore, it is notable that conventional research relating to speech emotion fails to provide or suggest a way for real-time processing (i.e., as opposed to off-line processing).

### Summary of the Invention

In accordance with an aspect of the disclosure, a method for signal processing of voice of a speaker is provided. The voice of the speaker can correspond to audio signals which include one or more voice parameters. The voice parameter(s) can be associated with emotion of the speaker.

The method can include receiving the audio signals, processing the audio signals, providing an output and providing an audible feedback.

The audio signals can be processed by adjusting one or more of the one or more voice parameters to produce an adjusted speech output signal.

In regard to providing an output, an output of the adjusted speech output signal can be provided so that the adjusted speech output signal is audibly perceivable.

In regard to providing an audible feedback, an audible feedback of the adjusted speech output signal can be provided to the speaker.

Speaker emotion can be audibly perceived to be adjusted, in accordance with the adjusted one or more voice parameters, when the adjusted speech output signal is output.

Moreover, the speaker is capable of receiving the audible feedback of the adjusted speech output signal and further adjusting the one or more voice parameters in a manner so as to accentuate speaker emotion.

#### **Brief Description of the Drawings**

Embodiments of the disclosure are described hereinafter with reference to the following drawings, in which:

Fig. 1 shows a system which can include an input module, a processing module and an output module, according to an embodiment of the disclosure;

Fig. 2 shows the system of Fig. 1 in further detail, according to an embodiment of the disclosure;

Fig. 3 shows exemplary signal waveforms associated with the system of Fig. 1, according to an embodiment of the disclosure; and

Fig. 4 shows a flow diagram illustrating a method for signal processing in association with the system of Fig. 1, according to an embodiment of the disclosure.

#### **Detailed Description**

Representative embodiments of the disclosure, for addressing one or more of the foregoing problems are described.

As mentioned earlier, conventional research relating to speech emotion (i.e., emotion(s) expressed in speech) is based on off-line processing.

The present disclosure contemplates that, in the context of speech emotion, real-time processing is avoided owing to the need to rely on predictive data (i.e., "future" data). Moreover, the present disclosure contemplates that there is a possibility that speech utterance may be rendered unnatural sounding during real-time processing (e.g. if speech rate is to be changed).

However, the present disclosure contemplates that real-time processing would be useful for real-time use type situation(s).

An example of a real-time use type situation is where a speaker desires a listener/listener(s) to perceive/discern, via the emotions expressed in the speaker's speech, the speaker's intended emotion (also, otherwise, referable to as "true" emotion) while the speaker is speaking. For example, in a classroom situation where a teacher is tired and is attempting to maintain order in the classroom, it is appreciable that the tired teacher's speech may not be capable of conveying the intended verbal firmness (i.e., intended emotion or "true" emotion). Therefore, it may be useful for the teacher that students in the classroom are capable of perceiving the intended verbal firmness from the teacher's speech despite the teacher's tiredness.

Another example of a useful real-time situation is where a speaker does not desire the listener/listener(s) to be able to perceive/discern, via the emotions expressed in the speaker's speech, the speaker's intended/"true" emotion while the speaker is speaking. For example, during a meeting session or during a teleconference session, it is generally a good idea to maintain a calm/neutral tone while having a discussion with a counterpart/counterpart(s). However, while having a sensitive/stressful discussion, a speaker's speech may inadvertently express negative emotions such as anger or frustration (i.e., "true" emotion). Therefore, it may be useful for the speaker that the speaker's true emotion/emotion(s) is/are not conveyed to his/her counterpart(s) during the teleconference session. Appreciably, a calm/neutral tone masks the speaker's true emotion (e.g., anger/frustration) and may help facilitate a favorable (or, at least, a more fruitful) outcome for the teleconference session.

Other examples are also useful. For example, during a karaoke session, it may be desired for a singer to sing in a manner matching the intended emotion of the song track being sung.

In this regard, there is a need to overcome the earlier contemplated problems associated with real-time processing so that application in real-time (i.e., real-time use type situation(s)) can be made possible. This will be discussed in further detail with reference to Fig. 1 to Fig. 4 hereinafter.

Referring to Fig. 1, a system 100 is shown in accordance with an embodiment of the disclosure. The system 100 can include an input module 102, a processing module 104 and an output module 106. The input module 102 can be coupled to the processing module 104. The processing module 104 can be coupled to the output module 106. As an option, the output module 106 can be coupled to the input module 102.

The input module 102 can be configured to generate input signals and communicate the input signals to the processing module 104 for processing.

The processing module 104 can be configured to receive and process the input signals to produce processed input signals. The processing module 104 can be further configured to communicate the processed input signals to the output module 106.

The output module 106 can be configured to receive and process the processed input signals in a manner so as to produce output signals.

In one embodiment, the system 100 can correspond to an ecosystem in which the input module 102, the processing module 104 and the output module 106 are three individual/independent devices which can be coupled as described earlier.

In another embodiment, the system 100 can correspond to a printed circuit board (PCB) on which the input module 102, the processing module 104 and the output module 106 are placed/positioned and coupled as described earlier. Specifically, the input module 102, the processing module 104 and the output module 106 can be three individual/independent elements carried by the PCB.

In yet another embodiment, the system 100 can correspond to a device such as an integrated circuit (IC) chip which includes the input module 102, the processing module 104 and the output module 106. Specifically, the input module 102, the processing module 104 and the output module 106 can form parts, which are coupled as described earlier, of the device (e.g., IC chip).

Other examples are also useful. In one example, in the system 100, the input module 102 and the processing module 104 can be parts of a device (i.e., a first device), and the output module 106 can be another device (i.e., a second device) coupled thereto. In another example, in the system 100, the processing module 104 and the output module 106 can be parts of a device (i.e., a first device), and the input module 102 can be another device (a second device) coupled thereto. In yet a further example, the input module 102 and the output module 106 can be parts of a device (i.e., a first device), and the processing module 104 can be another device (i.e., a second device) coupled thereto. Specifically, the first device can be coupled to the second device.

Additionally the aforementioned coupling can be based on one or both of wired coupling and wireless coupling.

Fig. 2 shows the system 100 of Fig. 1 in further detail, in accordance with an embodiment of the disclosure. As shown, the input module 102 can include an input portion 202 and a control portion 204. The input module 102 can, as an option, further include an analyzer portion 206.

Each of the input portion 202 and the control portion 204 can be coupled to the processing module 104. Moreover, the input portion 202 can, optionally, be coupled to one or both of the control portion 204 and the analyzer portion 206. Additionally, the analyzer portion 206 can, optionally, be coupled to the processing module 104.

The input portion 202 can be configured to receive audio signals. The audio signals can, for example, correspond to the voice of a speaker 208. Specifically, the audio signals can, for example, correspond to speaker voice (i.e., of the speaker 208). In this regard, the input portion 202 can, for example, be a microphone which can be configured to capture and receive the voice of a speaker 208 who may, for example, be speaking into the microphone. The input portion 202 can be further configured to process the received audio signals to produce source signals which can be communicated to one or both of the processing module 104 and the analyzer portion 206 for further processing.

The control portion 204 can be configured to generate control signals and communicate the control signals to one or both of the processing module 104 and the input portion 202. In general, the control signals can be communicated from the control portion 204 to the processing module 104 to control the manner in which the processing module 104 processes the source signals communicated

from the input portion 202 as will be discussed later in further detail in the context of an exemplary scenario. Additionally, the control portion 204 can be configured to communicate control signals to the input portion 202 to control the manner in which the input portion 202 receives and/or processes the audio signals. For example, control signals can be communicated from the control portion 204 to the input portion 202 to control sensitivity of the input portion 202 (e.g., microphone input sensitivity). In this regard, the control portion 202 can, for example, be a controller device having one or more controls. Examples of a control can include button(s), knob(s), slider(s), touch panel(s) and/or sensor(s). Moreover, examples of a sensor can include an accelerometer, a gyroscope, a magnetometer and a camera based tracker. The control(s) can, for example, be operated by the speaker 208 to generate control signals. For example, the speaker 208 can operate the slider (e.g., by adjusting the slider) to generate control signals. Therefore, the control portion 204 can, for example, be in the form of an “N” Dimensional (ND) controller device where “N” represents an integer and the integer represented by “N” can, for example, correspond to the number of controls available on the control portion 202 or the number of types of controls available on the control portion 202. In one example, where the control portion 202 features controls such as a knob and two sliders, the control portion 202 can be considered to be a 3 Dimensional controller device. In another example, where the control portion 202 features only one type control (e.g., two sliders), the control portion 202 can be considered to be a 1 Dimensional controller device.

The analyzer portion 206 can be configured to receive the source signals from the input portion 202. The analyzer portion 206 can be further configured to analyze the source signals to generate analysis signals. The analyzer portion 206 can yet be further configured to communicate the analysis signals to the processing module 104. In this regard, the analyzer portion 206 can, for example, correspond to a processor device configured to compute the source signals by way of analyzing to produce analysis signals. For example, the analyzer portion 206 can correspond to a processor device configured to perform the task of data mining (e.g., by way of analyzing) on the received source signals and produce analysis signals based on mined data. As will be discussed later in further detail in the context of an exemplary scenario, the processing module 104 can be configured to process one or both of the received control signals and the received source signals based on the analysis signals.

Appreciably, the input signals communicated from the input module 102 can include any one of the source signals, the control signals and the analysis signals, or any combination thereof.

The processing module 104 can be configured to receive the source signals, the control signals and/or the analysis signals for further processing to produce processed input signals which can, in turn, be further communicated (i.e., from the processing module 104) to the output module 106 for further processing to produce output signals.

For the sake of clarity, the system 100 will be discussed in further detail based on an exemplary scenario hereinafter.

In one exemplary scenario, the input portion 202 can correspond to a microphone, the processing module 104 can correspond to a processor device (e.g., in the form of an integrated circuit chip) and the output module 106 can correspond to one or more speaker drivers capable of sound output. Moreover, the speaker 208 can, for example, be a natural person generating and providing audio signals to the system 100 by way of speaking into the microphone. In this regard, the voice of the speaker 208 can correspond to the audio signals received by the input portion 202.

The voice of the speaker 208 can include one or more voice parameters which can be indicative of the emotion (e.g., anger, joy, sadness, surprise and calmness/neutral) of the speaker 208. In this regard, the one or more voice parameters can be associated with emotion of the speaker 208.

Examples of the aforementioned voice parameter(s) can include pitch contour, speech rate, formant, pitch & variability and energy.

Appreciably, the input portion 202 can be configured to process the audio signals to produce source signals. In this regard, the source signals can, by the same token, include one or more voice parameters. For example, audio signals, which can be analog based (i.e., voice of the speaker), can be received by the microphone (i.e., the input portion 202) and processed by the microphone to produce signals in a form (i.e., source signals) suitable for further processing by the processing module 104 and/or the analyzer portion 206. More specifically, the microphone can process the received audio signals by way of conversion to electric signals. Therefore, the source signals can, for example, correspond to electric signals which are based on conversion from the received audio signals which may be analog type signals.

As mentioned earlier, the source signals can be further communicated to the processing module 104 for further processing.

Preferably, the processing module 104 is configured to process the source signals by adjusting one or more of the voice parameters. Adjustment of the voice parameter(s) can, for example, be based on either the control signals or a combination of both of the control signals and the analysis signals. In this regard, the processing module 104 can be configured to process the source signals by adjusting one or more of the voice parameters based on, for example, either the control signals or a combination of the control signals and the analysis signals so as to produce processed source signals. The processed source signals can correspond to the aforementioned processed input signals in Fig. 1.

The processed source signals can be communicated from the processing module 104 to the output module 106 which can process the processed source signals in a manner so as to produce output signals. Additionally, as mentioned earlier, the output module 106 can correspond to one or more speaker drivers capable of sound output. In this regard, the output module 106 can be configured to receive and process the processed source signals in a manner so as to produce output signals which can be audibly perceived.

Since the source signals are based on the voice of the speaker 208 and can include one or more voice parameters, it is appreciable that the source signals can correspond to a speech input signal which can include one or more voice parameters. The speech input signal can be associated with one or more speaker emotions (i.e., of the speaker 208) based on the one or more voice parameters. By adjusting one or more voice parameters (i.e., associable with speaker emotion) at the processing module 104, the speech input signal can be considered to be adjusted in the context of speaker emotion. Therefore, output signals, which can be audibly perceived, communicated from the output module 106 can correspond to an adjusted speech output signal.

Therefore, the adjusted speech output signal, when output by the output module 106, can correspond to an audibly perceivable output signal where speaker emotion can be audibly perceived to be adjusted in accordance with the adjusted one or more voice parameters. Specifically, audibly perceivable speaker emotion associated with the adjusted speech output signal can be considered to be adjusted relative to speaker emotion associated with the input speech signal. The input speech signal is based on voice of the speaker 208.

Preferably, as signified by arrow 210 in Fig. 2, an audible feedback of the adjusted speech output signal can be provided to the speaker 208. In this regard, the speaker 208 can be capable of

receiving an audible feedback of the adjusted speech output signal. The audible feedback can, for example, be based on the speaker 208 hearing the adjusted speech output signal as being output from the output module 106. Appreciably, providing an audible feedback of the adjusted speech output signal to the speaker 208 can be useful in facilitating further adjustment of the one or more voice parameters. The one or more voice parameters can be further adjusted in a manner so as to accentuate, for example, an intended or a desired emotion (i.e., which can be audibly perceived) of the speaker 208. More specifically, the speaker 208 can be capable of receiving an audible feedback of the adjusted speech output signal and further adjusting the one or more voice parameters in a manner so as to accentuate speaker emotion. Yet more specifically, the speaker 208 can be capable of hearing the adjusted speech output signal as being output by the output module 106 and further adjusting, as desired by the speaker 208, the one or more voice parameters in a manner so as to accentuate speaker emotion.

For the sake of further clarity, the above discussed exemplary scenario will be put in context based on an exemplary real-time use type situation hereinafter.

Taking, as an example, a real-time use type situation where a teacher (i.e., the speaker 208) wishes to instill discipline during a lecture session in a lecture theatre/hall, it may be useful for the teacher (who may be tired after a long day) that students in the lecture theatre are able to perceive an emotion (i.e., from the teacher) associated with anger.

The teacher using/operating the input module 102 may speak into the microphone (i.e., the input portion 202) and generate control signals using the control portion 204 (e.g., by sliding the slider). The teacher's voice may be tired sounding. The tired sounding voice (audio signals which can be received by the microphone) may be associated with the emotion of calmness. Hence, the tired sounding voice of the teacher can be processed by the microphone to produce the speech input signal (i.e., source signals). In this regard, the speech input signal can be associated with the emotion of calmness which may not be the intended emotion of the teacher and, further, not ideal for the purpose of instilling discipline.

Appreciably, there may be a need for the teacher to make an adjustment.

Preferably, adjustment can be made by way of processing speech input signals based on control signals. The control signals can be generated based on the intent of the teacher to make an

adjustment to the speech input signal which can be associated with the emotion of calmness so as to produce an adjusted speech output signal which can be audibly perceived to be associated with the emotion of anger. To aid the teacher in using the control portion 204 to generate control signals as appropriate, labels can be provided. For example, a label can be provided at one end of the slider indicating "calm voice" and another label can be provided at another end of the slider indicating "extremely angry voice". In this manner, the teacher can slide the slider accordingly to vary the degree/strength of anger (i.e., more towards "calm voice" or more towards "extremely angry voice") which can be audibly perceived via the adjusted speech output signal.

Alternatively, adjustment can be made by way of processing speech input signals based on a combination of the control signals and the analysis signals. As mentioned earlier, the analyzer portion 206 can be configured to analyze the source signals (i.e., speech input signals) to generate analysis signals. Specifically, the analyzer portion 206 can be configured to analyze the teacher's voice and extract one or more voice characteristics such as pitch and formant. Based on the voice characteristic(s), it is possible for the analyzer portion 206 to determine useful information associated with the speaker 208 (i.e., the teacher) and generate analysis signals accordingly. An example of useful information associated with the speaker 208 is the gender of the speaker 208. For example, based on the voice characteristic(s) such as the pitch, it can be determined that the teacher is a female. When it is determined that the teacher is a female, the analyzer portion 206 can be configured to communicate analysis signals indicative of the teacher's gender being a female. The analysis signals can be used for adjusting the control signals as will be discussed later in further detail with reference to Fig. 3. Therefore, the speech input signals can be considered to be processed based on a combination of the control signals and the analysis signals. Additionally, the control signals can be adjusted by the analysis signals at either the control portion 204 or the processing module 104. Moreover, it is also possible for the control signals to be adjusted by the analysis signals at both the control portion 204 and the processing module 104.

The speech input signals can be communicated to the processing module 104 for further processing based on either the control signals or a combination of the control signals and the analysis signals. Specifically, one or more voice parameters of the speech input signal which can be associated with the emotion of anger (i.e., intended emotion of the speaker 208) can be adjusted based on the control signals or a combination of the control signals and the analysis signals.

Therefore, adjusting the voice parameter(s) based on the control signals can be the basis for a form of real-time processing to produce an adjusted speech output signal in which adjustment has been made in the context of the speaker emotion (i.e., associated with anger in this exemplary real-time use type situation). Moreover, since it is also possible for the processing module 104 to process the speech input signals based on a combination of the control signals and analysis signals (communicated from the analyzer portion 206) to produce the adjusted speech output signal, the analysis signals can be considered to be another basis for a form of real-time processing to produce an adjusted speech output signal in which adjustment has been made in the context of the speaker motion (i.e., associated with anger in this exemplary real-time use type situation). Processing of the speech input signal based on the control signals or a combination of the control signals and the analysis signals will be discussed later in further detail with reference to Fig. 3.

In this regard, it is appreciable that the speech input signals which can be associated with the emotion of calmness (e.g., owing to the tiredness of the teacher) can be adjusted by the teacher so that the adjusted speech output signals can be audibly perceived to be associated with the emotion of anger.

Therefore in this exemplary real-time use type situation, the teacher can be speaking into the microphone (i.e., the input portion 202) with a tired sounding voice (which can be associated with the emotion of calmness). However, instead of audibly perceiving the emotion of calmness, the students hearing the adjusted speech output signal from the one or more speaker drivers may audibly perceive the adjusted speech output signal to be associated with the emotion of anger.

Appreciably processing of the speech input signal to produce the adjusted speech output signal is based on real-time based processing.

Specifically, as discussed thus far, real-time processing of the received audio signals (i.e., voice of the teacher in accordance with the above exemplary real-time use type situation) to produce adjusted speech output signals can be made possible using the system 100 by virtue at least one of the following approaches:

- 1) Processing the received audio signals based on the control signals.
- 2) Processing the received audio signals based on the analysis signals in combination with the control signals.

Appreciably, apart from the two approaches mentioned above, real-time processing is also possible by providing an audible feedback (i.e., as signified by arrow 210) of the adjusted output signal to the speaker 208 (i.e., the teacher). Audible feedback of the adjusted output signal can be provided to the teacher via the one or more speaker drivers (i.e., the output module 106). Preferably, audible feedback corresponds to the teacher hearing the adjusted speech output signal as being output from the one or more speaker drivers (i.e., the output module 106). Thus, based on the adjusted speech output signal as being output by the one or more speaker drivers, the teacher is essentially hearing her (assuming the gender of the teacher is a female as discussed earlier) own voice which has been adjusted to be associated with the emotion of anger from of her original tired sounding voice (associated with the emotion of calmness) as received by the microphone (i.e. the input portion 202). In this regard, the speech input signal can be associated with an "original/input voice" (e.g., tired sounding voice as received by the microphone) and the adjusted speech output signal can correspond to a "modified voice" (e.g., angry sounding voice as being output by the one or more speaker drivers) relative to the "original/input voice".

The teacher, upon, hearing her own voice as output via the one or more speaker drivers (i.e., receiving audible feedback), may decide that the emotion of anger may not be clearly audibly perceived enough. The teacher may then decide to make further adjustment(s) to the one or more voice parameters so as to accentuate the emotion of anger.

The manner of further adjustment(s) can, for example, be by way of the teacher adjusting her own voice (i.e., speaker voice) when speaking into the microphone to be more angry sounding so as to accentuate the intended or the desired emotion of, for example, anger. Thus the teacher may realize that she has to make an effort to be less tired sounding when speaking into the microphone.

Other ways of further adjustment are also possible.

In one example, the teacher can make further adjustments via the control portion 204 by sliding the slider even more towards the "extremely angry voice" direction. This can be akin to dynamically adjusting the strength/impact of the effect of, for example, anger by making adjustment(s) (e.g., sliding the slider) using the control portion 204. Hence, effectively, further adjustment can be by manner of adjusting control signals generated and communicated to the processing module 104.

In another example, further adjustments can be by manner of both adjusting speaker voice communicated to, and received by, the microphone and adjusting control signals generated, and communicated to, the processing module 104.

Therefore, apart from the two earlier approaches mentioned above, it is also possible to aid real-time processing by providing audible feedback of the adjusted speech output signal to the speaker 208 so that the speaker 208 can make further adjustment(s), as desired or appropriate, to one or more voice parameters in a manner so as to accentuate speaker emotion as audibly perceived via the output module 106.

Hence it is appreciable that the system 100 effectively enables a user (e.g., the speaker 208) to change/adjust perceived (i.e., as perceived by one or more listeners such as the aforementioned students) emotion associated with the user's speech. As mentioned earlier as an example, the original tired sounding voice of the speaker 208 (e.g., a teacher) which can be associated with an emotion of calmness can be adjusted/changed so that listeners (e.g., the students) can audibly perceive an emotion of anger instead.

Furthermore, since the system 100 can be considered to be based on real-time processing by virtue of at least one of the following:

- 1) Processing the received audio signals based on the control signals; and
- 2) Processing the received audio signals based on the analysis signals in combination with the control signals,

there is no need to rely on predictive data (i.e., "future" data) as adjustment is made possible in real-time. Moreover, by providing audible feedback of the adjusted speech output signal to the speaker 208 so that the speaker 208 can make further adjustment(s), as desired or appropriate, to one or more voice parameters in a manner so as to accentuate speaker emotion as audibly perceived via the output module 106, it is also appreciable that there is no need to rely on predictive data (i.e., "future" data) as real-time compensation by the speaker 208 is made possible. It is also appreciable that based on the discussed approach(es), the possibility of speech utterance being rendered unnatural sounding during real-time processing (e.g. if speech rate is to be changed) can be significantly reduced since real-time adjustment and/or real-time compensation can be possible.

Earlier mentioned, the speech input signal (i.e., source signals based on the audio signals) can be processed based on the control signals or a combination of the control signals and the analysis signals. This will be discussed in further detail with reference to Fig. 3 hereinafter.

Fig. 3 shows exemplary signal waveforms associated with the system 100, in accordance with an embodiment of the disclosure.

Specifically, Fig. 3 shows an exemplary speech input signal waveform 310 and a corresponding data waveform 320 of a voice parameter. Fig. 3 also shows an adjusted data waveform 330 where the voice parameter has been adjusted and an exemplary speech output signal waveform 340.

The voice parameter can, for example, relate to pitch. Therefore, the data waveform 320 can show the pitch data associated with the speech input signal (i.e., as represented by the exemplary speech input signal waveform 310). The pitch data associated with the speech input signal can be indicative of audibly perceivable pitch (e.g., the degree of height or depth of voice tone) of the speaker's 208 voice. The data waveform 320 can be derived from the speech input signal waveform 310 by manner of, for example, a sampling process or an extraction process which can be performed by the processing module 104.

As shown, the data waveform 320 can include a plurality of data portions (e.g., labels 320a, 320b and 320c). Each of the data portions can include one or more data points (e.g., labels 320d and 320e within data portion 320c)

In operation, in accordance with an embodiment of the disclosure, the speaker 208 can provide audio signals to the system 100 by, for example, speaking into the input portion 202. The input portion 202 processes the audio signals and produces the speech input signal (i.e., source signals) as represented by the exemplary speech input signal waveform 310. The speech input signal can be processed by the processing module 104 in a manner so that one or more data points (e.g., labels 320f, 320g and 320h) within one or more data portions (e.g., labels 320a, 320b and 320c) is/are adjusted. Adjustment of the data point(s) can be based on instructions communicated to the processing module 104 in the form of the aforementioned control signals. In this manner, the speech input signal can be processed based on the control signals.

As mentioned earlier, the speech input signal can, as an option, be processed based a combination of the control signals and the analysis signals. This will be discussed in further detail using the earlier discussed example of the analysis signals being indicative of the gender of the speaker 208.

The analysis signals can be used to adjust the control signals (at the control portion 204 and/or the processing module 104), so that one or more of the data points can be adjusted, for example, to either a higher degree in or a lesser degree in comparison to the data point(s) being adjusted based on the control signals alone.

For example, where the speaker 208 is a female teacher (e.g., in the context of the earlier discussed exemplary real-time use type situation), it may be desired for data point 320f to be adjusted by a smaller degree as indicated by arrow 350. Conversely, where the speaker 208 is, for example, a male teacher, it may be desired for data point 320f to be adjusted by a larger degree as indicated by arrow 360.

In this above manner, the speech input signal can be processed based a combination of the control signals and the analysis signals. Specifically, the analysis signals can be used to adjust the control signals and the adjusted control signals can, in turn, be used to adjust the speech input signal to produce the adjusted speech output signal.

Referring to Fig. 4, a flow diagram 400 illustrating a method for signal processing in association with the system 100 is shown in accordance with an embodiment of the disclosure. The method for signal processing can be directed at the signal processing of voice of a speaker (e.g., earlier discussed speaker 208). The voice of the speaker can correspond to audio signals which include one or more voice parameters. The one or more voice parameters can be associated with emotion of the speaker.

The method can include receiving audio signals 410, processing audio signals to produce an adjusted speech output signal 420, providing an output of an adjusted speech output signal 430 and providing an audible feedback of the adjusted speech output signal 440.

In regard to the step of receiving audio signals 410, the audio signals can be received by the input portion 202 as discussed earlier with reference to the system 100.

In regard to the step of processing audio signals to produce an adjusted speech output signal 420, one or more voice parameters associated with the audio signals can be adjusted to produce an adjusted speech output signal. As mentioned earlier with reference to the system 100, the received audio signals can be processed by the input portion 202 to produce an input speech signal which can include one or more voice parameters (i.e., corresponding to the one or more voice parameters associated with the audio signals as received by the input portion 202). The input speech signal can be communicated to the processing module 104 for further processing. Specifically, the processing module 104 can be configured to process the speech input signal (i.e., source signals based on the audio signals) based on the control signals or a combination of the control signals and the analysis signals to produce the adjusted speech output signal.

In regard to the step of providing an output of an adjusted speech output signal 430, the adjusted speech output signal can, as discussed earlier with reference to the system 100, be communicated to the output module 106 which can be configured to process the speech output signal in a manner so as to produce corresponding output signals which can be audibly perceived. In this regard, the adjusted speech output signal can, by virtue of the output signals from the output module 106, be considered to be capable of being audibly perceived. More specifically, the adjusted speech output signal can be considered to be audibly perceivable by virtue of the output signals from the output module 106.

In regard to the step of providing an audible feedback of the adjusted speech output signal 440, it is preferable, in accordance with an embodiment of the disclosure, the audible feedback is provided to the speaker 208 and that the audible feedback corresponds to the speaker 208 hearing the adjusted speech output signal as being output from the output module 106 (e.g., from the one or more speaker drivers). Thus, for example, based on the adjusted speech output signal as being output by the one or more speaker drivers, the speaker 208 is essentially hearing his/her own voice which has been adjusted to be associated with the emotion of, for example, anger from his/her, for example, original tired sounding voice which can be associated with the emotion of calmness as received by the input portion 202 (e.g., microphone). In this regard, an audible feedback of the adjusted speech output signal can be considered to be provided to the speaker 208.

Preferably, speaker emotion is audibly perceivable to be adjusted, in accordance with the adjusted one or more voice parameters, when the adjusted speech output signal is output.

Also preferably, the speaker 208 is capable of receiving the audible feedback of the adjusted speech output signal and further adjusting the one or more voice parameters in a manner so as to accentuate speaker emotion.

In the foregoing manner, various embodiments of the disclosure are described for addressing at least one of the foregoing disadvantages. Such embodiments are intended to be encompassed by the following claims, and are not to be limited to specific forms or arrangements of parts so described and it will be apparent to one skilled in the art in view of this disclosure that numerous changes and/or modification can be made, which are also intended to be encompassed by the following claims.

For example, although it is earlier mentioned with reference to Fig. 2 that in accordance with an embodiment of the disclosure, the input module 102 can include the analyzer portion 206, it is appreciable that it is also possible that instead of the input module 102, the analyzer portion 206 can be included in the processing module 104. Specifically, in accordance with an embodiment of the disclosure, the processing module 104 can include the analyzer portion 206 (i.e., the analyzer portion 206 is absent from the input module 102). In such case, the analyzer portion 206 can correspond to a sub-processor within the processing module 104 and can operate in analogous manner per earlier discussion with reference to Fig. 2.

Of course, it is also possible that, in accordance with an embodiment of the disclosure, both the input module 102 and the processing module 104 can include an analyzer portion which can operate in like manner to the analyzer portion 206 per earlier discussion with reference to Fig. 2.

Additionally, the system 100 may also rely on facial visual (e.g., as captured in real-time by the earlier mentioned camera based tracker) of the speaker 208 to aid in real-time processing of the speech input signal so as to produce the adjusted speech output signal. This can be accomplished by detecting facial features and morphing the detected facial features based in data which correlates facial features with emotions (see, for example, "Real-time classification of evoked emotions using facial feature tracking and physiological responses. Bailenson J. N, Pontikakis E. D., Maus I. B. et al, Int. J. Human-Computer Studies 66 (2008) 303-317.").

**Claims**

1. A method for signal processing of voice of a speaker, the voice of the speaker corresponding to audio signals comprising one or more voice parameters associable with emotion of the speaker, the method comprising:
  - receiving the audio signals;
  - processing the audio signals by adjusting one or more of the one or more voice parameters to produce an adjusted speech output signal;
  - providing an output of the adjusted speech output signal so that the adjusted speech output signal is audibly perceivable; and
  - providing an audible feedback of the adjusted speech output signal to the speaker, wherein speaker emotion is audibly perceivable to be adjusted, in accordance with the adjusted one or more voice parameters, when the adjusted speech output signal is output, and wherein the speaker is capable of receiving the audible feedback of the adjusted speech output signal and further adjusting the one or more voice parameters in a manner so as to accentuate speaker emotion.
2. The method as in claim 1, wherein further adjustment of the one or more voice parameters to accentuate speaker emotion is based on the speaker adjusting speaker voice.
3. A device configured to receive and process voice of a speaker, the voice of the speaker corresponding to audio signals comprising one or more voice parameters associable with emotion of the speaker, the device comprising:
  - an input portion configured to receive the audio signals;
  - a controller operable to produce control signals;
  - a processor configured to receive process the audio signals by adjusting, based on the control signals, one or more of the one or more voice parameters to produce an adjusted speech output signal; and
  - an output portion configured to provide an output of the adjusted speech output signal, wherein the one or more voice parameters are adjusted in a manner so that speaker emotion is audibly perceived to be adjusted when the adjusted speech output signal is output, and

wherein the speaker is capable of receiving the audible feedback of the adjusted speech output signal and further adjusting the one or more voice parameters in a manner so as to accentuate speaker emotion.

1/3

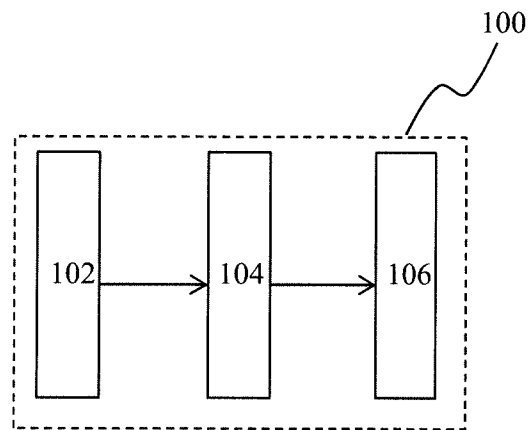


Fig. 1

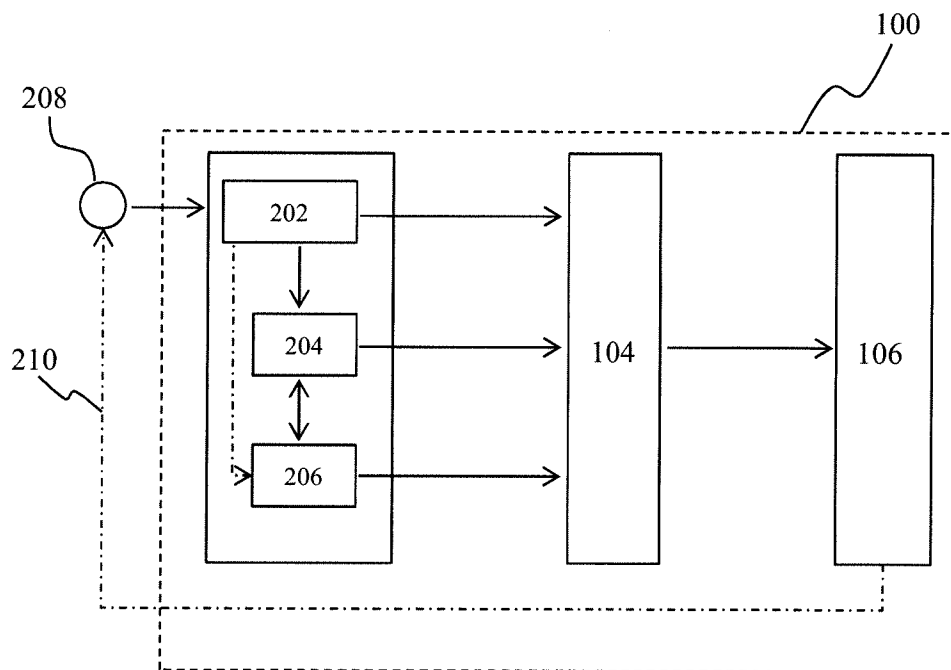


Fig. 2

2/3

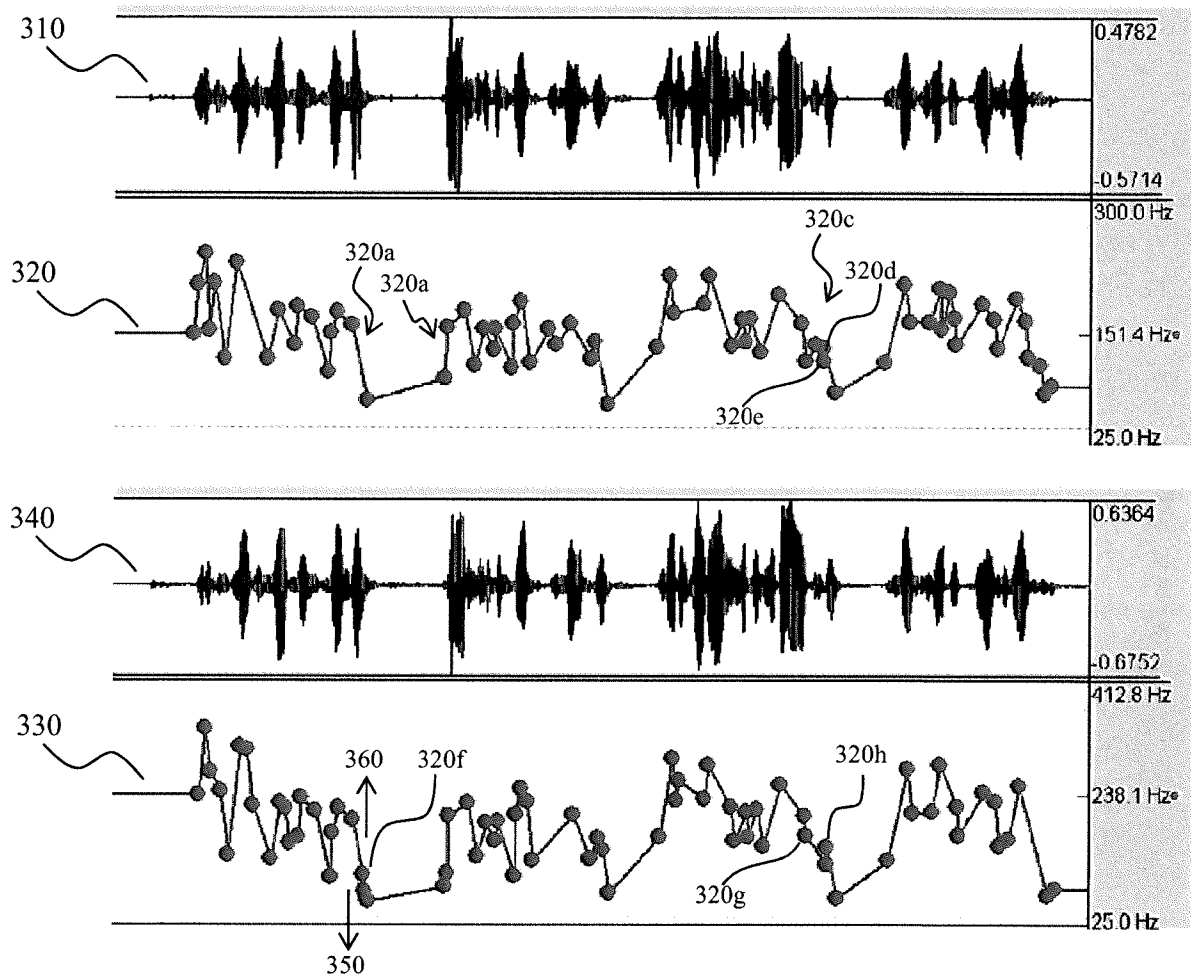


Fig. 3

3/3

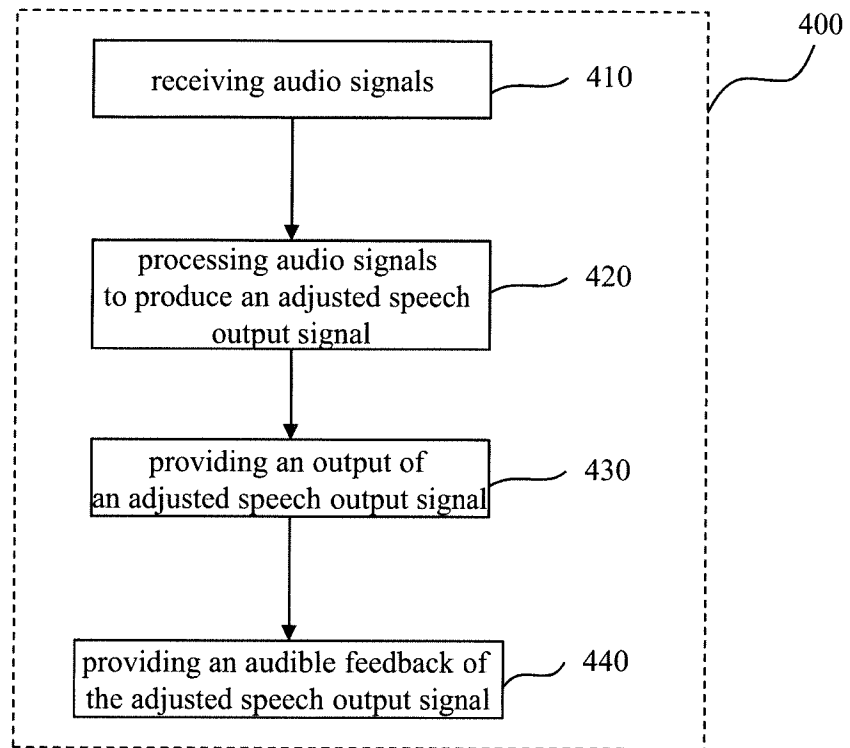


Fig. 4

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/SG2015/050519

## A. CLASSIFICATION OF SUBJECT MATTER

INV. G10L21/003  
ADD. G10L25/63      G10L25/15      G10L25/90

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, INSPEC, WPI Data

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                            | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | US 2008/147413 A1 (SOBOL-SHIKLER TAL [IL])<br>19 June 2008 (2008-06-19)<br>paragraphs [0004], [0005], [0006],<br>[0008], [0009]<br>paragraph [0012]<br>paragraph [0015]<br>paragraph [0023]<br>paragraphs [0040], [0041], [0050],<br>[0055] - [0057]<br>----- | 1-3                   |
| A         | US 2013/238337 A1 (KAMAI TAKAHIRO [JP] ET<br>AL) 12 September 2013 (2013-09-12)<br>paragraphs [0055], [0058], [0059];<br>figure 24<br>paragraph [0063]; figure 25<br>paragraphs [0141], [0164] - [0167],<br>[0174], [0185]; figure 8<br>-----                 | 1-3                   |



Further documents are listed in the continuation of Box C.



See patent family annex.

### \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

29 February 2016

Date of mailing of the international search report

09/03/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040,  
Fax: (+31-70) 340-3016

Authorized officer

Krembel, Luc

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/SG2015/050519

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date           |
|-------------------------------------------|---------------------|----------------------------|-------------------------------|
| US 2008147413                             | A1                  | 19-06-2008                 | NONE                          |
| -----                                     |                     |                            |                               |
| US 2013238337                             | A1                  | 12-09-2013                 | CN 103370743 A 23-10-2013     |
|                                           |                     |                            | JP 5194197 B2 08-05-2013      |
|                                           |                     |                            | JP WO2013008471 A1 23-02-2015 |
|                                           |                     |                            | US 2013238337 A1 12-09-2013   |
|                                           |                     |                            | WO 2013008471 A1 17-01-2013   |
| -----                                     |                     |                            |                               |