

에이전트의 표정과 목소리 정서의 교차양상지각*

Crossmodal Perception of Mismatched Emotional Expressions by Embodied Agents

조유숙** · 석지혜** · 한광희**†

Yu Suk Cho** · Ji He Suk** · Kwang Hee Han**†

연세대학교 심리학과**

Department of Psychology, Yonsei University**

Abstract

Today an embodied agent generates a large amount of interest because of its vital role for human-human interactions and human-computer interactions in virtual world. A number of researchers have found that we can recognize and distinguish between various emotions expressed by an embodied agent. In addition many studies found that we respond to simulated emotions in a similar way to human emotion. This study investigates interpretation of mismatched emotions expressed by an embodied agent (e.g. a happy face with a sad voice); whether audio-visual channel integration occurs or one channel dominates when participants judge the emotion. The study employed a 4 (visual: happy, sad, warm, cold) X 4 (audio: happy, sad, warm, cold) within-subjects repeated measure design. The results suggest that people perceive emotions not depending on just one channel but depending on both channels. Additionally facial expression (happy face vs. sad face) makes a difference in influence of two channels; Audio channel has more influence in interpretation of emotions when facial expression is happy. People were able to feel other emotion which was not expressed by face or voice from mismatched emotional expressions, so there is a possibility that we may express various and delicate emotions with embodied agent by using only several kinds of emotions.

Keywords : embodied agents, mismatched emotion, emotional information, crossmodal perception

요약

오늘날 가상 세계에서 인간과 인간, 컴퓨터와 인간의 상호작용을 좀 더 효율적으로 도와줄 수 있는 매개체로서 에이전트에 대한 관심이 증대되고 있다. 특히 에이전트의 정서표현도 인간의 정서 표현처럼 해석할 수 있고 그에 대해 인간의 정서표현에 반응하는 것과 마찬가지로의 방식으로 반응한다는 연구들이 있어왔다. 그렇다면 에이전트의 정서표현이 표정과 목소리에서 서로 불일치한다면 어떤 쪽에 더 치중해서 에이전트의 정서를 해석할지 또는 어느 한 쪽에 치중하는 것이 아니라 두 정서 모두 조합해서 받아들일지 알아보고자 했다. 또한 만약 표정과 목소리의 두 가지 정서를 조합해서 인지한다면 표정이나 목소리에서는 표현되지 않은 제 3의 정서를 느끼지는 않는 것인지 알아보았다. 실험은 표정(4) X 목소리(4)로 이루어졌다. 실험 결과 사람들은 에이전트의 한 가지 채널에 지배받아서 정서를 인지하는 것이 아니라 표정(시각)과 목소리(청각) 두 채널 모두를 조합해서 정서를 인지하는 것으로 나타났다. 또한 두 채널의 정서가 불일치할 때에는 혼란에 빠져 무슨

* 이 논문은 BK21사업(인지 및 심리과학 전문인력 양성 및 실용화 사업단)의 지원을 받아 수행되었음

† 교신저자 : 한광희 (연세대학교 심리학과 교수)

E-mail : khan@yonsei.ac.kr

TEL : 02-2123-2442

FAX : 02-2123-4723

정서인지 모르게 되는 것이 아니라 또 다른 실망, 지루함, 지친, 반가운 등의 제 3의 정서를 느끼는 것으로 나타났다.

주제어 : 에이전트, 교차양상지각, 정서통합, 정서인지, 불일치 정서

1. 서론

오늘날 컴퓨터뿐만 아니라 휴대용 기기들의 발달로 인터페이스를 통한 사용자 경험이 중요해졌다. 따라서 인터페이스를 통해 정보를 전달할 때 단순히 텍스트로만 제시하기 보다는 시각적 도구(아이콘, 그림), 청각적 도구(벨소리), 촉각적 도구(진동)등을 이용하여 기기와의 상호작용 효과를 높인다. 이런 도구 중에 하나가 에이전트(embodied agent)인데, 여기서 에이전트(embodied agent)라 함은 컴퓨터와 인간의 상호작용을 돕기 위해서 만들어진 그래픽적으로 형태를 가진 가상의 캐릭터로 정의한다. 에이전트는 시각적 도구(표정)와 청각적 도구(목소리)를 이용하여 사용자와 상호작용하기 때문에 특히, 교육¹⁾이나 운동²⁾, 영양보조³⁾로서의 역할을 하는 기기들에 등장하여 사용자에게 해야 할 일을 에이전트가 전달함으로써 상호작용의 실재감을 높였다.

인간과 비슷하게 표정과 목소리로 사용자와 의사소통한다는 자체가 상호작용의 실재감을 높이는 요소이지만 에이전트에 대한 관심이 점점 증가하면서 상호작용을 통해 사용자 경험을 증대시킬 수 있는 요소에 대한 연구들이 있는데,⁴⁾ 그중 하나가 정서이다.

Berry와 그의 동료들은 에이전트로 정보를 제시할 때 정보와 일치하는 정서로 표현해 주면 수행이 향상됨을 발견했다.⁶⁾ 또한 에이전트가 참가자의 수행에 맞게 공감의 정서를 표현하면 에이전트에 대한 선호도, 신뢰도 등이 증가하였다.⁷⁾

하지만, 지금까지 에이전트가 표현하는 정서는 주로 Ekman이 제시한 인간의 기본정서 여섯 가지(행복, 슬픔, 놀람, 분노, 공포, 역겨움)에 국한되어 있었다.⁸⁾ 인간이 인식할 수 있는 여러 가지 정서를 표현할 수 있다면 사용자는 기기와의 상호작용에서 실재감을 더 느낄 것이다. 그러나 실제로 기본 정서 이외의 정서를 에이전트가 표현하도록 설계하는 것은 정서의 복잡성을 고려하면 쉬운 일이 아니며, 에이전트가 기본 정서가 아닌 미묘한 정서를 표현하도록 설계되었다면 인간이 그대로 인지하는지도 고려해야 할 문제이다. 그러기 위해서는 단순히 사용자가 에이전트의 정서 표현을 구분할 수 있다는 사실⁹⁾¹⁰⁾에서 나아가 에이전트의 정서 표현을 인간이 어떤 방식으로 인지하는지 그 전체적인 양상에 대해 알아볼 필요가 있다.

인간은 실제 사람의 표정과 목소리를 동시에 지각하여 해석하게 되는데, 이때 교차 양상 지각(cross-modal perception)이 발생한다. 두 채널에서 각기 다른 두 정서가 입력되면 두 정서를 통합하여 판단하고, 하

- 1) Burleson, W. & Picard, R. (2004). Affective agents: sustaining motivation to learn through failure and a state of stuck. Social and Emotional intelligence in learning environments workshop, in conjunction with the 7th International Conference on Intelligent Tutoring Systems.
- 2) Bickmore, T. & Picard, R. (2005). Establishing and maintaining long-term human - computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI) 12 (2), 293 - 327.
- 3) Creed, C. (2006). Using computational agents to motivate diet change. In: Ijsselstein, W., De Kort, Y., Midden, C., & Van Den Hoven, E. (Eds.), First international conference on Persuasive Technology for human well-being. Springer-Verlag, Eindhoven University of Technology(the Netherlands), 100 - 103.
- 4) Elliot, C., & Brezezinski, J. (1998). Autonomous agents as synthetic characters. AI Magazine, 19(2), 13-30.
- 5) Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. Computers in Human Behavior,

24(4), 1494 - 1509.

- 6) Berry, D. C., Butler, L. T., & de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. International Journal of Human-Computer Studies, 63(3), 304-327.
- 7) Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies, 62(2), 161-178.
- 8) Ekman, P. (1992) An argument for basic emotions. Cognition and emotion, 6, 169-200.
- 9) Bartneck, C. (2001). Affective expressions of machines. Extended Abstracts of CHI'01 : Conference on Human Factors in Computing Systems, 189 - 190.
- 10) Massaro, D. W., Cohen, M. M., Beskow, J., & Cole, R. (2000). Developing and evaluating Conversational Agents. In: Chassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.), Embodied conversational agents, MIT Press(Cambridge, MA), 287 - 318.

나의 채널에서 모호한(중립적) 정서가 입력되면 나머지 채널의 정보에 치우쳐서 판단한다.¹¹⁾ 그렇다면 에이전트와 상호작용할 때도 사람들은 교차 양상 지각을 할 가능성이 있으므로 에이전트 정서의 교차 양상 지각에 관한 실험을 통해 정서 인지의 양상에 관해 살펴볼 수 있겠다.

따라서 본 연구는 에이전트 정서의 지각 양상을 살펴보기 위해 얼굴 표정과 목소리 두 채널을 모두 고려하여 에이전트가 표현하는 두 채널의 정서가 불일치하는 경우에 표정이나 목소리 중 어느 한쪽 채널이 우월하게 작용하는지 혹은 두 채널이 모두 영향을 미치는지 알아보고자 한다. 또한 두 채널의 정서가 불일치할 때 단순히 정서의 부조화를 느껴 혼란에 빠지는지, 혹은 제3의 정서를 느끼는지에 대해서도 살펴보고자 한다.

2. 이론적 배경

2.1. 교차 양상 지각(crossmodal perception)

두 가지 이상의 감각 채널에 정보가 동시에 입력되면 두 채널의 정보를 통합하는 교차양상 지각(cross-modal perception)이 발생한다. Saldaña와Rosenblum은 음악의 톤이 올라가는 것으로 들리는지 내려가는 것으로 들리는지에 시각적 정보(얼굴 표정)가 영향을 주는 것을 발견했다.¹²⁾ 뿐만 아니라 표정은 상대적인 음의 높이 지각에도 영향을 미치는 것으로 나타났다.¹³⁾ 언어지각에도 시각적 정보가 큰 영향을 미친다.¹⁴⁾ ‘ga’를 말하는 입 모양 시각자극에 ‘ba’의 청각자극을 덧붙이면 ‘da’로 인지하게 된다. 이것은 시각과 청각 자극이 통합되어 처리되는 것을 의미한다.

정서 역시 두 가지 이상의 감각 채널에 동시에 입력되면 교차 양상 지각이 발생한다. 얼굴 표정(시각)과 목소리(청각)의 정서가 불일치하는 경우 어느 한

채널이 우세하기 보다는 두 채널의 정서가 통합되어 처리된다.¹⁵⁾¹⁶⁾

de Gelder과 Vroomen의 연구를 보면 기쁜(happy) 목소리와 두려운(fear) 목소리의 정서를 목소리만 듣고 평가하게 했을 때는 100% 정서를 구분하였다. 그런데 이 목소리 자극에 해당 목소리 자극의 정서와 일치하지 않는 얼굴 표정자극을 더 해주면 결과가 달라진다(예, 기쁜 목소리와 두려운 표정 결합자극). de Gelder과 Vroomen의 연구에서는 기쁜 표정과 두려운 표정을 양 끝점으로 놓고 컴퓨터를 이용하여 그 두 표정을 10 단계로 연속적으로 변화시킨 표정자극을 만들어냈다. 그렇게 만들어진 자극들에 기쁜 목소리와 두려운 목소리를 결합하면 목소리만 제시했을 때와 감정 평가가 달라진다. 목소리만 제시했을 때는 기쁨이라고 평가했는데, 동일한 목소리에 두려운 표정 쪽의 표정들을 제시할수록 기쁨과 두려움 두 정서의 중간 정도로 평가된다. 이는 목소리의 정서뿐 아니라 표정에서 전달되는 정서가 전체적인 정서평가에 영향을 미친 것을 의미한다.¹⁷⁾

또한, 오페라 가수가 기쁜 노래를 부르는 영상(시각)과 슬픈 노래(청각)를, 슬픈 노래를 부르는 영상과 기쁜 노래를 매치시킨 것을 본 사람들은 연구자가 노래(청각)에만 집중해서 정서를 평가하게 했음에도 영상자극(시각)에 영향을 받아 정서를 평가하였다. 즉 슬픈 노래가 들림에도 오페라 가수가 기쁜 노래를 부르는 영상을 보여주면 슬픈 노래에 슬픈 노래를 부르는 영상을 보여줄 때보다 기쁜 정서로 평가된 것이다.¹⁸⁾ 이는 목소리(청각)와 표정(시각)간의 정서의 교차 양상 지각(crossmodal perception)을 암시한다.

그렇다면 인간은 에이전트에 대해서도 동일한 반응을 보일까. 사람들에게 그들의 식습관에 대해 조언을 하면서 정서를 표현하는 영양 전문가 에이전트에 대

11) Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review*, 3, 215- 221.
 12) Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception and Psychophysics*, 54, 406-416.
 13) Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, 18, 756-757.
 14) McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

15) Massaro, D. W., & Egan, P. B. (1996) Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, 3(2), 215-221.
 16) Hietanen, J. K. Leppänen, J. M., Illi, M. & Surakka, V. (2004). Evidence for the integration of audiovisual emotional information at the perceptual level of processing. *European Journal of Cognitive Psychology*, 16(6), 769 - 790.
 17) de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and emotion*, 14 (3), 289-311.
 18) Thompson, W. F., Russo, F. A. & Quinto L. (2008). Audio-visual integration of emotional cues in song. *Cognition and emotion*, 22(8), 1457-1470.

한 연구가 있다. 에이전트가 시각과 청각 채널에서 일치하지 않는 정서를 표현하면 사람들은 시각이나 청각 자극 한 가지에만 의존하여 해석하는 것이 아니라 시각과 청각 두 채널에서 전달되는 정서를 혼합하여 받아들인다는 결과이다.¹⁹⁾ 이는 인간이 에이전트가 표현하는 정서인지에서도 인간이 표현하는 정서인지에서도 마찬가지로 교차양상지각(crossmodal perception)을 한다는 것을 의미한다.

2.2. 동적 얼굴 표정

몇 개의 예외를²⁰⁾²¹⁾제외하고 정서의 교차양상지각(crossmodal perception)에 관한 대부분의 연구들은 고정된 얼굴을 자극으로 사용하였다. 그러나 정서표현에 있어서 동적인 얼굴 표현이 정적인 표현보다 더 잘 인지된다는 연구²²⁾가 있을 뿐만 아니라 우리가 일상생활에서 마주하게 되는 얼굴 또한 가만히 있는 순간의 얼굴이 아니라 움직이며 역동적으로 정서를 표현하는 얼굴이다. 신경과학 분야의 연구에 따르면 뇌의 후부상측두구(posterior superior temporal sulcus)나 편도(amygdala)와 같이 얼굴의 정서를 처리하는 것으로 알려진 뇌의 부분이 고정되어 있는 정서 표현보다 움직이는 정서 표현에 더 반응하는 것으로 드러나고 있다.²³⁾²⁴⁾²⁵⁾

고정된 자극보다 역동적인 자극에 더 잘 반응하는 정서인지특성은 두 채널에서 정보를 받아들일 때에도 유지된다. 위에 언급한 de Gelder과 Vroomen의 연구에서 볼 수 있듯이 두 채널 간의 정보 연합은 움직임이 없는 얼굴 사진과 목소리가 주어졌을 때에도 일어나지만, 움직임이 있는 얼굴이 제시되었을 때에 훨씬 강하게 일어난다.²⁶⁾²⁷⁾²⁸⁾²⁹⁾

본 연구에서는 이런 정서인지의 특성을 고려하여 고정된 표정의 사진을 제시하기 보다는 중립적인 표정에서 해당 정서를 표현하는 표정으로 변화하는 애니메이션 자극을 사용하기로 하였다. 사진이 목소리를 낼 수 있다고 가정하는 것보다 동영상으로 제시하는 캐릭터가 목소리를 낸다는 것이 더 자연스럽다. 따라서 움직이는 얼굴 표정에 목소리가 동시에 제시되는 것은 교차 양상 지각 연구에 현실감을 더 높여줄 것이다.

2.3. 일관성

사람들은 어떤 현상을 대할 때 비일관성(inconsistency)보다 일관성(consistency)을 기대하는 경향이 있다. 만약 비일관성이 지각되면 인간의 지각체계는 비일관성을 가능한 일관되게 지각하려고 한다.³⁰⁾ HCI 분야에서도 위의 이론이 적용되는 예를 찾아볼 수 있는데, 목소리의 정서와 그 목소리가 읽어주는 내용의 정서가 일치하는 조건(기쁨-기쁨/슬픈-슬픈)과 일치하지 않는 조건(기쁨-슬픈)이 있을 때 사람들은 불일치 조건에서의 내용을 더 신뢰롭다고 평가했다.³¹⁾ 여기

facial expressions. *Neuroimage*, 18, 156 - 168.

- 19) Creed, C., & Beale, R. (2008). Psychological responses to simulated displays of mismatched emotional expressions. *Interacting with computers*, 20, 225-239.
- 20) de Gelder, B., Bo'cker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience Letters*, 260, 133 - 136.
- 21) Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage*, 37, 1445 - 1456.
- 22) Collier, G. (1985). *Emotional Expression*. Lawrence Erlbaum Associates.
- 23) Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, 4, 223 - 233.
- 24) LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*, 13, 1023 - 1033.
- 25) Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *Neuroimage*, 18, 156 - 168.
- 26) Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Science*, 11, 535 - 543.
- 27) Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, 25, 5004 - 5012.
- 28) Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly Journal of Experimental Psychology*, 60, 1446 - 1456.
- 29) Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M., (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *Journal of Neuroscience*, 26, 11138 - 11147.
- 30) Aronson, E. (1997). Back to the future: retrospective review of Leon Festinger's "A Theory of Cognitive Dissonance". *The American Journal of Psychology*, 110 (1), 127 - 37.

서 알 수 있는 것은 두 채널의 정서가 일치하지 않아도 사람들은 이를 모순적으로 여기지 않고 일치하지 않는 정서에서 의미를 찾는다라는 것이다. 실제로 사람들은 표정과 목소리에서 일치하지 않는 정서를 표현할 때가 있다. 아주 친한 친구가 지방에 취직이 되어서 이사가 가게 되었을 때 기쁘면서도 슬플 것이다. 그럼 표정은 기쁘게 짓지만 목소리는 슬프게 나올 수 있다. 이런 표정과 목소리에서 상대방은 섭섭함의 정서를 느낄 수 있을 것이다. 표정에서 전달하는 정서는 기쁘고 목소리가 전달하는 정서는 슬픔일지라도 이 둘이 합쳐지면서 섭섭함으로 인지되는 것이다.

위와 같은 사실에 따르면 인간이 에이전트의 정서를 파악할 때 두 채널로부터 각기 다른 정서가 입력된다면 교차 양상 지각이 발생해 제3의 정서를 인지할 가능성이 있다. 즉, 정서의 불일치로부터 단순히 인지적 부조화를 경험하는 것이 아니라 두 정서의 조합으로 하나의 정서를 느낄 수 있다. 만약 제3의 정서가 생성된다면, 이는 인간의 기본정서의 범주에 포함되기보다는 미묘한 정서 표현에 가까울 것이다. 인공물인 에이전트는 지금까지는 이런 미묘한 감정 표현(예, 섭섭함)이 거의 불가능했다. 왜냐하면 특정 정서에 맞는 특정 표정과 목소리를 전부 개별적으로 구현해야 하는 것은 매우 수고스러운 일이기 때문에 일반 기업에서는 극히 단순화된 에이전트의 정서 표현만을 제공해 왔다. 만약 몇 개의 기본정서를 가진 독립채널들의 조합으로 미묘한 정서 표현이 가능하다면 에이전트의 정서 표현도 풍부해질 것이다.

본 연구에서는 에이전트의 정서가 목소리와 표정 두 채널에서 서로 다양하게 조합되었을 때, 채널 간의 상호작용은 어떤 양상으로 사람들에게 어떤 정서로 인지될지와 두 채널의 서로 다른 정서 조합으로 창출되는 또 다른 제3의 정서가 있는지 알아보려고 한다.

가설 1 : 에이전트의 정서 인지에 있어서 어느 한 채널의 지배를 받는 것이 아니라, 두 채널(시각, 청각) 모두의 영향을 받을 것이다.

가설 2 : 표정과 목소리 두 채널에서 표현되는 정서가 다르면 각각의 독립 채널에서 표현된 정서와는 다른 또 다른 제3의 정서가 인지된다.

3. 방법

3.1. 실험 참가자

연구 참가자들은 심리학 과목을 수강하는 연세대학교 학부생 45명(M=20세, 남학생 26명/ 여학생 19명)이었다. 실험은 평균 20분 정도 소요 되었으며, 실험이 끝난 후 실험에 대한 간단한 설명과 그에 대한 보상으로 1 크레딧씩 주어졌다. 크레딧은 심리학실험에 참가했다는 것에 대한 확인의 개념이다.

3.2. 자극

3D 프로그램 MAYA로 제작된 서로 다른 정서를 표현하는 얼굴 표정(facial) 애니메이션 4조건(그림 1)과 서로 다른 정서를 표현하는 목소리 4조건이 조합된 16개의 자극(기쁜 표정-기쁜 목소리, 기쁜 표정-따뜻한 목소리, 기쁜 표정-차가운 목소리, 기쁜 표정-슬픈 목소리, 따뜻한 표정-기쁜 목소리 등등)을 이용하였다. 목소리자극은 정서를 포함하고 있지 않으면서 여러 가지 정서로 말해질 수 있는 “11시다”로 제작하였다. 에이전트의 얼굴 표정만을 자극으로 선택한 이유는 다른 행동적인 단서들이 끼어들어가는 것을 차단하고 순수하게 표정과 목소리로만 판단하게 하기 위함이었다. 얼굴 표정 애니메이션은 Ekman의 기술에 따른 정서 표현 표정의 예에 따라 제작하였다.³²⁾ 이때 얼굴 표정은 중립적 표정에서 각자의 정서를 표현하는 표정으로 움직이는 애니메이션이기 때문에 참가자는 표정에 나타난 정서를 정지된 화면보다 더 정확히 인식할 수 있다.³³⁾ 표정변화에서 정서를 전달해야했기에 립싱크는 하지 않았다.

31) Nass, C., Foehr, U., Brave, S., & Somoza, M. (2001). The effects of emotion of voice in synthesized and recorded speech. Paper presented at the Proceedings of the AAAI Symposium: Emotional and Intelligent II: The Tangled Knot of Social Cognition, North Falmouth, MA.

32) Ekman, P. (2004). Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life. Henry Holt & Co.

33) Desmet, P. (2002). Designing Emotions. Doctoral dissertation. Delft University of Technology.

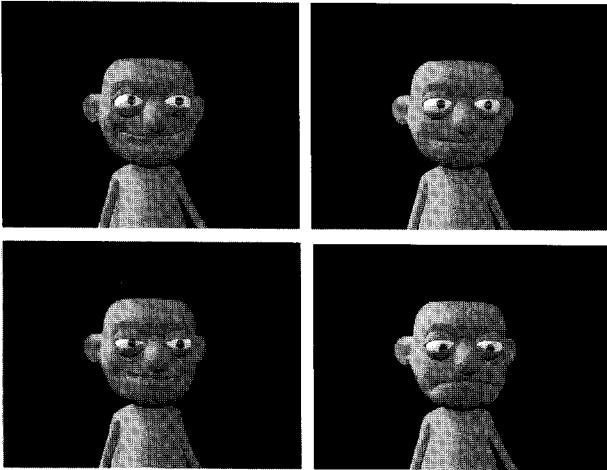


그림 1. 왼쪽 위부터 시계방향으로 기쁜 표정, 따뜻한 표정, 슬픈 표정, 차가운 표정(실제 자극은 정지화면이 아니라 중립적 표정에서 각자의 정서를 표현하는 표정으로 움직이는 애니메이션)

정서 평가는 양쪽에 형용사를 배치한 7점 리커트(likert) 척도 질문 4개(기쁜-슬픈 / 따뜻한-차가운 / 각성-이완 / 쾌-불쾌)와 정서를 자유롭게 기술하게 한 개방형 질문 1개로 받았다. 자극과 척도는 비주얼 베이직 6.0을 이용하여 한 화면에 구성되었다.

3.3. 조작검증

3.3.1. 실험 참가자

심리학 과목을 수강하는 연세대학교 학부생 12명(M=20세)으로 진행하고, 실험 종료 후 1크레딧씩 배부하였다.

3.3.2. 절차

목소리와 표정 자극을 단독으로 보여주고 5점 리커트 척도(기쁜/ 슬픈/ 따뜻한/ 차가운)로 그 에이전트가 표현하고 있는 표정이 각각의 정서에서 어느 정도 점수인지 평정 받았다. 목소리는 각각 다른 3사람의 목소리로 녹음된 기쁨, 따뜻함, 차가움, 슬픔의 4가지 정서 표현을(총 12 자극) 들려주었다. 표정은 기쁨, 따뜻함, 차가움, 슬픔의 4가지 정서를 표현한 총 4개 자극을 제시하였다. 표정 4개를 한 세트, 목소리 12개를 한 세트로 구성하여 세트 내에서 자극은 랜덤하게 제시하였다. 2개의 세트는 역균형을 적용하였다.

3.3.3. 결과

각 표정에 해당되는 5점 리커트 척도 질문 4개(기쁜/슬픈/따뜻한/차가운)에서 그 평균이 중간점수 3점보다 높은 정서를 선택해 One-sample test를 실시하였다. 차가운 표정과 슬픈 표정은 4개 종속측정치 중 각각 차가운, 슬픈 정서의 평균만 3점보다 높았다. 이에 대해 중간점수 3점과 One-sample test 결과 차가운 표정, 슬픈 표정은 중간점수 3점보다 유의미하게 높았다, $t(11)=4.062, p<.05$; $t(11)=2.708, p<.05$. 반면 기쁜 표정과 따뜻한 표정은 기쁜 정서와 따뜻한 정서 항목에서 동일하게 평균이 3점보다 높았다. 따라서 기쁜 표정에서 기쁜 정서와 따뜻한 정서 항목에 대해 중간점수 3점과 One-sample test 실시했고 그 결과는 기쁜 정서만 3점보다 유의미하게 높았다, $t(11)=7.707, p<.001$. 따뜻한 표정 역시 따뜻한 정서 항목에 대해서만 3점보다 유의미하게 높았다, $t(11)=3.546, p<.05$. 각각의 표정에 그에 해당되는 정서가 잘 표현된 것으로 나타났다.

목소리는 세 사람 중 한 명만이 4가지 목소리(기쁜, 따뜻한, 차가운, 슬픈) 각각에 대응되는 정서의 평균만 3점 이상이었다.(기쁜: M=4.5, SD=0.9 / 슬픈: M=4.08, SD=1.1/ 따뜻한: M=4.08, SD=0.6 / 차가운: M=3.91, SD=1.2) 표정과 달리 목소리는 각각의 정서를 표현하는 그 정서 종속측정치에서만 평균 3점 이상으로 나타났다. 더 정확한 분석을 위해 위의 평균들이 중간점수 3점과 실제로 차이가 나는지 알아보기 위해 One-sample test를 실시했다. One-sample test 결과 기쁜, 슬픈, 따뜻한, 차가운 목소리는 각각 중간점수 3점보다 유의미하게 높았다, $t(11)=5.745, p<.001$; $t(11)=3.223, p<.05$; $t(11)=5.613, p<.001$; $t(11)=2.561, p<.05$.

3.4. 실험설계방안

표정(4) X 목소리(4) 요인 방안이 실시되었다. 첫 번째 요인은 에이전트의 표정(기쁜, 따뜻한, 차가운, 슬픈)이었고, 두 번째 요인은 목소리(기쁜, 따뜻한, 차가운, 슬픈)였다. 실험은 피험자 내 설계로 모든 참가자들은 16 조건의 자극들을 보고 들으면서 그 조건에서의 에이전트의 정서를 평가하게 된다.

3.5. 절차

실험 참가자들은 자신이 신청한 시간에 실험에 참여했다. 참가자들은 실험에 참가하기에 앞서 실험의 목적, 진행절차, 동의서에 대한 설명을 듣고 연세대학교 심리학과 연구심의위원회의 승인을 받은 참가 동의서에 서명하였다.

실험은 전 과정이 비주얼 베이직6.0에 의해 제작된 양식을 가지고 진행이 되었다. 우선 참가자는 자신의 나이와 성별을 입력한다. 그 후에 참가자가 에이전트에 친근해지도록 에이전트가 소개되고, 본 실험과 동일한 화면이 제시되면서 실험에 대한 자세한 설명이 제공되었다. 이때 에이전트를 보고 참가자가 느낀 감정이 아니라 에이전트의 감정 자체를 평가하는 것이 목적임을 강조했다. 본 실험이 시작되면 왼쪽에 플래시 창이 나타나게 되고 오른쪽에 7점 리커트 척도로 표시된 정서 평가를 위한 4문항(기쁨-슬픔 / 따뜻한-차가운 / 각성-이완 / 쾌-불쾌)이 있고 그 아래 자유롭게 느껴지는 정서를 기술하게 한 개방형 문항이 1개 있다. 참가자들은 이 5개 문항에 모두 대답해야 다음 조건으로 넘어갈 수 있고, 개방형 문항에 생각나는 것이 없으면 ‘없다’라고 적도록 했다. 참가자들이 표정 4조건 목소리 4조건이 조합된 총 16개의 조건의 애니메이션을 보고 그에 답하는 것으로 실험은 종료되었다. 그리고 다시보기 버튼이 있어 참가자들은 원하는 만큼 자극을 다시 볼 수 있었다. 실험자는 이와 관련하여 에이전트의 표정과 목소리를 동시에 듣고 종합하여 응답하라고 강조하면서 잘 모르겠으면 다시보기 버튼을 눌러서 정서가 느껴질 때까지 자극을 재생하고, 그래도 모르겠다고 생각하면 7점 척도에서 중간값인 4점을 체크하도록 설명했다. 4점은 리커트 척도의 양 끝에 배치된 두 정서 모두 느껴지지 않음을 의미한다. 이를 강조한 것은 참가자들이 정서가 느껴지지 않음에도 불구하고 체크해야 한다는 압박감에 억지로 정서에 체크함을 방지하기 위해서이다. 다시보기 버튼을 통해 참가자들은 자신이 경험하는 정서를 어느 정도 확신할 수 있을 때에 반응할 수 있었다.

4. 결과

참가자들이 정서 표현을 어떻게 받아들였는지 알아보기 위해 반복표본변량 분석과 다차원 척도법을 사

용하였다. 특히 ‘기쁨-슬픔’과 ‘따뜻한-차가운’ 문항은 반복표본변량분석을 통해 서로 일치하지 않는 표정의 정서와 목소리의 정서가 결합하였을 때 정서인지에 있어서 어떤 경향이(표정이나 목소리 어느 한쪽의 정서가 지배적으로 인지되는지, 그렇지 않다면 정서가 어떤 방식으로 인지되는지) 나타나는지 알아보고자 하였다. 다차원 척도법을 통해서 다양한 표정과 목소리의 결합을 통한 각 자극의 위치를 살펴봄으로서 각 자극에서 인지되는 정서를 좀 더 자세히 살펴보고자 하였다.

4.1. 기쁨-슬픔

표정의 주효과와 목소리의 주효과가 둘 다 통계적으로 유의미한 것으로 나타났다. 각각 $F(3, 132) = 184.618, p < 0.001$, 그리고 $F(2.305, 101.399) = 69.663, p < 0.001$ 이었다. 표정과 목소리의 상호작용도 통계적으로 유의미한 것으로 나타났다. $F(9, 396) = 3.572, p < 0.001$. 그림 2에서 보듯이 표정의 정서나 목소리의 정서 어느 한쪽의 지배를 받는 것이 아니라 표정과 목소리의 조합에 따라 기쁨이나 슬픔을 느끼는 정도가 달라졌다. 기쁨 표정의 조건에서는 목소리에 따라 느껴지는 기쁨의 정도가 넓게 4.9에서 6.9까지 퍼져 있는 반면에 슬픈 표정의 조건에서는 목소리에 따라 느껴지는 슬픔의 정서가 상대적으로 그리 넓게 퍼져 있지 못하다. 이를 통계적으로 분석하기 위해서 기쁨 표정과 슬픈 표정 조건에서 목소리와의 상호작용에 대해 *within-subjects contrasts* 분석을 실시하였다. 그림 2를 보면 두 표정 조건에서 동일하게 기쁨 목소리와 슬픈 목소리가 최고, 최저 점수를 나타내고 있기 때문에, 기쁨 표정과 슬픈 표정에서 기쁨 목소리와 슬픈 목소리의 상호작용이 유의미하다면 각 표정에서 퍼져 있는 정서의 정도가 차이가 있다고 말할 수 있다. 기쁨 표정과 슬픈 표정에서 기쁨 목소리와 슬픈 목소리의 상호작용이 통계적으로 유의미하였다, $F(1, 44) = 22.495, p < 0.001$. 이는 기쁨 표정에서 기쁨 정서의 폭이 슬픈 표정에서의 슬픈 정서의 폭보다 넓음을 의미한다.

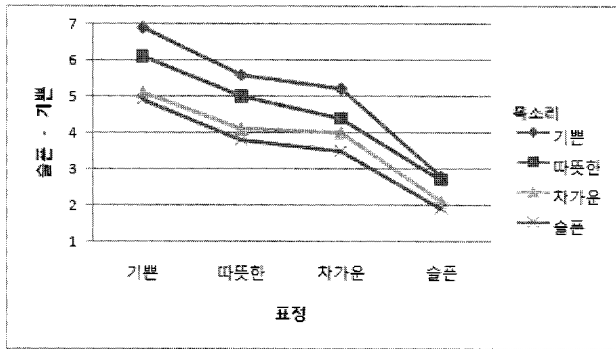


그림 2. 표정과 목소리에 따른 기쁨/ 슬픔의 정도

4.2. 따뜻함-차가움

표정과 목소리의 주효과가 모두 통계적으로 유의미한 것으로 나타났다. 각각 $F(3, 132) = 62.907, p < 0.001$, 그리고 $F(2.431, 106.974) = 28.972, p < 0.001$ 이었다. 표정과 목소리의 상호작용도 통계적으로 유의미한 것으로 나타났다. $F(6.590, 289.955) = 5.030, p < 0.001$.

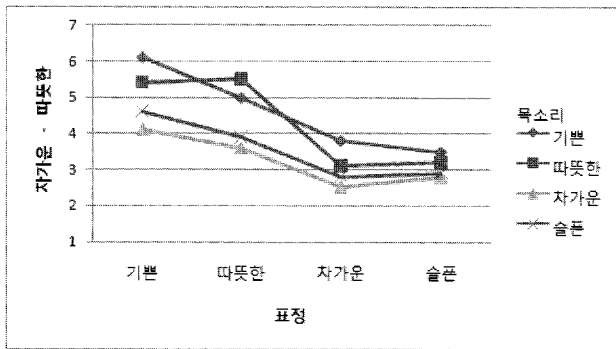


그림 3. 표정과 목소리에 따른 따뜻함과 차가움의 정도

그림 3에서 보듯이 기쁨-슬픔 척도와는 달리 따뜻한 표정 조건일 때와 차가운 표정 조건일 때의 정서가 느껴지는 정도가 확 떨어져 있는 것을 알 수 있다. 또한 따뜻한 표정에 따뜻한 목소리 조건일 때 가장 따뜻함의 정서를 많이 느낀 것을 볼 수 있다. 자극이 표현하고자 하는 해당 정서일 때 정서가 가장 잘 인지된다는 결과이다.

4.3. 또 다른 정서

표정과 목소리 정서들의 여러 조합을 통해서 목소리나 표정 단독에서는 느껴지지 않는 또 다른 정서가 느껴질지도 모른다는 두 번째 가설을 검증해 보기 위

해서 개방형 문항을 정리해 보았다. 우선 정서가 일치하는 조건을 정리하면, 기쁨 표정-기쁨 목소리에서는 기쁨, 기분 좋음이 슬픈 표정-슬픈 목소리에서는 슬픔, 우울이 따뜻한 표정-따뜻한 목소리에서는 호뭇함, 편안함이 차가운 표정-차가운 목소리에서는 차가움, 비웃음이 가장 많이 느껴졌다고 보고했다. 두 번째로 정서가 불일치하는 조건을 정리하면, 슬픈 표정-따뜻한 목소리, 슬픈 표정-차가운 목소리에서는 실망이 차가운 표정-기쁨 목소리, 따뜻한 표정-기쁨 목소리에서는 흥분, 들뜸이 느껴졌다는 반응이 많았다. 차가운 표정-슬픈 목소리, 따뜻한 표정-슬픈 목소리에서는 지친, 지루한, 씁쓸한 정서가 가장 많이 느껴졌다. 기쁨 표정-슬픈 목소리, 행복한 표정-차가운 목소리에서는 반가움과 기쁘게 슬픈 척한다는 기술이 많았다. 정서가 일치할 때 주로 기술된 정서는 조작된 정서와 일치한다. 이는 개방형 문항의 분석이 타당함을 시사한다. 그렇다면 정서가 불일치 할 때 기술된 정서를 참가자가 실제로 느꼈을 가능성이 높다. 수치적으로 이런 제3의 정서들에 대해 증명하기 위해 다차원 척도법을 사용했다.

그림 4에서 위쪽 그림은 Russell의 정서 동그라미이고,34) 아래쪽은 다차원 척도법을 이용하여 16가지 자극들을 사분면에 표현한 것이다. 그림에서 앞의 영어 철자는 표정(h: 기쁨, s: 슬픈, w: 따뜻한, c: 차가운), 뒤의 철자는 목소리를 나타낸다(예, hs: 기쁨 표정-슬픈 목소리). 가로축은 쾌-불쾌, 세로축은 각성-이완 차원으로 볼 수 있겠다. 위치상으로 두 그림을 비교해보면 상당히 잘 맞아 떨어지는 것을 알 수 있다. 더 정확히 비교하기 위해 실험에 사용된 4개의 정서 중에 Russell의 정서 동그라미에 포함되어 있는 기쁨(happy)과 슬픔(sad)을 선택했다. 기쁨 표정-기쁨 목소리 조건은 Russell의 정서 동그라미의 happy위치와 슬픈 표정-슬픈 목소리 조건은 Russell의 정서 동그라미의 sad위치와 맞아 떨어진다. 그러나 다른 위치에 찍힌 조건들을 Russell의 정서 동그라미상의 위치와 맞추어보는 것을 통해 그 조건의 정서를 예측하기에 좀 더 구체적인 증거가 필요했다. 예를 들어 cs(차가운 표정-스픈 목소리)가 3사분면에 찍혀 있기는 하지만, 이 위치와 Russell의 정서 동그라미의 위치 비교만으로 cs에서 느껴지는 정서가 지루함이나 우울한 것이라고 추측하기는 힘들었다. 그래서 개방형 문항을 통

34) Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.

해 얻은 형용사들의 빈도를 기준으로 느껴지는 정서를 그림에 함께 표시해 보았다.

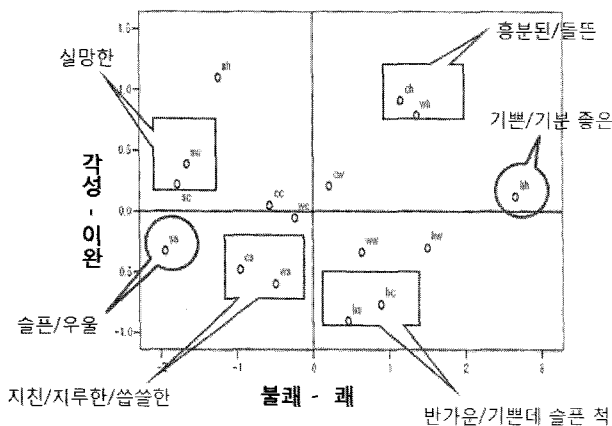
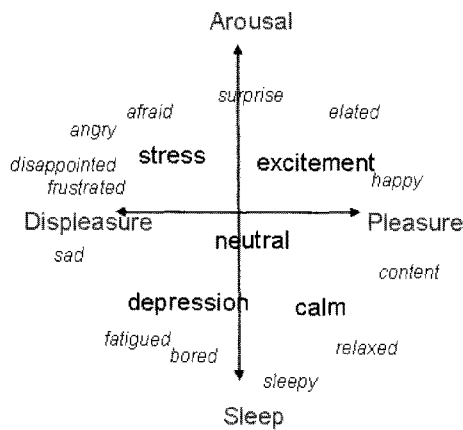


그림 4. Circumplex model of affect(Russell, 1980) / MDS결과 (해당 조건들에 대한 개방형 질문의 결과들을 말풍선으로 표시: 사각형 말풍선은 정서 불일치 조건, 원형 말풍선은 정서 일치 조건)

그림의 사분면상에서 각각의 조건들의 위치는 리커트 척도에 의해 얻어진 점수들에 근거한 결과이고 그 옆에 말풍선 안에 써넣은 정서들은 개방형 문항에서 뽑아낸 것인데도 위치상 비교해 보면 Russell의 정서 동그라미상의 정서와 크게 다르지 않은 것을 알 수 있다. 위에서 기술한 바와 같이 다차원 척도법상의 위치와 Russell의 정서 동그라미 위치상의 비교만을 통해서 목소리의 정서와 표정의 정서가 결합되어 해당 위치의 정서를 느꼈을 것이라고 예측하는 것은 불안정하다. 하지만 본 연구는 개방형 질문을 통해 참가자에게 직접 느껴지는 정서를 기술하도록 했고 그것이 위치상의 정서와 맞아떨어진다는 것은 위치상의 비교에 의미를 더해주는 것이다. 이렇게 서로 일관되지 않

은 표정상의 정서와 목소리상의 정서가 어우러졌을 때 알아볼 수 없거나 이상한 자극이 되어버리는 것이 아니라 또 다른 제3의 정서를 유발 수 있다는 것을 알 수 있었다.

5. 논의

본 연구의 결과 에이전트의 표정과 목소리에서 표현되는 정서가 다를 때 사람들은 어느 한 채널, 즉 표정이나 목소리 어느 한쪽의 정서를 따라가는 것이 아니라 표정과 목소리 두 채널의 정서 모두를 받아들여서 인지한다는 것을 알 수 있었다.

에이전트가 슬픈 표정일 때보다 기쁜 표정일 때 목소리에 따라서 기쁨을 느끼는 정도가 달랐다. 즉 기쁜 표정 일 때에는 목소리에 따라서 기쁨을 느끼는 정도가 다양했으나 슬픈 표정일 때에는 그 폭이 줄어들었다(그림 3). 이는 따뜻함-차가움 척도에서도 마찬가지로 결과였다. 이것은 사람들이 기쁜 표정을 지으면서 진심이 아닌 말들을 할 때가 더 많기 때문이라 추측해 볼 수 있다. 일상생활에서 기쁜 표정만으로는 상대방의 진심을 알 수가 없기에 에이전트 정서 판단에도 이것을 적용시켜서 기쁜 표정일 때는 슬픈 표정일 때보다 에이전트의 목소리를 더 많이 고려하는 것이다. 이와 같은 결과는 문화적 차이에 근거한 것이라고도 볼 수 있는데, 예를 들어 집단주의와 상호의존적 성향이 강한 동양에서는 집단 구성원 간의 조화로운 관계를 유지하기 위해 감정 표현을 통제할 수 있는 것이 중요하다고 여겨진다.³⁵⁾ 그래서 동양인들은 대체적으로는 기쁜 표정을 짓고 있지만, 다른 감정들이 목소리에 표현되는 경우가 있다. 따라서 동양인들은 타인의 진실한 감정 상태를 파악하기 위해서 두 채널의 정보를 종합적으로 고려하는 경향이 강할 것이라 짐작해 볼 수 있다. 실험의 참가자들의 이와 같은 성향이 에이전트의 감정 상태를 파악할 때도 동일하게 작용했다. 사전 조작 체크에서는 차가운 표정일 때 기쁨의 감정이 거의 느껴지지 않는다는 결과가 있었는데, 본 실험에서 차가운 표정과 기쁘거나 따뜻한 목소리가 결합하면 기쁨의 감정을 느끼는 것으로 나타났다. 이것은 사람들이 항상 자신의 얼굴에 정서를 확연하게 드러내면서 표현하지는 않기 때문이라고 생각된다. 특

35) Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S., (1999). Is there a universal need for positive self-regard. *Psychological Review*, 106, 766 - 794.

하나 애니메이션으로 표현된 차가운 정서라는 것이 진짜로 그렇게 매몰차 보이는 차가움이 아니라 자신의 목적을 달성한 듯 어느 정도 음흉해 보이기도 하는 종류의 차가움이라서 기쁨이 느껴진 것이라 생각한다.

언뜻 보기에는 기쁜 표정과 따뜻한 표정에 큰 차이가 없다고 생각할 수 있으나 따뜻함-차가움 척도를 통해 얻어진 결과에 의하면 기쁜 표정-따뜻한 목소리 조건에서보다 따뜻한 표정-따뜻한 목소리 조건에서 따뜻한 정서를 가장 많이 느꼈다고 평가되었다. 이 결과는 표정과 목소리의 정서가 정확히 일치할 때 정확한 정서 전달이 잘된다는 것을 의미한다. 동시에 사람들은 에이전트에서도 미묘한 정서의 차이를 경험할 수 있다는 것을 증명했다.

개방형 문항과 다차원 척도법의 분석 결과에 의하면 표정과 목소리의 정서가 일치할 때에는 일치하는 해당 정서를 경험하였다는(기쁜 표정-기쁜 목소리 조건에서는 기쁨, 행복함 등의 정서를, 슬픈 표정-슬픈 목소리 조건에서는 슬픈, 우울한 등의 정서를 보고함) 기술이 대부분이었다. 표정과 목소리의 정서가 일치하지 않을 때 사람들은 혼란스러워서 정서를 알 수 없게 되는 것이 아니라, 표정이나 목소리 단독 채널들에서는 표현되고 있지 않은 또 다른 제3의 정서를 느끼게 되는 것을 알 수 있었다. 구체적으로 사람들은 두 채널간의 조합에서 지루함, 지침, 실망, 반가움 등의 정서를 느끼는 것으로 나타났다. 그렇다면 에이전트로 표현 할 수 있는 표정의 종류나 목소리의 종류의 부족함을 그 둘 간의 조합을 통해 해결할 수 있을 것이다. 즉 다른 종류의 정서를 표현하기 위해서 일일이 다른 표정과 목소리를 다시 제작하는 것이 아니라 일정 수의 표정과 목소리의 조합으로 다른 정서들을 전달하는 것이 가능해질 것이다. 그렇게 된다면 훨씬 더 손쉽고 풍부하게 사람과 컴퓨터가 상호작용할 수 있을 것이다.

본 연구에서 사용된 에이전트는 인간과 그리 흡사하지는 않고 캐릭터의 특징이 보는 사람에게 어떤 감정을 불러일으킬 수 있다는 한계가 있다. 정서의 선택에 있어서도 따뜻함과 차가움은 일면 모호할 수 있는 정서였다. 또한 따뜻함을 표현하는 표정은 기쁨과 많이 닮아 있어 조작에 어려움이 있었던 것이 사실이다. 그러므로 일반적이고 뚜렷이 구별되는 정서를 선택해 보는 것도 좋을 듯하다.

향후에는 본 연구에서 얻어진 개방형 문항의 자료

들을 좀 더 분석하여 구체적으로 어떤 정서들의 조합이 어떤 특정 정서를 유발하는지 정리해 볼 수 있을 듯하다. 또한 에이전트의 정서를 맞춰보는 것이 아니라 에이전트가 표현한 정서를 보고 참가자가 느끼는 정서를 알아보는 것도 재미있는 연구가 될 것 이다. 보통 코미디 프로에서도 무표정으로 전혀 상상이 가지 않는 목소리로 말을 하면 웃음이 유발되는 것을 볼 수 있고 복화술 같은 것도 일종의 표정과 목소리가 일치하지 않음에서 오는 어떤 신기함이라고 생각 되기 때문이다.

참고문헌

- Aronson, E. (1997). Back to the future: retrospective review of Leon Festinger's "A Theory of Cognitive Dissonance". *The American Journal of Psychology*, 110(1), 127 - 37.
- Bartneck, C. (2001). Affective expressions of machines. *Extended Abstracts of CHI'01 : Conference on Human Factors in Computing Systems*, 189 - 190.
- Berry, D. C., Butler, L. T., & de Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies*, 63(3), 304-327.
- Bickmore, T. & Picard, R. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12 (2), 293 - 327.
- Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human Computer Studies* 62(2), 161-178.
- Bursleson, W. & Picard, R. (2004). Affective agents: sustaining motivation to learn through failure and a state of stuck. *Social and Emotional intelligence in learning environments workshop, in conjunction with the 7th International Conference on Intelligent Tutoring Systems*.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Science*, 11, 535 - 543.
- Collier, G. (1985). *Emotional Expression*. Lawrence

- Erlbaum Associates.
- Creed, C. (2006). Using computational agents to motivate diet change. In: Ijsselstein, W., De Kort, Y., Midden, C., & Van Den Hoven, E. (Eds.), *First international conference on Persuasive Technology for human well-being*. Springer-Verlag, Eindhoven University of Technology(the Netherlands), 100 - 103.
 - Creed, C., & Beale, R. (2008). Psychological responses to simulated displays of mismatched emotional expressions. *Interacting with computers*, 20, 225-239.
 - de Gelder, B., Boecker, K. B. E., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience Letters*, 260, 133 - 136.
 - de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and emotion*, 14 (3), 289-311.
 - Desmet, P. (2002). *Designing Emotions*. Doctoral dissertation. Delft University of Technology.
 - Ekman, P. (1992) An argument for basic emotions. *Cognition and emotion*, 6, 169-200.
 - Ekman, P. (2004). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. Henry Holt & Co.
 - Elliot, C., & Brezezinski, J. (1998). Autinimous agents as synthetic characters. *AI Magazine*, 19(2), 13-30.
 - Ghazanfar, A .A., Maier, J. X., Hoffman, K .L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, 25, 5004 - 5012.
 - Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior*, 24(4), 1494 - 1509.
 - Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Science*, 4, 223 - 233.
 - Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S., (1999). Is there a universal need for positive self-regard. *Psychological Review*, 106, 766 - 794.
 - Hietanen, J. K. Leppänen, J. M., Illi, M. & Surakka, V. (2004). Evidence for the integration of audiovisual emotional information at the perceptual level of processing. *European Journal of Cognitive Psychology*, 16(6), 769 - 790.
 - Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *Neuroimage*, 18, 156 - 168.
 - Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage*, 37, 1445 - 1456.
 - LaBar, K. S., Crupain, M. J., Voyvodic, J. T., & McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*, 13, 1023 - 1033.
 - Massaro, D. W., & Egan, P. B. (1996) Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, 3(2), 215-221.
 - Massaro, D. W., Cohen, M. M., Beskow, J., & Cole, R. (2000). Developing and evaluating Conversational Agents. In: Chassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.), *Embodied conversational agents*, MIT Press(Cambridge, MA), 287 - 318.
 - McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
 - Nass, C., Foehr, U., Brave, S., & Somoza, M. (2001). The effects of emotion of voice in synthesized and recorded speech. Paper presented at the Proceedings of the AAI Symposium: Emotional and Intelligent II: The Tangled Knot of Social Cognition, North Falmouth, MA.
 - Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6), 1161.
 - Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception and Psychophysics*, 54, 406-416.
 - Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing facial identities. *The Quarterly*

Journal of Experimental Psychology, 60, 1446 - 1456.

- Sugihara, T., Diltz, M .D., Averbek, B. B., & Romanski, L. M., (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. Journal of Neuroscience, 26, 11138 - 11147.
- Thompson, W. F., & Russo, F. A. (2007). Facing the music. Psychological Science, 18, 756-757.
- Thompson. W. F., Russo. F. A. & Quinto L. (2008). Audio-visual integration of emotional cues in song. Cognition and emotion, 22(8), 1457-1470.

원고접수 : 09.04.21

수정접수 : 09.06.22

게재확정 : 09.09.03